Optimal and efficient algorithms for learning high-dimensional, Banach-valued functions from limited samples

by

Sebastian Alfonso Moraga Scheuermann

Civil Mathematical Engineer, Universidad de Concepción, 2019
B.Sc., Universidad de Concepción, 2017

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Mathematics
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2024

Declaration of Committee

Name:	**Sebastian Alfonso Moraga Scheuermann**
Degree:	**Doctor of Philosophy**
Thesis title:	**Optimal and efficient algorithms for learning high-dimensional, Banach-valued functions from limited samples**
Committee:	**Chair:** Ben Ashby Assistant Professor, Mathematics

Ben Adcock
Supervisor
Professor, Mathematics

Nilima Nigam
Committee Member
Professor, Mathematics

Nick Dexter
Committee Member
Assistant Professor, Scientific Computing
Florida State University

Weiran Sun
Examiner
Associate Professor, Mathematics

Fabio Nobile
External Examiner
Professor, Mathematics
Swiss Federal Institute of Technology Lausanne

Abstract

Learning high- or infinite-dimensional functions from limited samples is a key task in Computational Science and Engineering (CSE). For example, in Uncertainty Quantification for CSE, a fundamental problem involves approximating solutions to parametric (or stochastic) Partial Differential Equations whose solutions take values in abstract spaces. While this problem has been studied for several decades, it remains difficult due to the significant challenges posed. Specifically, pointwise samples are expensive to acquire, errors may corrupt the data, and the ranges of these functions lie in Banach spaces while their domain is usually high- or infinite-dimensional. In this work, we combine recently developed approximation theory for holomorphic, high-dimensional functions asserting exponential (in finite dimensions) and algebraic (in both finite and infinite dimensions) convergence rates, along with recent methods from convex optimization and deep learning (DL) for computing approximations based on ℓ^1- or weighted ℓ^1-minimization strategies. We focus on overcoming the aforementioned challenges and closing key gaps between smooth high-dimensional function approximation theory and practice. First, we establish the existence of efficient algorithms based on the Chambolle-Pock algorithm for computing polynomial approximations to Hilbert-valued functions, achieving the same theoretical rates as current benchmarks. Our first main results account for all sources of error, i.e., polynomial approximation, sampling, algorithmic and physical discretization errors. Second, we present a novel result on DL for approximating smooth Banach-valued functions with known and unknown parametric dependence. Here, we extend key results from Compressed Sensing (CS) theory to Banach spaces. In summary, our second results assert the existence of a dimension-independent class of DNNs, whose training procedure is based on minimizing a regularized or unregularized ℓ^2-loss function, achieving near-optimal algebraic rates of convergence for holomorphic, infinite-dimensional Banach-valued functions. Next, we use the theory of m-widths to show that these convergence rates are near-optimal for infinite-dimensional Banach-valued, holomorphic functions. Finally, we present numerical experiments demonstrating the practical efficacy of DL on challenging problems including the parametric diffusion, Navier-Stokes-Brinkman and Boussinesq equations. In other words, the methods and algorithms developed in this thesis are essentially optimal for approximating holomorphic functions in high dimensions from limited samples.

Keywords: Banach-valued functions; high-dimensional approximation; compressed sensing; parametric PDEs; deep neural networks; optimal learning

Dedication

To my beloved wife, Ingrid Wang, for her patience and understanding during the long nights of writing. This thesis was only possible with her support and comforting hugs.

This work is also dedicated to my parents, Jeanette Scheuermann Renalqueo and José Hernán Moraga Valdivia, for their endless love and care throughout my life.

To my niece, Fernanda, and my sisters, Nixtala and Nathalie. I would not be where I am today without their support and encouragement.

Acknowledgements

Over the last decade, the concept of mathematics, in my mind, has become increasingly similar to an orchestra. I started this math journey in 2013, comparable to playing an out-of-tune "*Twinkle, Twinkle, Little Star*", by studying my first formal course in algebra and calculus. Slowly, a single quiet note in a peculiar frequency guided me to choose a more complex composition, a field in math called analysis, and to discover the applied mathematics that today plays a significant role in this thesis.

As the years pass, I realize that doing applied mathematics is never a one-person orchestra. You focus on the deadlines, the exams and assignments, staying up long nights trying to figure out an exercise, a theorem, or a class you missed. Slowly, the people around you start to add more strings to your instrument or stay to play along with you. You realize that your PhD thesis is a composition of learned lessons and time spent with the most important people in your life, arranging a beautifully complicated manuscript that, in the best-case scenario, will also help others in their journey. Today, I want to thank everyone who made this thesis possible.

From the academic side, I am profoundly grateful to my advisor, Prof. Ben Adcock, whose wisdom and patience have shaped the professional I am today. His guidance encouraged my perseverance and broadened my vision of communicating my research. I also want to thank Nick Dexter (Florida State University), who is part of my committee and contributed to every aspect of my PhD journey. The three of us met almost weekly to discuss our research progress and hold ourselves accountable, ultimately making this manuscript possible. Special thanks to Prof. Nilima Nigam, who always welcomed me with a big smile and the most amazing words of wisdom I could hear. Additionally, thanks to Prof. Gabriel Gatica (Udec). Without them, I wouldn't be where I am now.

I would love to dedicate a whole paragraph to everyone, but by doing so, this section would never end. I extend my gratitude to the fantastic people who shared their time with me, discussing math and other academic topics: Mahsa Faizrahnemoon, Simone Brugiapaglia, (Concordia University), Justin Gray, Petra Menz, John Stockie, Weiran Sun, Rita Li, Christie Carlson, and Rachel Tong. Overall, special thanks to the SFU Department of Mathematics.

Now, from my personal non-academic life, I want to thank the people who made these years the best of my life so far. First, my genuine thanks to my lovely wife, Ingrid Wang.

I am the luckiest husband to have such an amazing person by my side. As Prof. Nilima says, “You create better math when your heart is happy.” Ingrid, you made this possible, and I am very excited about what is to come for us as a family. I also want to thank my family in Chile: my parents, José and Jeanette, my niece Fernanda, and my sisters Nathalie and Nixtala. They supported me in every aspect throughout this journey. Additionally, a warm thanks to my family in Taiwan: my parents-in-law, Ken Wang and Grace Su, and my sister-in-law Jamie Wang. Thank you for opening your hearts and welcoming me into your beautiful family.

I want to thank my friends and people who, in one way or another, supported me these five years: Kshitij Patil, Paola Mazlum, Larissa Nicolau, Juan Manuel Cardenas, Javier Almonacid, Peter Phuong, Aneta Remesova, Patricio Alvarez, Fernando Bravo, Sebastian Dominguez, Daniela Soledad, Erik Mella, Chloe Huang, Matthew Spragge, Daniel Venn, Hannah Potgieter, Maria Flavia, Earl Decanay, Brendan Lee, Ehsan Far and Yasi Ahmadi.

Finally, I want to acknowledge the financial support of Simon Fraser University through the awards “TARA” (received many times to present our work at international and national conferences), the “Graduate Fellowship” (received twice) and the “Dept. of Mathematics Grad Scholarship” (received once). I want to thank my supervisor, Ben Adcock, who funded part of my research through research assistance funding.

Table of Contents

Declaration of Committee ii

Abstract iii

Dedication v

Acknowledgements vi

Table of Contents viii

List of Tables xiv

List of Figures xv

Notation and symbols xviii

1 Introduction 1

1.1 Problem statement . . . 1
1.2 Motivations . . . 2
1.3 Challenges . . . 4
1.4 Key considerations in this work . . . 7
1.5 Methods considered in this thesis . . . 9
1.5.1 Polynomial approximation . . . 9
1.5.2 Deep Learning . . . 9
1.6 Key questions in this thesis . . . 9
1.7 Outline and contributions of this thesis . . . 11
1.8 Literature review . . . 12
1.8.1 Chapter 3 . . . 13
1.8.2 Chapter 4 . . . 14
1.8.3 Chapter 5 . . . 14
1.8.4 Chapter 6 . . . 16
1.8.5 Chapter 7 . . . 16

2 General preliminaries 19

2.1 Notation . . . 19
2.2 Function spaces . . . 20
2.2.1 Measures and parametric domain . . . 20
2.2.2 $\mathcal{V}$-valued sequence spaces . . . 21
2.2.3 Lebesgue-Bochner and Sobolev spaces . . . 21
2.3 Holomorphy . . . 25
2.3.1 The class of $(\boldsymbol{b}, \varepsilon)$-holomorphic functions . . . 26
2.3.2 Parametric dependence: known and unknown cases . . . 30
2.3.3 Holomorphy and polynomial approximation . . . 31
2.4 Best s-term polynomial approximation . . . 31
2.4.1 Orthogonal polynomial expansions . . . 31
2.4.2 Sparsity and best s-term approximation error . . . 33
2.4.3 Rates of best s-term polynomial approximation . . . 34
2.4.4 Lower and anchored sets . . . 36
2.4.5 Weights and error bounds for Banach-valued functions . . . 37
2.4.6 Hyperbolic cross index sets . . . 41
2.5 Recovery of orthogonal polynomial coefficients . . . 42
2.5.1 Finite-dimensional approximation . . . 44
2.5.2 Unknown anisotropy recovery . . . 44
2.5.3 Known anisotropy recovery . . . 45
2.6 Deep learning . . . 46
2.6.1 Deep neural networks . . . 47
2.6.2 Recovery of coefficients via DNNs . . . 48
2.7 Main sources of error . . . 48

3 Compressed sensing for near-best polynomial approximation from limited samples 50
3.1 Preliminaries . . . 50
3.1.1 Setup . . . 51
3.1.2 Problem statement . . . 51
3.2 Contributions . . . 53
3.3 Main results . . . 54
3.4 Discussion . . . 58
3.5 The methods and proofs setup . . . 60
3.5.1 The methods in Theorems 3.3.1– 3.3.3 . . . 60
3.6 Compressed sensing . . . 60
3.6.1 The weighted robust Null Space Property . . . 61
3.6.2 Matrices satisfying the weighted rNSP over Banach spaces . . . 63
3.7 Error bounds for polynomial approximation via weighted SR-LASSO . . . 66

3.7.1 Overview . . . 66
3.7.2 The wRIP for the polynomial approximation problem . . . 67
3.7.3 Bounds for polynomial approximations obtained as inexact minimizers 68
3.8 Proofs of the main results: Theorems 3.3.1– 3.3.3 . . . 74
3.8.1 Theorem 3.3.1: algebraic rates of convergence, finite dimensions . . . 74
3.8.2 Theorem 3.3.2: algebraic rates of convergence, infinite dimensions . . 75
3.8.3 Theorem 3.3.3: exponential rates of convergence, finite dimensions . 76
3.9 Conclusions . . . 77
3.10 Future work . . . 78

4 Efficient algorithms for computing near-best polynomial approximations via compressed sensing from limited samples 79
4.1 Preliminaries . . . 79
4.1.1 Setup . . . 80
4.1.2 Problem statement . . . 80
4.2 Contributions . . . 81
4.3 Main results . . . 83
4.4 Discussion . . . 88
4.5 The construction of the algorithms in Theorems 4.3.1, 4.3.3 and 4.3.5 . . . 89
4.5.1 Equivalent minimization problems and method well-definedness . . . 90
4.5.2 The primal-dual iteration . . . 91
4.5.3 The primal-dual iteration for the weighted SR-LASSO problem . . . 93
4.5.4 The algorithms in Theorems 4.3.1, 4.3.3 and 4.3.5 . . . 95
4.6 An efficient restarting procedure; the algorithms used in Theorems 4.3.2, 4.3.4 and 4.3.6 . . . 96
4.7 The computational cost of the algorithms . . . 98
4.8 Error bounds and the restarting scheme for the primal-dual iteration . . . 101
4.8.1 Overview . . . 101
4.8.2 Error bounds for the primal-dual iteration . . . 101
4.8.3 The restarting scheme . . . 103
4.9 Proofs of the main results: Theorems 4.3.1–4.3.6 . . . 105
4.9.1 Theorems 4.3.1–4.3.2: algebraic rates of convergence, finite dimensions 105
4.9.2 Theorems 4.3.3–4.3.4: algebraic rates of convergence, infinite dimensions 108
4.9.3 Theorems 4.3.5–4.3.6: exponential rates of convergence, finite dimensions . . . 109
4.10 Conclusions . . . 110
4.11 Future work . . . 111

5 Deep neural networks for Banach- and Hilbert-valued approximations from limited samples 113

5.1 Preliminaries . . . 113
5.1.1 Setup . . . 114
5.1.2 Problem statement . . . 115
5.2 Contributions . . . 116
5.3 Main results . . . 118
5.3.1 Learning in the case of unknown anisotropy . . . 118
5.3.2 Learning in the case of known anisotropy . . . 120
5.4 Discussion . . . 122
5.5 Formulating the training problems and proof strategy . . . 125
5.5.1 Formulation as a vector recovery problem . . . 125
5.5.2 The class of DNNs $\mathcal{N}$ and the approximate measurement matrix . . . 126
5.5.3 Unknown anisotropy recovery . . . 126
5.5.4 Known anisotropy recovery . . . 127
5.6 Matrices satisfying the weighted rNSP over Banach spaces . . . 128
5.7 Deep neural network approximation . . . 129
5.7.1 Approximate multiplication via DNNs . . . 130
5.7.2 Emulation of orthogonal polynomials via DNNs . . . 132
5.8 Proofs of main results: Theorems 5.3.1–5.3.4 . . . 136
5.8.1 Theorem 5.3.1: unknown anisotropy, Banach-valued case . . . 136
5.8.2 Theorem 5.3.2: unknown anisotropy, Hilbert-valued case . . . 146
5.8.3 Theorem 5.3.3: known anisotropy, Banach-valued case . . . 149
5.8.4 Theorem 5.3.4: known anisotropy, Hilbert-valued case . . . 154
5.9 Conclusions . . . 157
5.10 Future works . . . 158

6 Optimal learning of holomorphic functions 160
6.1 Preliminaries . . . 160
6.1.1 Setup . . . 161
6.1.2 Sampling operators . . . 162
6.1.3 Adaptive m-widths . . . 163
6.1.4 Adaptive m-widths in the case of known and unknown anisotropy . . . 164
6.1.5 Problem statement . . . 165
6.2 Contributions . . . 165
6.3 Main results . . . 166
6.3.1 Lower bounds . . . 166
6.3.2 Upper bounds . . . 168
6.4 Discussion . . . 170
6.5 Proofs setup . . . 171
6.5.1 Notation . . . 171

6.5.2 Widths and standard results on widths . . . 172
6.6 Proof of main results: Theorems 6.3.1–6.3.4 . . . 173
6.6.1 Proof of Theorem 6.3.1: known anisotropy, lower bound . . . 173
6.6.2 Proof of Theorem 6.3.2: unknown anisotropy, lower bound . . . 178
6.6.3 Proof of Theorem 6.3.3: known anisotropy, upper bound . . . 179
6.6.4 Proof of Theorem 6.3.4: unknown anisotropy, upper bound . . . 181
6.7 Conclusions . . . 189
6.8 Future work . . . 190

7 Numerical approximation to parametric PDEs using DNNs 193
7.1 Preliminaries . . . 193
7.2 Setup . . . 194
7.2.1 Deep neural network approximation . . . 194
7.2.2 Methodology . . . 194
7.3 Contributions . . . 199
7.4 Main formulations . . . 199
7.4.1 The parametric coefficients . . . 200
7.4.2 The parametric diffusion equation . . . 200
7.4.3 The Navier-Stokes-Brinkman equations . . . 204
7.4.4 The stationary parametric Boussinesq equations . . . 207
7.5 Numerical results . . . 209
7.6 Additional discussion . . . 212
7.7 Conclusions . . . 212
7.8 Future work . . . 213

Bibliography 235

Appendix A Holomorphic maps for mixed formulations 254
A.1 Babuska-Brezzi theory for real and complex variables . . . 254
A.1.1 The inf-sup condition for b . . . 255
A.1.2 The BNB theorem for complex-valued Hilbert spaces . . . 256
A.2 The parametric diffusion equation . . . 258
A.2.1 Affine parametric dependence . . . 261
A.3 Holomorphic extension of the solution map . . . 261

Appendix B Legendre coefficients summability and best s-term polynomial approximation rates 268
B.1 Setup . . . 268
B.2 ℓ^p-summability and best s-term rates . . . 268
B.3 ℓ^p_{A}-summability and best s-term rates in anchored sets . . . 271

Appendix C Widths of weighted ℓ^p-norm balls in $\mathbb{R}^N$ **274**

List of Tables

Table 3.1 The methods $\mathcal{M} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K}$ used in Theorems 3.3.1, 3.3.2 and 3.3.3 61

Table 4.1 The algorithms $\mathcal{A} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K}$ used in Theorem 4.3.1, Theorem 4.3.3 and 4.3.5. 97

Table 4.2 The algorithms $\mathcal{A} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K}$ used in Theorems 4.3.2, 4.3.4 and 4.3.6. 100

Table 7.1 Shows the maximum requested and used resources for the Poisson problem for a single data point trought 12 different trials. 198

Table 7.2 Shows the maximum requested and used resources for the NSB problem for a single data point trought 12 different trials. 198

Table 7.3 Shows the maximum requested and used resources for the Boussinesq problem for a single data point trought 8 different trials. 199

List of Figures

Figure 7.1 Shows the domain for the parametric difussion equation. 203

Figure 7.2 Shows the solution $(\boldsymbol{u})(\boldsymbol{y})$ to the parametric Poisson problem in (7.4.4)–(7.4.5) for a given parameter $\boldsymbol{y} = (1, 0, 0, 0)^\top$ with affine coefficient a_1, utilizing a total of 732 degrees of freedom (DoF) for $\boldsymbol{u}$. The left displays the solution given by the FEM solver, while the right column shows the 4×40 ELU DNN approximation after 60000 epochs of training with $m = 500$ sample points. 204

Figure 7.3 The figure shows the solution $(\boldsymbol{u}, p)(\boldsymbol{y})$ to the parametric NSB problem in (7.4.9) for a given parameter $\boldsymbol{y} = (1, 0, 0, 0)^\top$ with affine coefficient a_1, utilizing a total of 1464 degrees of freedom (DoF) for $\boldsymbol{u}$ and 244 DoF for p. The left column displays the solution given by the FEM solver, while the right column shows the 4×40 ELU DNN approximation after 60000 epochs of training with $m = 500$ sample points. The top displays the magnitude of the vector field $\boldsymbol{u}$ and its direction with white arrows. On the bottom side, we show the pressure p. 215

Figure 7.4 The figure shows the solution $(\boldsymbol{u}, \varphi, p)(\boldsymbol{y})$ to the parametric Boussinesq problem in (7.4.13) for a given parameter $\boldsymbol{y} = (1, 0, 0, 0)^\top$ with an affine coefficient a_1. The solution utilizes a total of 18480 degrees of freedom (DoF) for $\boldsymbol{u}$ and 528 DoF for both φ and p. The left column displays the solution given by the FEM solver, while the right column shows the 4×40 ELU DNN approximation after 60000 epochs of training with $m = 500$ sample points. The top row displays streamlines of the vector field $\boldsymbol{u}$ and their direction with colored arrows. In the middle row, we visualize the temperature distribution inside the cube using colored spheres, with the hottest region at the center of the cube. The bottom row illustrates the points of highest pressure p. 216

Figure 7.5 Average relative $L^2_\varrho([-1, 1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ELU DNNs solving the Hilbert-valued diffusion equation in (7.4.4)–(7.4.5). 217

Figure 7.6 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ReLU DNNs solving the Hilbert-valued diffusion equation in (7.4.4)–(7.4.5). 218
Figure 7.7 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for tanh DNNs solving the Hilbert-valued diffusion equation in (7.4.4)–(7.4.5). 219
Figure 7.8 Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued NSB problem in (7.4.10). 220
Figure 7.9 Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued NSB problem in (7.4.10). 221
Figure 7.10 Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued NSB problem in (7.4.10). 222
Figure 7.11 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued NSB problem in (7.4.10). 223
Figure 7.12 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued NSB problem in (7.4.10). 224
Figure 7.13 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued NSB problem in (7.4.10). 225
Figure 7.14 Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 226
Figure 7.15 Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 227
Figure 7.16 Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 228
Figure 7.17 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^4(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $\varphi \in \mathrm{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 229
Figure 7.18 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^4(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $\varphi \in \mathrm{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 230

Figure 7.19 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^4(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $\varphi \in \mathrm{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 231
Figure 7.20 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 232
Figure 7.21 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 233
Figure 7.22 Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13). 234

Notation and symbols

The notation described below is used multiple times throughout the thesis

Indices, multi-indices, and vector notation

$\mathbb{N}$, $\mathbb{N}_0$	The sets of positive and nonnegative integers, respectively
ν	An index in $\mathbb{N}_0$
$\mathcal{F}$	The set of multi-indices in $\mathbb{N}_0^{\mathbb{N}}$ with at most finitely many nonzero terms
$\boldsymbol{\nu}$	A multi-index in $\mathbb{N}_0^d$ (or $\mathcal{F}$, if $d = \infty$)
$a \lesssim b$	There exists a constant $c > 0$ independent of a and b such that $a \leq cb$
$\boldsymbol{a}^{\boldsymbol{\nu}}$	The product $\prod_{k\in[d]} a_k^{\nu_k}$ for $\boldsymbol{a} = (a)_{k\in[d]}$ and $\boldsymbol{\nu} = (\nu)_{k\in[d]}$
$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d$	The canonical basis of $\mathbb{R}^d$
$[d]$	Set notation to denote $\{1, \ldots, d\}$, equal to $\mathbb{N}$ if $d = \infty$
$\mathbb{R}_+$	The set of positive numbers
$\operatorname{div}(\cdot)$, $\mathbf{div}(\cdot)$	Vector and tensor divergence operator

Function approximation

d	Number of parametric variables, either $d \in \mathbb{N}$ or $d = \infty$
$\mathcal{U}$	Parameter domain, a subset of $\mathbb{R}^d$ (or $\mathbb{R}^{\mathbb{N}}$, if $d = \infty$)
Ω	Physical domain, e.g., $\Omega = (0,1)^2 \subset \mathbb{R}^2$
$\partial\Omega$	Boundary of the physical domain Ω
$\mathcal{V}$	Output space, typically a scalar field, or a Banach space
$\mathcal{V}^*$	Continuous dual of the output space
$\mathcal{V}_K$	Finite dimensional subspace of the output space $\mathcal{V}$
$\boldsymbol{y} = (y_k)_{k\in[d]}$	parametric variable, $\boldsymbol{y} \in \mathcal{U}$
$f = f(\boldsymbol{y})$	Generic Banach-valued parametric map from $\mathcal{U}$ to $\mathcal{V}$
$u = u(\boldsymbol{y})$	Banach-valued parametric map from $\mathcal{U}$ to $\mathcal{V}$, typically the solution to a parametric differential equation
$\hat{f}, \hat{u}$	Approximation to f or u
m	Number of samples
$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$	Sample points in $\mathcal{U}$
n_i	Measurement error; noisy samples take the form $f(\boldsymbol{y}_i) + n_i$
$\mathcal{P}_K(\cdot)$	Bounded and linear operator $\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K$ (if $\mathcal{V}$ is a Hilbert space, $\mathcal{P}_K$ is the orthogonal projection from $\mathcal{V}$ to $\mathcal{V}_K$)

Polynomial approximation

Ψ_ν	νth element of the one-dimensional orthonormal polynomial basis of $L^2_\varrho([-1,1])$
$\Psi_{\boldsymbol{\nu}}$	$\boldsymbol{\nu}$th element of the d-dimensional orthonormal polynomial basis of $L^2_\varrho(\mathcal{U})$
$\boldsymbol{c}_{\boldsymbol{\nu}}$	Coefficients of a function with respect to $\Psi_{\boldsymbol{\nu}}$
S	Typically, a finite subset of $\mathbb{N}_0^d$ (or $\mathcal{F}$, if $d = \infty$) of size $\lvert S\rvert$
$\Lambda^{\mathsf{HC}}_{n,d}$	The hyperbolic cross index set of order $n \in \mathbb{N}_0$
f_Λ	Truncated expansion of f with respect to $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$
$\Theta(n,d)$	The cardinality of the index set employed

Indices, multi-indices, and vector notation

$\mathcal{O}$	Typically, an open subset of $\mathbb{C}^d$ or $\mathbb{C}^{\mathbb{N}}$ (if $d = \infty$) in which a function has a holomorphic extension
$\rho, \boldsymbol{\rho} = (\rho_k)_{k\in[d]}$	Parameter $\rho > 1$ or $\boldsymbol{\rho} > 1$ defining a Bernstein ellipse or polyellipse
$\mathcal{B}_\rho, \mathcal{B}_{\boldsymbol{\rho}}$	The Bernstein ellipse or polyellipse with parameter ρ or $\boldsymbol{\rho}$
$\mathcal{E}_\rho, \mathcal{E}_{\boldsymbol{\rho}}$	The corresponding filled-in Bernstein ellipse or polyellipse with parameter ρ or $\boldsymbol{\rho}$
$\mathcal{R}(\boldsymbol{b}, \varepsilon)$	Complex region defined by a union of filled-in Bernstein polyellipses
$\mathcal{H}(\boldsymbol{b}, \varepsilon)$	Class of $(\boldsymbol{b}, \varepsilon)$-holomorphic functions with L^∞-norm at most one
$\mathcal{H}(p)$	The union of $\mathcal{H}(\boldsymbol{b}, 1)$ for $\|\boldsymbol{b}\|_p \le 1$ with $0 < p < 1$
$\mathcal{H}(p, \mathsf{M})$	The union of $\mathcal{H}(\boldsymbol{b}, 1)$ for $\|\boldsymbol{b}\|_{p,\mathsf{M}} \le 1$ with $0 < p < 1$

Sequences

Λ	A multi-index set in $\mathbb{N}_0^d$ (or $\mathcal{F}$, if $d = \infty$), possibly finite or infinite
$\boldsymbol{c}_\Lambda = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$	A vector or sequence with indices in Λ
$\operatorname{supp}(\boldsymbol{c})$	The support of $\boldsymbol{c}$, i.e., the set of multi-indices $\boldsymbol{\nu}$ for which $c_{\boldsymbol{\nu}} \neq 0$
$\boldsymbol{\nu}$	A multi-index in $\mathbb{N}_0^d$ (or $\mathcal{F}$, if $d = \infty$)
$\boldsymbol{w} = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$	A sequence of nonnegative weights
$\boldsymbol{u} = (u_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$	The intrinsic weights, defined by $u_{\boldsymbol{\nu}} = \|\Psi_{\boldsymbol{\nu}}\|_{L^\infty(\mathcal{U})}$
$\|S\|_{\boldsymbol{w}}$	The weighted cardinality of a set $S \subseteq \Lambda$
$m(\Lambda)$	$m(\Lambda) = \max_{\boldsymbol{\nu}\in\Lambda} \|\boldsymbol{\nu}\|_1$

Neural networks

$\Phi_{\boldsymbol{\nu}}$	DNN approximation to a basis function $\Psi_{\boldsymbol{\nu}}$
$\mathcal{N}$	A class of DNN
$\boldsymbol{A}'$	Approximate measurement matrix defined by $\Phi_{\boldsymbol{\nu}}$
$\mathcal{J}(\cdot)$	Regularization functional $\mathcal{J} : \mathcal{N} \to [0, \infty)$
$\sigma(\cdot)$	Nonlinear activation function
$\mathcal{A}_\ell(\cdot)$	Affine linear map $\mathcal{A}_\ell : \mathbb{R}^{N_\ell} \to \mathbb{R}^{N_{\ell+1}}$, acting on the ℓ-th layer
$\boldsymbol{W}_\ell$	Weight matrix acting on the ℓ-th layer
$\boldsymbol{b}_\ell$	Bias vector acting on the ℓ-th layer
$\mathcal{L}(\cdot)$	Loss function
$\mathcal{T}_\Theta(\cdot)$	Variable restriction operator $\mathcal{T}_\Theta : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^n$ with $\lvert\Theta\rvert = n$

Optimization

argmin	The set of (global) minimizers of an optimization problem
$\boldsymbol{A}$	Measurement matrix defined by $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\Lambda}$
$\boldsymbol{f}$	(Noisy) measurement vector
λ	Regularization parameter in an optimization problem
$\hat{\boldsymbol{c}}$	The reconstruction of a vector $\boldsymbol{c}$ via an optimization problem
$\mathcal{G}$	Objective function associated to a minimization problem

Optimal learning

$\boldsymbol{\mathcal{R}}(\cdot)$	Arbitrary reconstruction map $\mathcal{R} : \mathcal{V}^m \to L^2_\varrho(\mathcal{U}; \mathcal{V})$
$\boldsymbol{\mathcal{S}}(\cdot)$	Adaptive sampling operator $\boldsymbol{\mathcal{S}} : \mathcal{Y} \to \mathcal{V}^m$
$\boldsymbol{\Delta}(\cdot)$	Scalar-valued reconstruction mapping $\boldsymbol{\Delta} : \mathbb{R}^m \to \mathbb{R}^N$
$\boldsymbol{\Gamma}(\cdot)$	Scalar-valued adaptive sampling operator $\boldsymbol{\Gamma} : \mathbb{R}^N \to \mathbb{R}^m$
$B^p_N(\boldsymbol{w})$	Weighted unit ball of elements in $\ell^p_N(\boldsymbol{w})$
$\theta_m(\boldsymbol{b})$	(Adaptive) m-width $\Theta(\mathcal{H}(\boldsymbol{b}); \mathcal{Y}, L^2_\varrho(\mathcal{U}; \mathcal{V}))$ where $\mathcal{Y}$ is a subspace of $\mathcal{X}$
$\theta_m(p)$	(Adaptive) m-width $\Theta(\mathcal{H}(p); \mathcal{Y}, L^2_\varrho(\mathcal{U}; \mathcal{V}))$ where $\mathcal{Y}$ is a subspace of $\mathcal{X}$
$\theta_m(p, \mathsf{M})$	(Adaptive) m-width $\Theta(\mathcal{H}(p, \mathsf{M}); \mathcal{Y}, L^2_\varrho(\mathcal{U}; \mathcal{V}))$ where $\mathcal{Y}$ is a subspace of $\mathcal{X}$
$\overline{\theta_m}(p)$	Supremum of $\theta_m(\boldsymbol{b})$ over $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with $\|\boldsymbol{b}\|_p \leq 1$
$\overline{\theta_m}(p, \mathsf{M})$	Supremum of $\theta_m(\boldsymbol{b})$ over $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1$
$d^m(\mathcal{K}, \mathcal{X})$	Gelfand m-width of a subset $\mathcal{K}$ of a normed space $(\mathcal{X}, \|\cdot\|_\mathcal{X})$
$E^m_{\mathsf{ada}}(\mathcal{K}, \mathcal{X})$	Adaptive compressive m-width of a subset $\mathcal{K}$ of a normed space $(\mathcal{X}, \|\cdot\|_\mathcal{X})$

Sequences spaces and products

$\ell^p(\Lambda)$	The space of ℓ^p-summable, scalar-valued sequences indexed over Λ, where $0 < p \leq \infty$
$\ell^p_N(\boldsymbol{w})$	The finite dimensional space $(\mathbb{R}^N, \|\cdot\|_{p,\boldsymbol{w}})$
$\langle \cdot, \cdot \rangle$	The Euclidian inner product on $\ell^2(\Lambda)$
$\sigma_s(\cdot)_p$	The ℓ^p-norm best s-term approximation error
$\sigma_{s,\mathsf{L}}(\cdot)_p$	The ℓ^p-norm best s-term approximation error in lower sets
$\tilde{\boldsymbol{z}}$	The minimal monotone or anchored majorant of a sequence $\boldsymbol{z} \in \ell^\infty(\Lambda)$, where $\Lambda = \mathbb{N}_0^d$ or $\Lambda = \mathcal{F}$ (if $d = \infty$)
$\ell^p_{\mathsf{M}}(\Lambda; \mathcal{V})$, $\ell^p_{\mathsf{M}}(\Lambda)$	The space of Banach-valued or scalar-valued sequences with ℓ^p-summable monotone majorants, where $0 < p \leq \infty$ and $\Lambda = \mathbb{N}_0^d$ or $\Lambda = \mathcal{F}$ (if $d = \infty$)
$\ell^p_{\mathsf{A}}(\Lambda; \mathcal{V})$, $\ell^p_{\mathsf{A}}(\Lambda)$	The space of Banach-valued or scalar-valued sequences with ℓ^p-summable anchored majorants, where $0 < p \leq \infty$ and $\Lambda = \mathbb{N}_0^d$ or $\Lambda = \mathcal{F}$ (if $d = \infty$)
$(k, \boldsymbol{w})$	Weighted sparsity $k \geq 0$ with respect to the weights $\boldsymbol{w} \geq 0$
$\sigma_k(\cdot)_{p,\boldsymbol{w}}$	The $\ell^p_{\boldsymbol{w}}$-norm best $(k, \boldsymbol{w})$-term approximation error
$\boldsymbol{u} \cdot \boldsymbol{v}$	Component-wise product of vector for functions in $\mathbf{L}^p(\Omega)$
$\boldsymbol{u} : \boldsymbol{v}$	Component-wise product of tensors for functions in $\mathbb{L}^p(\Omega)$
$\boldsymbol{w} \odot \boldsymbol{v}$	The Hadamard product $\boldsymbol{w} \odot \boldsymbol{v} = (w_i v_i)_{i \in [N]}$ for $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^N$
$\rho \otimes \rho$	Tensor product for measures, e.g., $\rho^{(d)} = \rho \otimes \cdots \otimes \rho$
$\boldsymbol{w} \otimes \boldsymbol{v}$	Tensor product of vectors, e.g., $\boldsymbol{w} \otimes \boldsymbol{v} = (w_i v_j)_{i,j \in [N]}$ for $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^N$
$; \mathcal{V}$	Indicates spaces, norms, and approximation errors of sequences that take values in a Banach space $\mathcal{V}$, such as $\ell^p(\Lambda; \mathcal{V})$, $\|\cdot\|_{p;\mathcal{V}}$, or $\sigma_s(\cdot)_{p;\mathcal{V}}$

Function spaces

ϱ	Either a probability measure on $[-1,1]$ or the resulting tensor-product probability measure on $\mathcal{U} = [-1,1]^d$ (or $\mathcal{U} = [-1,1]^{\mathbb{N}}$, if $d = \infty$)
$L^p_\varrho(\mathcal{U})$	The space of p-integrable scalar-valued functions $f : \mathcal{U} \to \mathbb{C}$ with respect to the measure ρ, with $1 \leq p < \infty$
$L^\infty(\mathcal{U})$	The space of essentially bounded scalar-valued functions $f : \mathcal{U} \to \mathbb{C}$
$C(\mathcal{U})$	The space of continuous scalar-valued functions $f : \mathcal{U} \to \mathbb{C}$
$L^p_\varrho(\mathcal{U};\mathcal{V})$	The space of p-integrable Banach-valued functions $f : \mathcal{U} \to \mathcal{V}$ with respect to the measure ρ, with $1 \leq p < \infty$
$L^\infty(\mathcal{U};\mathcal{V})$	The space of essentially bounded Banach-valued functions $f : \mathcal{U} \to \mathcal{V}$
$\mathcal{Y}$	Normed vector subspace of the Lebesgue-Bochner space $L^2_\varrho(\mathcal{U};\mathcal{V})$
$C(\mathcal{U};\mathcal{V})$	The space of continuous Banach-valued functions $f : \mathcal{U} \to \mathcal{V}$ with respect to the uniform norm
$\mathrm{L}^p(\Omega)$	Standard Lebesgue space of functions with a physical domain Ω
$W^{s,p}(\Omega)$	Standard Sobolev spaces with $s \in \mathbb{R}$ and $p \geq 1$
$\mathrm{H}^1(\Omega)$	Sobolev space with $s = 1$ and $p = 2$
$\mathrm{H}^{1/2}(\partial\Omega)$, $\mathrm{H}^{-1/2}(\partial\Omega)$	Space of traces of functions in $H^1(\Omega)$, and its dual
$\mathrm{H}^1_0(\Omega)$	Sobolev space of functions in $\mathrm{H}^1(\Omega)$ with zero trace on the boundary Ω
$\mathrm{L}^2_0(\Omega)$	Space of functions in $\mathrm{L}^2(\Omega)$ with zero mean
$\mathbb{L}^2(\Omega)$	Space of tensor functions with each component in $\mathrm{L}^2(\Omega)$
$\mathbb{L}^2_{\mathrm{skew}}(\Omega)$	Space of skew-symmetric tensor functions in $\mathbb{L}^2(\Omega)$
$\mathbb{L}^2_{\mathrm{tr}}(\Omega)$	Space of trace-free tensor functions in $\mathbb{L}^2(\Omega)$
$\mathbf{H}(\mathrm{div}_p;\Omega)$	Space of vector functions $\boldsymbol{v} \in \mathbf{L}^2(\Omega)$ with $\mathrm{div}(\boldsymbol{v}) \in \mathrm{L}^p(\Omega)$, with $p \geq 1$
$\mathbb{H}(\mathbf{div}_p;\Omega)$	Space of tensor functions $\boldsymbol{v} \in \mathbb{L}^2(\Omega)$ with $\mathbf{div}(\boldsymbol{v}) \in \mathbf{L}^p(\Omega)$, with $p \geq 1$
$\mathbb{H}_0(\mathbf{div}_p;\Omega)$	Space of tensor functions in $\mathbb{H}_0(\mathbf{div}_p;\Omega)$ with zero mean trace
$\mathbb{H}(\mathbf{div};\Omega)$, $\mathbb{H}(\mathbf{div};\Omega)$	Special case of vector and tensor functions in $\mathbf{H}(\mathrm{div}_2;\Omega)$ and $\mathbb{H}(\mathbf{div}_2;\Omega)$, respectively

Norms

$\|\cdot\|_p$	The usual vector ℓ^p-norm (if $1 \le p \le \infty$) or quasi-norm (if $0 < p < 1$)
$\|\cdot\|_{p,q}$	The matrix $\ell^{p,q}$-norm (if $1 \le p \le \infty$)
$\|\cdot\|_{p,\mathsf{M};\mathcal{V}}$, $\|\cdot\|_{p,\mathsf{M}}$	The norms on $\ell^p_{\mathsf{M}}(\Lambda;\mathcal{V})$ and $\ell^p_{\mathsf{M}}(\Lambda)$ (if $0 < p \le \infty$)
$\|\cdot\|_{p,\mathsf{A};\mathcal{V}}$, $\|\cdot\|_{p,\mathsf{A}}$	The norms on $\ell^p_{\mathsf{A}}(\Lambda;\mathcal{V})$ and $\ell^p_{\mathsf{A}}(\Lambda)$ (if $0 < p \le \infty$)
$\|\cdot\|_{p,\boldsymbol{w}}$	The $\ell^p_{\boldsymbol{w}}$-norm (if $1 \le p \le 2$) or quasi-norm (if $0 < p < 1$)
$\|\cdot\|_{\mathcal{V}}$	The norm on a Banach space $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$
$\|\cdot\|_{p;\mathcal{V}}$	The norm on $\ell^p(\mathcal{F};\mathcal{V})$ (if $1 \le p \le \infty$)
$\|\cdot\|_{L^\infty(\mathcal{U})}$	The norm on $L^\infty(\mathcal{U})$
$\|\cdot\|_{L^p_\varrho(\mathcal{U})}$, $\langle\cdot,\cdot\rangle_{L^2_\varrho(\mathcal{U};\mathcal{V})}$	The norm and inner product on $L^p_\varrho(\mathcal{U})$ and $L^2_\varrho(\mathcal{U})$, respectively

Errors

E_{app}	Approximation error in the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm
E^∞_{app}	Approximation error in the $L^\infty(\mathcal{U};\mathcal{V})$-norm
$E_{\mathsf{app},\mathsf{UB}}$, $E_{\mathsf{app},\mathsf{UH}}$, $E_{\mathsf{app},\mathsf{KB}}$, $E_{\mathsf{app},\mathsf{KH}}$	Variant of E_{app} in unknown (U) or known (K) anisotropy case and the Banach (B)- or Hilbert (H)-valued case
$E^\infty_{\mathsf{app},\mathsf{UB}}$, $E^\infty_{\mathsf{app},\mathsf{UH}}$, $E^\infty_{\mathsf{app},\mathsf{KB}}$, $E^\infty_{\mathsf{app},\mathsf{KH}}$	Variant of E^∞_{app} in unknown (U) or known (K) anisotropy case and the Banach (B)- or Hilbert (H)-valued case
E_{disc}	Physical discretization error in the $L^\infty(\mathcal{U};\mathcal{V})$-norm
E_{samp}	Sampling error
E_{opt}	Optimization error

Miscellaneous

$\mathbb{P}$	Probability
$\mathbb{E}$	Expectation
$\sim$	Indicates that a variable follows a certain distribution
$\cup, \cap$	Union and intersection operators, respectively

Chapter 1

Introduction

In order to understand our world, we rely on science to make forecasts and look for patterns in nature, making observations and comparing them to our description of the universe through the language of mathematics. Computers have become fundamental tools for achieving better and more reliable predictions of physical phenomena. We have seen scientific computing providing support and faster simulations in various fields, such as economics, biology, engineering, chemistry and physics, to name a few. However, modern scientific computing still faces challenges describing complex phenomena involving specific quantities of interest in those fields.

An essential task in scientific computing is the approximation of specific quantities of interest involving physical phenomena depending on parameters. The setting for this thesis is high-dimensional problems coming from computational science and engineering, especially Uncertainty Quantification (UQ) [27,56,119,176,209,239,246]. Its contributions are optimal and efficient methods for learning high-dimensional, Banach-valued functions from limited data with theoretical error guarantees and sample complexity bounds. The material in this thesis is based on [7,8,10,18].

1.1 Problem statement

In particular, we study parametric models that take a parametric variable $\boldsymbol{y}$ and give as an output $f(\boldsymbol{y})$. We assume that $\boldsymbol{y}$ belongs to a parametric domain $\mathcal{U}$, typically a subset of $\mathbb{R}^d$, where $d \in \mathbb{N} \cup \{\infty\}$, and f takes values in a space $\mathcal{V}$. Specifically, given m sample points $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in \mathcal{U}$ we consider approximations to

$$f : \mathcal{U} \to \mathcal{V}, \boldsymbol{y} \mapsto f(\boldsymbol{y}),$$

from its m sample values (or snapshots)

$$f(\boldsymbol{y}_1), \dots, f(\boldsymbol{y}_m) \in \mathcal{V}. \tag{1.1.1}$$

In this thesis, the mechanism computing the sample values $f(\boldsymbol{y})$ is assumed to be available and is treated as a black box. As noted in [12, Rmk. 1.1], we refer to methods that construct an approximation of f from its sample values as nonintrusive methods. In practice, one may only want to approximate a certain scalar-valued *quantity of interest* $\mathcal{Q}(f(\boldsymbol{y}))$ of f. Typical examples include the mean of f over some physical domain, or its evaluation at a specific point. However, approximating f itself allows one to subsequently approximate arbitrarily many quantities of interest of it. See, e.g., the last paragraph of [12, §1.2.3].

Due to their relevance in the main motivating applications, we consider smooth functions in this thesis. The formal definition of a *holomorphic* function will be presented in section §2.3. Informally, we say that a target function f of a parametric model is smooth when it is holomorphic with respect to its parameters $\boldsymbol{y}$. Hence, it is natural to compare our theoretical results with those current benchmark methods for smooth functions, e.g., the *best s-term polynomial approximation* as suggested in [12].

1.2 Motivations

Although our work applies more generally, this thesis is primarily motivated by parametric Differential Equations (DEs) [63, 71]. In UQ, many applied problems are posed in terms of (systems of) Ordinary Differential Equations (ODEs) or Partial Differential Equations (PDEs) that are smooth with respect to their parameters (see §1.6 and §2.3). In this setting, given a parameter $\boldsymbol{y} \in \mathcal{U}$, obtaining the output $f(\boldsymbol{y}) \in \mathcal{V}$ may involve an expensive numerical simulation to approximately solve the PDE. This fact motivates the construction of efficient algorithms to accurately approximate f from the least amount of samples m as possible.

Parametric DEs

In most applications involving the solution of a DE, obtaining measurements from the real world may introduce uncertainties in some parameters. For example, in fluid dynamics, some mechanical properties of the fluid may be uncertain due to changes in the environment, errors in measurement devices or lack of computational resources. These uncertainties can be modelled as random variables, and the resulting stochastic differential equations can be reformulated as deterministic PDEs [156, §2.1.1]. Then, using this framework, relevant and valuable statistics from solutions of PDEs with random input data can be computed [268, Chp. 1]. As mentioned in [156, §1.1.2], in practice, this is done by computing the solution of the underlying PDE for a large set of values of $\boldsymbol{y}$ using expensive iterative solvers, e.g., fixed-point iteration methods.

In such problems, one considers a function $u = u(\boldsymbol{x}, \boldsymbol{y})$, depending on parametric and physical variables $\boldsymbol{y} \in \mathcal{U}$ and $\boldsymbol{x} \in \Omega$, respectively, that arises as the solution of a DE system

$$\mathcal{D}_{\boldsymbol{x}}(u, \boldsymbol{y}) = 0. \tag{1.2.1}$$

Here $\mathcal{D}_x(\cdot, \boldsymbol{y})$ is a differential operator in $\boldsymbol{x}$ that depends on the parameters $\boldsymbol{y}$, that also encodes the relevant boundary or initial conditions for the problem. Hereafter, we use the notation $u = u(\boldsymbol{y})$ instead of $f = f(\boldsymbol{y})$ when refering specifically to the case where the function to be approximated arises a the solution of a DE.

Common examples include parametric diffusion problems, natural convection in porous media, Boussinesq approximations, Navier-Stokes equations, and various other generally nonlinear and coupled systems. Understanding these problems is relevant, for example, in fluid mechanics, where the viscosity of the fluid may be unknown [222,240] and its parametric dependence may be caused by an undetermined contaminant, an error in measurements or external factors. Below we provide two of the three motivating examples of parametric PDEs considered in this thesis.

Parametric elliptic diffusion equation

A classic example is the parametric elliptic diffusion equation: Given $\boldsymbol{y} \in \mathcal{U}$, find u such that

$$-\text{div}(a(\boldsymbol{x}, \boldsymbol{y})\nabla u(\boldsymbol{x}, \boldsymbol{y})) = F(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega, \qquad u(\boldsymbol{x}, \boldsymbol{y}) = 0, \quad \text{on} \quad \partial\Omega, \tag{1.2.2}$$

where the diffusion coefficient $a(\boldsymbol{x}, \boldsymbol{y})$ is parametric, and the forcing term $F(\boldsymbol{x})$ is non-parametric. In electrostatics, it describes the potential field generated by a given charge distribution [153]. It is also crucial in modeling the conduction of heat in solids [221] and has various applications in fluid dynamics [33, 109, 269].

Parametric Navier-Stokes-Brinkmann equations

Another example that will be considered in this thesis is the parametric nonlinear stationary Navier-Stokes-Brinkmann (NSB) equations with random viscosity. Consider a bounded and Lipschitz physical domain $\Omega \subseteq \mathbb{R}^2$. Given $\boldsymbol{y} \in \mathcal{U}$, the modelling of a viscous fluid in a porous medium within Ω can be described by the incompressible NSB equations with random viscosity: find $\boldsymbol{u} : \Omega \times \mathcal{U} \to \mathbb{R}^2$ and $p : \Omega \times \mathcal{U} \to \mathbb{R}$ such that

$$\begin{aligned} \eta\boldsymbol{u}(\boldsymbol{y}) - \lambda\mathbf{div}(a(\boldsymbol{y})\boldsymbol{e}(\boldsymbol{u}(\boldsymbol{y}))) + (\boldsymbol{u}(\boldsymbol{y}) \cdot \nabla)\boldsymbol{u}(\boldsymbol{y}) + \nabla p(\boldsymbol{y}) &= f, \quad \text{in } \Omega \\ \text{div}(\boldsymbol{u}(\boldsymbol{y})) &= 0, \quad \text{in } \Omega \\ \boldsymbol{u} &= \begin{cases} \boldsymbol{u}_D, & \text{on } \partial\Omega_{\text{in}} \\ 0, & \text{on } \partial\Omega_{\text{wall}} \end{cases} \\ (a\nabla\boldsymbol{e}(\boldsymbol{u}) - p\mathbb{I})\nu &= 0, \quad \text{on } \partial\Omega_{\text{out}} \\ \int_\Omega p &= 0, \end{aligned} \tag{1.2.3}$$

where $\lambda = \mathrm{Re}^{-1}$, where Re is the Reynolds number, $a : \Omega \times \mathcal{U} \to \mathbb{R}_+$ is the random viscosity of the fluid, $\eta \in \mathbb{R}_+$ is the scaled inverse permeability of the porous media, $\boldsymbol{u}$ is the velocity of the fluid, $\boldsymbol{e}(\boldsymbol{u}) = \frac{1}{2}(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^t)$ is the symmetric part of the gradient and p is the pressure of the fluid. Moreover, $f : \Omega \to \mathbb{R}$ is a forcing term that is independent of the parameters. Here, we consider a nonzero inflow on the inlet $\partial\Omega_{\text{out}}$, a zero normal Cauchy stress on the outlet $\partial\Omega_{\text{out}}$ and a no-slip condition on the walls $\partial\Omega \setminus \{\partial\Omega_{\text{out}} \cup \partial\Omega_{\text{in}}\}$. Note that, to simplify the notation, we have dropped the $\boldsymbol{x}$ term from (1.2.3).

The deterministic version of this problem is taken from [115, §5.2] and was introduced in [256]. The study of the effect of a random viscosity on Navier-Stokes type of equations is important to determine measurement errors or uncertainties in porous media [177]. In particular, if the viscosity is not known precisely, introducing it as a random variable in the NSB equations can be used to characterize such uncertainties statistically instead of relying on measurements which may not capture the variability of the viscosity of the fluid through porous media.

Returning to the general problem (1.2.1), we assume that the (numerical) PDE solver is a *black box* (see also §1.4) that generates the m sample values (1.1.1) of $f(\boldsymbol{y}) = u(\cdot, \boldsymbol{y})$. The construction of approximations to the parametric map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$ without modifying the numerical PDE solver, as mentioned on §1.1, is known as a nonintrusive approach. In contrast, this thesis does not aim to cover the so-called *intrusive methods*, where the specific DE is an essential part of the construction of the approximation. See [96, 122, 239, 246] for more details on intrusive methods.

1.3 Challenges

Towards the end of 1957, when Richard Bellman published his book *Dynamic Programming* [38], the mathematical community was already starting to understand the computational limitations when dealing with high-dimensional problems. Many problems in UQ, such as those stemming from (1.2.1), require us to construct approximations to functions perform approximations of functions depending on many (or even infinitely-many) parameters that describe even more complex phenomena. For example, it is common to use Karhunen-Loève (KL) expansions to model diffusion coefficients as in (1.2.2). See, e.g., [176, §2.1] further discussion and examples of modeling random fields with KL expansions. Most of the time, it is either impossible or prohibitely expensive to exactly evaluate the underlying function or relevant quantity of interest. Thus, techniques for computing accurate approximations of such functions have become indispensable for modern CSE.

With this in mind, we now describe the main challenges this thesis aims to tackle.

The challenge of high dimensions

As mentioned in [12], Bellman coined the term *curse of dimensionality* (see also [38,39,63]) to describe computational situations that appear when certain approximations that work well in lower dimensions start to show undesirable results in higher dimensions. It can also impact the problem in different ways. For some sampling-based approaches (e.g., some stochastic collocation rules) it can require a number of points depending exponentially on the dimension [209, Rmk. 3.12]. This could lead to undesirable algorithm performance or results without meaningful representation of the physical phenomenon.

Hence, the foremost challenge is developing efficient efficient methods for approximating computational models in UQ that maintain their performance as the dimension of the model increases, thereby defeating the curse of dimensionality.

The challenge of Banach-valued functions

Obtaining the sample values in (1.1.1) is not a straightforward task, since the output space $\mathcal{V}$ is typically an infinite-dimensional function space, and therefore requires discretization. We typically know relevant details of the algorithms used for computing such discretizations (e.g., via a Finite Difference (FD), Finite Element (FE) or spectral method), such as convergence rates with respect to the number of degrees of freedom used in the discretization. However, this process always incurs an error, and therefore it is vital that the subsequent approximation method is robust, and has quantifiable bounds with respect to this error. In particular, as an input of the process to simulate, we have a parameter $\boldsymbol{y} \in \mathcal{U}$, and the output $f(\boldsymbol{y}) \in \mathcal{V}$, where $\mathcal{V}$ is a function space, e.g., Lebesgue and Sobolev spaces. Most current works consider models where the output is a scalar- or Hilbert-valued function.

While standard PDE problems, such as (1.2.2), are naturally formulated in weak form in Hilbert spaces, the efficient numerical solution of more complex PDEs increasingly involves weak formulations in Banach spaces [52,81,110,116,149]. As an example, a primal formulation (in the sense of FEM) of the steady-state Navier-Stokes equations (NSE) usually gives a solution belonging to a Hilbert space [240, §3]. However, a mixed formulation of the NSEs could be posed in Banach spaces as in [149]. More generally, problems of Banach-valued function approximation arise naturally in the context of UQ within the framework of *parametric operator equations* [98,99,225,234]. In addition, when providing recovery guarantees and approximation rates for smooth function approximation, certain results are easily extendible from the scalar- to the Hilbert-valued setting, due to the presence of an inner product. However, extending results from the Hilbert-valued setting to the Banach-valued case is more delicate. The absence of an inner product heavily impacts the analysis of such methods.

Specifically, this thesis investigates the approximation properties, convergence and limits of approximability of methods of computation depending on samples for smooth function

approximation in Hilbert and Banach spaces. As mentioned above, our focus is mainly on parametric solution maps of PDEs where different types of weak formulations may lead to different solution maps $\boldsymbol{y} \mapsto f(\cdot, \boldsymbol{y}) \in \mathcal{V}$.

The challenge of obtaining samples

With modern technological advances, data may seem unlimited, available and free of errors. For instance, the classification of images in social media benefits from a large source of high-quality training samples from users. This is far from what happens in practice in CSE. Here, many problems are data-starved or work with scarce data since most of them rely on large, costly simulations. For instance, in practical parametric DEs applications, generating sample values of the form (1.1.1) involves using computationally expensive solvers or acquiring expensive real-world data. Furthermore, to get better approximations to the underlying physical process, CSE requires ever more complex mathematical models, with more features to analyze (more parameters). As mentioned in [12, §1.1], there is consequently a need for algorithms whose sample complexity does not scale poorly as the parametric dimension increases.

This also gives rise to another challenge: namely, the measurements of $f(\boldsymbol{y})$ are always inexact. Here different sources of error must be considered, such as random noise, numerical errors, and calibration errors in physical devices, to name a few. Hence, the last challenge is developing methods so that corruptions in the samples do not lead to a drastic deterioration in the resulting approximations. We refer §2.7 for further details on these errors.

Identifying the challenges

In order to give a clear understanding of how these main challenges in UQ affect the main results in the thesis, we divide them into four key challenges. First, rather than considering sample values as in (1.1.1), we now consider m noisy *sample values* given by

$$d_i = f(\boldsymbol{y}_i) + n_i \in \mathcal{V}, \quad i = 1, \dots, m, \tag{1.3.1}$$

where n_i is a measurement error term in the ith measurement.

We now state the four key challenges that motivate this work:

(C-i) The domain $\mathcal{U} \subseteq \mathbb{R}^d$ of f high- or infinite-dimensional, e.g., typical parametric DEs depend on many parameters.

(C-ii) Generating data is expensive, e.g., acquiring each sample d_i may involve running a numerical PDE solver.

(C-iii) The data d_i is corrupted by errors, e.g., numerical errors resulting from the PDE solver.

(C-iv) The function f takes values in an infinite-dimensional function space $\mathcal{V}$, typically a Hilbert or Banach space.

1.4 Key considerations in this work

To frame our main contributions, we need several further fundamental considerations.

Lebesgue-Bochner spaces

In general, this thesis considers a function $f \in \mathcal{X}$, where in most cases $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ is a Lebesgue-Bochner space. Typically, we consider $\mathcal{X}$ as the Bochner space $L^2_\varrho(\mathcal{U};\mathcal{V})$ (see Chapter 2 for the definition). The parameter domain $\mathcal{U}$ is a subset of $\mathbb{R}^d$ or $\mathbb{R}^{\mathbb{N}}$, and the output space is a Banach or Hilbert space. As we describe next, the parameters $(y_k)_{k=1}^d$ are usually assumed to be independent and vary in bounded intervals $\mathcal{U}_k \subset \mathbb{R}$ (which, up to rescaling, can be taken to be equal to $[-1,1]$). Thus $\boldsymbol{y} = (y_k)_{k=1}^d$ belongs to a hyperrectangle $\mathcal{U} = \prod_{k=1}^d \mathcal{U}_k$. In particular, we focus on the symmetric hypercube $\mathcal{U} = [-1,1]^d$ of side length 2 with $d \gg 1$ or $d = \infty$.

The sampling strategy

This thesis considers sample points $\boldsymbol{y}$ drawn from a probability measure ϱ on $\mathcal{U}$ [12]. Specifically, we obtain the samples using Monte Carlo sampling (MCS), which draws independent and identically distributed random samples $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in \mathcal{U}$. This is very common in practice, particularly in high-dimensional UQ settings. As mentioned in [122, §3.1], by using MCS methods one can obtain error decay that scale mildly with (or are independent of) d, which is a clear advantage in view of (C-i). However, in practice, generating enough data and sample values $f(\boldsymbol{y})$ to do a useful study may take seconds or minutes, to days or even weeks. Thus, challenges (C-ii) and (C-iii) remain present.

Remark 1.4.1 As a further consideration, we note that in many scenarios one may have substantial flexibility to choose the sample points $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in \mathcal{U}$ in (1.1.1). However, in other scenarios they may be fixed, e.g., when dealing with legacy data or when we have no control over the sampling procedure. As mentioned above, we consider Monte Carlo sampling – which may be considered either as a chosen sampling strategy or a fixed one, depending on the setting.

A black box model

Given an input $\boldsymbol{y}_i$, we assume that we have access to a black box model (or solver in a PDE setting) producing an output that approximates $f(\boldsymbol{y}_i)$. These approximations must belong to a finite-dimensional space and be characterized by a finite set of numbers. Specifically, the sample values d_i in (1.3.1) are assumed to be elements of a finite-dimensional space $\mathcal{V}_K \subseteq \mathcal{V}$ of dimension $K \in \mathbb{N}$. For example, in a PDE setting $\mathcal{V}_K$ may be chosen as a FE space. Here, we do not specify precisely how such an approximation is performed, nor how large an error this results in. In other words, we consider the computation that evaluates f at $\boldsymbol{y}_i$ as a black box. A particular case of interest is when the d_i are the orthogonal projections of the exact sample values $f(\boldsymbol{y}_i)$. However, we do not assume this in what follows since, in practice, the numerical computation that yields d_i may not involve computing a projection. Our objective is to develop approximations for which the error scales linearly in $\|\boldsymbol{n}\|_{2;\mathcal{V}}$, the norm of the noise $(n_i)_{i=1}^m$ defined in (1.3.1), thus accounting for any black box mechanism for computing the samples.

Known versus unknown anisotropy

In general, high-dimensional approximation is only possible under some type of *anisotropy* assumption, i.e., the function in question depends more strongly on certain variables than others. In some cases the nature of this anisotropy – i.e., the order of importance of the variables and the relative strengths of the interactions between them – may be known in advance. However, in more realistic scenarios, it is unknown. On the one hand, in the known anisotropy setting, a reconstruction procedure can use this information to achieve a good approximation uniformly over a certain class of smooth functions. On the other hand, in the unknown anisotropy case, the reconstruction procedure has no access to this information. In this work, we consider both the *known* and *unknown anisotropy* settings.

Uniform versus nonuniform recovery

We now describe uniform and nonuniform recovery, which are subtely different concepts. The latter describes the situation where a single draw of the points $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ is sufficient for recovery of a fixed function f with high probability up to the specified error bound. Conversely, *uniform* recovery concerns the situation where a single draw of the sample points $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ is sufficient for recovery of any function with high probability up to the specified error bound. The reason for this difference in our theoretical results stems from bounding a certain *discrete* error term (see, e.g., the discussion in Remark 3.3.4), which is a random variable depending on f and the sample points.

1.5 Methods considered in this thesis

This thesis focuses on the use of polynomial and Deep Learning (DL)-based methods for smooth function approximation. Polynomials lie at the heart of many well-established approximation schemes. On the other hand, machine learning techniques based on Deep Neural Network (DNNs) have achieved some impressive results in inverse problems in imaging, PDEs, protein structure prediction, and biological applications, to name a few [87, 118, 158, 163, 172, 235].

1.5.1 Polynomial approximation

Polynomial-based methods are a convenient tool for deriving explicit convergence rates for function approximation in high- or infinite-dimensions. These convergence rates are fast when assuming that the function f is smooth with respect to its variables. For instance, best polynomial approximation rates can be as fast as exponential in the number of terms in the polynomial expansion [12, §1.3]. In this thesis, we consider the best polynomial approximations of holomorphic functions. Such functions are relevant for the parametric DE problems considered in this thesis (see §1.2) but are not necessary for using polynomial-based methods.

1.5.2 Deep Learning

Mathematically, the first construction of a DNN appeared in [193]. In recent years, Deep Learning (DL) and DNNs have begun to have a significant impact on scientific computing and its many applications. Impressive results have been achieved by training large models with billions of parameters using vast datasets on large distributed computing resources. However, many still question their use in critical applications that require rigorous safety standards. As the process of training DNNs on real-world or synthetic data is increasingly considered for applications in medicine, science, and engineering, it is important to quantify the efficiency and reliability of DL from both theoretical and practical standpoints.

1.6 Key questions in this thesis

In this thesis, we focus on overcoming the challenges (C-i) to (C-iv) above through the development of stable, accurate and efficient approximations to holomorphic high- or infinite-dimensional Banach-valued functions, based on polynomials and DNNs. We also study fundamental limits of approximatibility for such functions, using the theory of m-widths, Gelfand widths and Kolmogorov widths.

To guide the reader and highlight the main contributions in each chapter, we pose nine fundamental questions. Our answers to these question will involve polynomial approxima-

tions based on compressed sensing and efficient restart schemes for first-order optimization solvers.

Question 1 *Are there methods for computing approximations to holomorphic, finite- or infinite-dimensional, Hilbert-valued functions from limited samples that achieve similar theoretical rates as benchmarks such as the best s-term polynomial approximation?*

Question 2 *Can these methods be implemented efficiently as algorithms, and what is their computational cost?*

In particular, keeping in mind challenge (C-i), from Question 2 we pose two more specific questions.

Question 3 *Is it possible to approximate a holomorphic Hilbert-valued function of infinitely-many variables with error decaying algebraically fast in m via an algorithm whose computational cost is subexponential in m?*

Question 4 *For finite-dimensional problems, i.e., $d < \infty$, is it possible to approximate a Hilbert-valued holomorphic function with error decaying exponentially fast in m via an algorithm whose computational cost is polynomial in m?*

Note that Questions 1–4 will constitute a substantial part of this thesis. Specifically Chapters 3–4 deal with algorithms computing near-best polynomial approximations in Hilbert spaces. We also broaden our focus to considering Banach-valued functions in Chapters 5–7, where we turn our atention to an an alternative approach. The answer to the four following questions involve DL and more specifically DNNs, key extensions of compressed sensing for Banach-valued functions and emulation of polynomials via DNNs.

Question 5 *In the known or unknown anisotropy case, is it possible to learn smooth high or infinite-dimensional Hilbert- or Banach-valued approximations from a limited dataset using DNNs with a complete theoretical understanding of the sample complexity and approximation rates?*

Question 6 *Is it possible to efficiently apply DL to learn smooth parametric functions with architectures commonly used in practice to approximate Banach-valued solutions of challenging DEs from limited samples, and does it achieve close to the theoretical rates espoused by Question 5?*

As mentioned in §1.3, there is a growing interest in the study of Banach-valued solutions of parametric PDEs in the context of parametric PDEs. Consequently, it is crucial to identify and extend key results from smooth function approximation from the Hilbert- to the Banach-valued case. A key question involves investigating whether, in practice, approximating more general Banach-valued functions results in a worse decay rate compared to Hilbert-valued approximation.

Question 7 *Is there any preliminary empirical evidence indicating that learning a Banach-valued solution of a parametric DE exhibits a decay rate in terms of the number of samples that is worse than in the Hilbert-valued case?*

Next, the answer to the following question involves optimal learning and a natural question that arises when deriving the approximation rates in the previous context. More precisely, we seek to determine the underlying limits of approximability from limited samples for smooth functions.

Question 8 *Are the approximation rates derived in answering Questions 1–5 optimal?*

Our final inquiry is related to the challenges (C-iii) and (C-iv) from §1.3. In this work, it is important to understand the impacts of different types of errors in the approximations. Various factors, including numerical errors, physical discretization errors, sampling errors and optimization errors (commited when certain minimization problems are not solved exactly) can influence the accuracy of the approximations.

Question 9 *How do various types of errors, such as numerical errors, physical discretization errors, sampling errors and optimization errors, impact the accuracy of the approximations constructed in answering Questions 1–5?*

1.7 Outline and contributions of this thesis

We now briefly outline the remainder of this thesis and describe the contribution of each chapter in relation to the key questions above.

Chapter 2 is a crucial starting point, providing a comprehensive presentation of the general notation used in this thesis. It also introduces significant function spaces, relevant examples, and additional considerations. This chapter formally defines the concepts of smoothness, holomorphy, and parametric dependence, along with the best s-term polynomial approximation and DNNs. These concepts lay the foundation for the analysis in the subsequent chapters.

Chapter 3 is dedicated to addressing Questions 1 and 9 in the context of polynomial approximation. It focuses on the methods used for computing polynomial approximations to Hilbert-valued functions. We present three main theorems that deal with finite and infinite-dimensional cases, focusing on unknown anisotropy.

The methods developed in Chapter 3 are not algorithms, as they rely on exact minimizers of certain convex optimization problems. In Chapter 4, we address this issue, by developing efficient algorithms that approximately solve these problems. In particular, we show that these algorithms achieve the desired approximation rates, thereby answering Questions 2–4, and remain robust to the other sources of error, thus answering Question and 9.

In Chapter 5 we switch focus and consider DL-based methods. We also extend from Hilbert-valued to Banach-valued functions. Our goal is to answer Questions 5 and 9, and to derive similar rates to those shown in the previous chapters. The key idea in this chapter involves emulating polynomial approximation methods, specifically, those based on least squares and compressed sensing, as DNN training problems.

In Chapter 6, we introduce new concepts based on optimal learning. Our aim is to determine whether or not the algebraic approximation rates obtained by the methods developed in Chapters 3-5 are optimal. Specifically, this chapter aims to answer Question 8 by using the theory of (adaptive) m-width, and by reducing the problem to one of determining lower bounds on the Gelfand and Kolmogorov widths of certain weighted unit balls in finite dimensions.

Chapter 7 presents three numerical experiments that address Questions 6–7. The experiments consist of three parametric PDEs: an elliptic diffusion equation, a Navier-Stokes-Brinkman equation and a Boussinesq equation. We provide a detailed methodology to ensure reproducibility, describe each problem and its main challenges, and analyze the empirical results, highlighting their significance and relevance to the thesis.

The thesis also includes several appendices that supplement the main content. In Appendix A, we utilize the arguments presented in [12] to demonstrate that, by employing a mixed formulation (in the sense of FE), the parametric solution map of the Poisson problem is a mapping for which the theory developed in the previous chapters applies. Appendix B presents several key lemmas necessary to achieve the desired error bounds discussed in Chapter 6.

1.8 Literature review

In this section, we describe how our works relate to several existing areas of research and previous works. We commence with a general overview of works relating parametric DEs and smooth function approximation.

A substantial body of literature has demonstrated that solution maps for various parametric DEs are *holomorphic* functions of their parameters [27,56,63,71,77,122,144,157,258]. To mention a few examples: elliptic PDEs with both affine and nonaffine parametric dependencies, parabolic PDEs, PDEs defined over parametrized domains dealing with shape uncertainty, parametric Initial Value Problems (IVPs), parametric hyperbolic problems, and parametric control problems. More classical results in this area can be found, for instance, in [265] and related references. For surveys on more recent findings, we direct readers to [71], [12], and related references.

Alongside efforts to establish holomorphic regularity of parametric DEs, there has been a focus on applying polynomial approximation, particularly *best s-term polynomial approximation*, to construct finite approximations to such functions. As mentioned in §1.5.1, best

s-term approximation involves approximating the function f using an s-term expansion that corresponds to its largest s coefficients (measured in a suitable norm) with respect to a polynomial basis. Common choices for the basis include Taylor polynomials, tensor-product Legendre and Chebyshev polynomials on bounded hypercubes, or tensor-product Hermite and Laguerre polynomials on $\mathbb{R}^d$ or $[0,\infty)^d$. See [234] for a detailed overview.

In this thesis we focus on nonintrusive surrogate model constructions where various methods have been developed. A partial list includes sparse polynomial approximation [36, 43, 63, 64, 71, 72, 136, 137], Gaussian process regression (or kriging) [239, 246], radial basis functions [159, 239], reduced-basis methods [143, 223], and most recently DNNs and DL [7, 16, 82, 84, 87, 141, 180, 214, 215, 236].

1.8.1 Chapter 3

The systematic study of best s-term polynomial approximation of high- or infinite-dimensional holomorphic functions began around 2010 with the works of [43, 72, 73, 136, 255]. For reviews, see [71] and [12, Chpt. 3]. Note that many of these works assume that the function is a solution of a parametric PDE, and therefore first demonstrate that such a function is holomorphic. However, other works avoid this step and use specific properties of the DE to obtain refined estimates. See, e.g., [29, 30] for results of this type. Other recent works such as [47] also study the problem without assuming the function is a solution of a parametric PDE.

The study of least-squares method for constructing such approximations from sample points began in the early 2010s [61, 70, 195, 200]. Many works have pursued various extensions, such as enhanced sampling strategies [124, 199, 204, 237, 251, 277, 278], near-optimal sampling strategies [13, 74, 132], optimal sampling strategies [32, 101, 102, 160, 182, 253], methods for general domains [20, 100, 198], optimal and adaptive methods [76, 196, 197] and multilevel strategies [131]. See [75, 126, 130] and [12, Chpt. 5] for reviews.

Compressed sensing was introduced in the context of image and signal processing by modelling image and signals as sparse vectors [19, 54, 103, 112]. Its use in polynomial approximation started early in the last decade with the works of [45, 104, 191, 226, 271]. This has also led to substantial research. See [94, 95, 104, 191, 225, 273] and references therein for applications to parametric PDEs. Various extensions include refined sampling strategies [23, 97, 127, 133, 155, 183, 250], iterative methods and basis selection techniques [22, 135, 262, 274–276], nonconvex optimization methods [123, 257, 270, 272], sublinear-time algorithms [65, 66], gradient-enhaced minimization techniques [21, 125, 154, 218, 245, 249], methods for dealing with corrupted samples [3, 4, 145, 238] and multilevel and multifidelity strategies [48, 208]. For additional information and reviews, see [134, 166, 190, 205] and [12, Chpt. 7].

Here, weighted ℓ^1-minimization plays an important role and has been developed in works such as [1–3, 12, 64, 217, 227, 273] and [12, Chpts. 6-7]. Alongside with the notions of

lower and *anchored* sets (see §2.4.4). These have been extensively studied in the best s-term polynomial approximation literature. Compressed sensing techniques aiming to exploit such structures were first considered in [2, 3, 64] and [12, Chpt. 7]. Moreover, the extension of classical compressed sensing theory from vectors in $\mathbb{R}^N$ (or $\mathbb{C}^N$) to Hilbert-valued vectors in $\mathcal{V}^N$ was first developed in works such as [95, 234]. In Chapter 3 in order to prove our main results, we also extend this framework to the weighted setting and Banach-valued case.

1.8.2 Chapter 4

Primal-dual methods have, over the last decade, been established as efficient methods for solving (convex) optimization problems, especially for image reconstruction [279]. In [207] the authors show convergence of the ergodic sequence of the form $\mathcal{O}(1/t)$ for t number of steps. In [57] the authors show a over-relaxation step with a similar rate to [207]. See [58, 59] for more on the primal-dual iteration and [228–230] for the general notion of restarts in continuous optimization. We use primal-dual methods to solve (weighted) ℓ^1-minimization problems in compressed sensing. Note that there are also various non-optimization based techniques in the compressed sensing literature (see, e.g., [112]), including iterative threshold and greedy methods (the latter are closely related to the adaptive least-squares methods discussed earlier [12, §6.2.5]). However, these do not currently possess theoretical guarantees in the weighted setting.

There have been several previous attempts to connect compressed sensing theory for analyzing the sample complexity of polynomial approximations via (weighted) ℓ^1-minimization and best s-term polynomial approximation theory. In [225], the authors consider approximating scalar quantities of interest of solutions to affine parametric operator equations in Banach spaces. Assuming a certain weighted summability criterion, they first show holomorphy of the parametric solution map and then use a weighted ℓ^1-minimization procedure in combination with Chebyshev polynomials to derive algebraic rates of convergence in the L^2_ϱ-norm of the form $\mathcal{O}\left((m/\text{polylog}(\text{m}))^{1/2-1/p}\right)$. The work in Chapter 4 is more general, since its starting point is a holomorphic function, not a solution of a parametric operator equation. We also consider Hilbert-valued functions, i.e., the whole solution map, not a scalar quantity of interest of it. Moreover, the work of [225] is based on exact minimizers of certain constrained, weighted ℓ^1-minimization problems, whereas we construct full algorithms. Recently, at the same time as writing the theory for this chapter, some similar results were presented in the book [12]. However, these only consider the scalar-valued case and do not address algorithms, which is the main focus of Chapter 4.

1.8.3 Chapter 5

In CSE, there is increasing empirical evidence that DL is a promising tool for solving challenging problems, such as UQ problems where the underlying model is described in terms

of parametric DEs [7, 37, 42, 68, 85, 90, 118, 121, 138, 162, 168, 178, 181]. In addition, theoretical results on DNNs overcoming the curse of dimensionality (in some suitable sense) have emerged [202], but they do not address how the training procedure learns from limited samples. Obtaining these samples is often expensive, involving physical or numerical experiments. For example, in uncertainty quantification problems where the underlying model is described by the solution of DEs, costly numerical solvers are typically used to obtain these sample evaluations. Therefore, understanding the sample complexity of training DNNs is crucial.

As we explain later, Chapter 5's main results are established by emulating polynomials using DNNs. Early approaches on approximation methods for smooth functions involving multivariate polynomials include interpolation schemes using sparse grids (see [62, 71, 75] and [12, Chpt. 1] and references therein). As discussed in [12, §1.7], these methods are best suited to the known anisotropy setting, since they generally require a priori knowledge of a good polynomial subspace in which to construct the approximation. They also do not generally obtain optimal rates in terms of m, due to growing Lebesgue constants [62, 75]. More recently, there has been significant focus on least-squares methods [61, 70, 200] and methods based on compressed sensing [104, 191, 226]. See [12, Chpts. 5 & 7] for reviews. The former are suitable for the known anisotropy case since they require knowledge of a good polynomial subspace. In contrast, the latter can handle the unknown anisotropy setting.

Using polynomial techniques to establish theoretical guarantees for DNN training from limited samples was previously considered in [7, 16]. These works consider either scalar- or Hilbert-valued functions in finite dimensions and approximation via ReLU DNNs. Chapter 5 can be considered a theoretical extension to the (significantly more challenging) infinite-dimensional and Banach-valued setting, using other families of DNNs. As noted above, emulation of DNNs by polynomials is a well-established technique in DNN approximation theory. In this chapter, we use ideas from [91, 180, 215, 235] to establish our main results. For more on polynomial-based methods for high-dimensional approximation, see [12, 64, 71] and references therein. See also [172] for a different approach based on reduced bases to derive DNN approximation results for parametric PDEs.

Another line of recent DL research involves learning operators between function spaces [42, 53, 174, 181, 189, 206, 267]. This is motivated in great part by parametric PDEs, where the operator is, for example in the case of (1.2.2), the mapping from the diffusion coefficient to the PDE solution $a \in \mathrm{L}^\infty(\Omega) \to u = u(\cdot, a) \in \mathrm{H}^1_0(\Omega)$. Chapter 5 is related to this line of investigation in that we assume a parametrization of a in terms of an infinite vector $\boldsymbol{y} \in [-1, 1]^{\mathbb{N}}$ for which the map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$ is holomorphic. However, it is also different in scope, as we consider approximating an arbitrary Banach-valued, holomorphic function $f : \mathcal{U} \to \mathcal{V}$ which may or may not arise as the solution map of a parametric DE. We note also that many of the aforementioned works assume a Hilbert space formulation, whereas we consider Banach spaces.

1.8.4 Chapter 6

When applied to infinite-dimensional, holomorphic, Hilbert-valued functions, the methods developed in Chapters 3-5 of this thesis exhibit algebraic rates of convergence of the form $\mathcal{O}\left((m/\text{polylog}(m))^{1/2-1/p}\right)$ when applied to the class of Hilbert-valued smooth functions. Here, m represents the number of pointwise samples, typically Monte Carlo samples, and $0 < p < 1$ is a smoothness parameter. However, to the best of our knowledge, few works have sought to determine the underlying limits of approximability from samples for such function classes.

The study of Kolmogorov widths began around 1936, as mentioned in [187] and [165]. Since then, it has been widely used to study how well the worst element of a space can be approximated. Other works, such as [220] and references therein, have contributed to the essential theory of widths. Recently, works like [28, 44] have provided a detailed mathematical description of the concept of optimal learning in the context of PDEs, and [101] has addressed the approximation of functions based on pointwise data.

Gelfand widths, as discussed in [111], have become key to understanding the limits of approximation and performance bounds for sparse recovery methods. For methods using compressed sensing and results on m-widths providing matching upper and lower estimates for certain classes of functions in ℓ^p balls, we refer to [69] and references therein.

It is worth noting that while [28] considers Kolmogorov widths for Hilbert-valued functions $u(\boldsymbol{y})$ arising as solutions to certain parametric elliptic PDEs, it does not address the question of finite samples.

Generally speaking, our work in Chapter 6 is related to recent advances [44, 253] in optimal recovery [93, 194]. We use concepts and ideas from information-based complexity [210–212, 260], in particular m-widths, to understand the aforementioned limits of approximability from samples for smooth functions. More specifically, we use lower estimates for the Gelfand widths, Kolmogorov widths and in general m-widths theory (see, e.g., [220]).

1.8.5 Chapter 7

Note that the majority of the content in this thesis is theoretical. However, Chapter 7 focuses on the practical implementation of DL for approximating complex parametric PDEs. As previously mentioned, recent advances have demonstrated that DNNs can efficiently learn solution maps of parametric PDEs, significantly improving computational speed. Yet, a gap remains between theory and practice [16]. After reviewing recent developments, highlighting significant works, methodological advancements, and emerging trends in the application of DNNs to parametric PDEs, we now turn our attention to the specific PDEs addressed in Chapter 7.

The Poisson problem

The Poisson problem, described by the equation $-\text{div}(\nabla u) = f$, has numerous significant applications in science and engineering. As mentioned in §1.2 it is applied in electrostatics [153], heat transfer in solids [221] and fluid dynamics [33, 109, 269].

When using FEM to solve the Poisson problem, some variables are obtained by direct differentiation (e.g., the gradient of u), which can result in a loss of accuracy [173]. In many practical applications, it is important to obtain accurate approximations of the gradient or other variables of interest. To address this, various methods have been employed mixed formultions, or other methods to solve the Poisson problem. For detailed discussions on these methods, we refer to [71, 108, 113, 114, 219, 233] and the references therein.

Many recent works have shown that DL is effective at learning the solution of the parametric Poisson equation. There are numerous studies on this topic, including [42, 83, 85, 139, 167, 172], to name a few.

Navier-Stokes-Brinkman equations

The Navier-Stokes-Brinkman (NSB) equations model a wide variety of viscous fluids through porous media, phase change models, and fluids in complex geometries [150, 161, 188]. In some cases, the velocity flow or the geometry becomes too complex to model via Darcy's or Navier-Stokes equations, and the incorporation of the Brinkman model provides good accuracy in a variety of contexts [152, §1.1]. For instance, NSB equations appear in [266] where the authors present a 2D numerical model for natural convection in a square cavity for a melting process. For a review of phase change models, we refer to [105]. A model closer to the one used in this thesis appears in the prediction of a filter and absorption of contaminants in water purification problems [201]. Several works have employed different methods to numerically solve similar Stokes, Navier-Stokes, and Brinkman flows. See, e.g., [31, 46, 115, 117, 129, 152] and references therein.

Boussinesq approximation

The analysis of an incompressible viscous fluid governed by the Navier-Stokes equations coupled with a nonlinear heat equation is considered in [24] as the Boussinesq approximation for a constant viscosity. An early derivation of the Boussinesq approximation in a non-stationary combustion theory framework can be found in [192]. Years later, in [41] the authors prove the existence and uniqueness of an analogous non-parametric Boussinesq approximation. For further details we refer to [55, 128, 175, 185]. The nonlinear coupled problem is considered through a temperature-dependent viscosity and a buoyancy term approximation given in [86, 216], the latter is based on the theoretical derivation in [184]. A primal formulation (in the sense of FEMs) for the viscosity and thermal conductivity of the

fluid, which depend on the temperature, are discussed in [81]. The weak formulation of the Boussinesq problem in Chapter 7 is based on the fully-mixed formulation in [116] taking ideas from [81].

Chapter 2

General preliminaries

In this chapter, we describe the main notation, definitions and technical details required throughout this thesis.

2.1 Notation

We now introduce the main notation used in this thesis. We denote $\mathbb{N}$ and $\mathbb{N}_0$ as the sets of positive and nonnegative integers, respectively, and $\mathbb{R}_+$ for the set of positive numbers. As usual for $d \in \mathbb{N}$, we write $[d]$ to denote $\{1, \dots, d\}$. In the case where $d = \infty$, we write $[d] = \mathbb{N}$ as the set of positive integers. When $d \in \mathbb{N}$, we write $\mathbb{K}^d$ for the scalar vector space (real space when $\mathbb{K} = \mathbb{R}$ or complex when $\mathbb{K} = \mathbb{C}$) with d components, and when $d = \infty$ we write $\mathbb{K}^{\mathbb{N}}$ for the vector spaces of real- or complex- valued sequence indexed over $\mathbb{N}$. Similarly, we write $\mathbb{N}_0^d$ or $\mathbb{N}_0^{\mathbb{N}}$ as the set of nonnegative multi-indices of length $d \in \mathbb{N} \cup \{\infty\}$. Let $d \in \mathbb{N}$. We define the multi-index set $\mathcal{F}$ as the set of nonnegative multi-indices, i.e.,

$$\mathcal{F} := \mathbb{N}_0^d = \{\boldsymbol{\nu} = (\nu_k)_{k=1}^d : \nu_k \in \mathbb{N}_0\}, \quad d < \infty. \tag{2.1.1}$$

In the infinite-dimensional case we define $\mathcal{F}$ as the set of multi-indices in $\mathbb{N}_0^{\mathbb{N}}$ with at most finitely-many nonzero terms. That is

$$\mathcal{F} := \{\boldsymbol{\nu} = (\nu_k)_{k=1}^\infty \in \mathbb{N}_0^{\mathbb{N}} : |\{k : \nu_k \neq 0\}| < \infty\}. \tag{2.1.2}$$

We also write $\boldsymbol{e}_j = (\delta_{j,k})_{k \in [d]}$ for the standard basis vectors, where $j \in \mathbb{N}$ or $j \in [d]$. We write $\boldsymbol{w} \odot \boldsymbol{v} = (w_i v_i)_{i \in [d]}$ for the Hadamard product and $\boldsymbol{w} \otimes \boldsymbol{v} = (w_i v_j)_{i,j \in [d]}$ for the tensor product of vectors $\boldsymbol{w} = (w_i)_{i \in [d]}$ and $\boldsymbol{v} = (v_i)_{i \in [d]}$.

We write $\mathbf{0}$ and $\mathbf{1}$ for the multi-indices consisting of all zeros and all ones, respectively. The inequality $\boldsymbol{\mu} \leq \boldsymbol{\nu}$ is understood componentwise for any multi-indices $\boldsymbol{\mu} = (\mu_i)_{i \in [d]}$ and $\boldsymbol{\nu} = (\nu_i)_{i \in [d]}$, i.e., $\boldsymbol{\mu} \leq \boldsymbol{\nu}$ means that $\mu_k \leq \nu_k$ for all $k \in [d]$. Let $\boldsymbol{\nu} = (\nu_i)_{i \in [N]}$ and

$\boldsymbol{\mu} = (\mu_i)_{i\in[N]}$. Then we also write

$$\boldsymbol{\nu}^{\boldsymbol{\mu}} = \prod_{i\in[N]} \nu_i^{\mu_i}, \quad \text{and} \quad \boldsymbol{\nu}! = \prod_{i\in[N]} (\nu_i!),$$

with the convention that $0^0 = 1$.

Let $N \in \mathbb{N}$, $S \subseteq [N]$ and $\boldsymbol{x} \in \mathbb{C}^N$. We write $\boldsymbol{x}_S \in \mathbb{C}^N$ for the vector with ith entry equal to x_i if $i \in S$ and zero otherwise. We also write S^c for the complement $[N] \setminus S$ of S.

Given $0 < p \leq \infty$ and we write $\|\cdot\|_p$ for the usual vector ℓ^p-norm (if $p \geq 1$) or ℓ^p-quasinorm (if $0 < p \leq 1$). For $0 < p, q < \infty$ we define the matrix $\ell^{p,q}$-(quasi)norm of an $m \times n$ matrix $\boldsymbol{G} = (G_{i,j})_{i,j=1}^{m,n}$ as $\|\boldsymbol{G}\|_{p,q}^q := \sum_{j=1}^n (\sum_{i=1}^m |G_{i,j}|^p)^{q/p}$, and similarly when $p = \infty$ or $q = \infty$.

2.2 Function spaces

In this thesis, we employ a variety of function spaces. We now define the main spaces used.

2.2.1 Measures and parametric domain

First, we consider measures that arise as tensor products of probabilities measures supported on the interval $[-1, 1]$, and we write $\varrho^{(1)}$ for such a measure.

Here we focus mainly on two examples, the uniform and Chebyshev (arcsine) measures. These are defined by

$$\mathrm{d}\varrho^{(1)}(y) = 2^{-1}\,\mathrm{d}y, \quad \text{and} \quad \mathrm{d}\varrho^{(1)}(y) = \frac{1}{\pi\sqrt{1-y^2}}\,\mathrm{d}y, \quad y \in [-1, 1], \tag{2.2.1}$$

where $\mathrm{d}y$ is the Lebesgue measure. In finite dimensions, we let the parametric domain $\mathcal{U} = [-1, 1]^d$ be the symmetric d-dimensional hypercube of side length 2 and define the probability measure on $\mathcal{U}$ by tensoring the one-dimensional measure:

$$\varrho = \varrho^{(d)} := \varrho^{(1)} \otimes \ldots \otimes \varrho^{(1)}. \tag{2.2.2}$$

In particular, the d-dimensional uniform and Chebyshev measures are given by

$$\mathrm{d}\varrho(\boldsymbol{y}) = 2^{-d}\,\mathrm{d}\boldsymbol{y}, \quad \text{and} \quad \mathrm{d}\varrho(\boldsymbol{y}) = \prod_{k\in[d]} \frac{1}{\pi\sqrt{1-y_k^2}}\,\mathrm{d}\boldsymbol{y}, \quad \forall \boldsymbol{y} = (y_k)_{k=1}^d \in \mathcal{U}, \tag{2.2.3}$$

respectively. In infinite dimensions, we consider the infinite-dimensional symmetric hypercube $\mathcal{U} = [-1, 1]^{\mathbb{N}}$ of side length 2 and write $\boldsymbol{y} = (y_j)_{j\in\mathbb{N}} \in \mathcal{U}$ for the variable in this domain. For the uniform or Chebyshev measures, the Kolmogorov extension theorem guarantees the existence of a probability measure on $\mathcal{U}$ as the infinite tensor-product of their

one-dimensional measure (see, e.g., [252, §2.4]), which we denote as

$$\varrho = \varrho^{(\infty)} = \varrho^{(1)} \otimes \varrho^{(1)} \otimes \dots . \tag{2.2.4}$$

2.2.2 $\mathcal{V}$-valued sequence spaces

Throughout, we let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be a (separable) Banach space over $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$) and write $(\mathcal{V}^*, \|\cdot\|_{\mathcal{V}^*})$ for its continuous dual. We consider $v \in \mathcal{V}$ as an arbitrary element of $\mathcal{V}$ and $v^* \in \mathcal{V}^*$ as an arbitrary element of $\mathcal{V}^*$. We let $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ be the duality pairing between $\mathcal{V}$ and $\mathcal{V}^*$, $\langle v^*, v \rangle_{\mathcal{V}} = v^*(v)$. Note that

$$\|v\|_{\mathcal{V}} = \sup_{\substack{v^* \in \mathcal{V}^* \\ \|v^*\|_{\mathcal{V}^*} \leq 1}} |v^*(v)| = \max_{\substack{v^* \in \mathcal{V}^* \\ \|v^*\|_{\mathcal{V}^*} = 1}} |v^*(v)|, \quad \forall v \in \mathcal{V}, \tag{2.2.5}$$

(see, e.g., [50, Cor. 1.4]). We let $\mathcal{V}^N$ be the vector space of Banach-valued vectors of length N, i.e., $\boldsymbol{v} = (v_i)_{i=1}^N$ where $v_i \in \mathcal{V}$, $i = 1, \dots, N$. More generally, let $\Lambda \subseteq \mathbb{N}_0^d$ denote a (possibly infinite) multi-index set. We write $\boldsymbol{v} = (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ for a sequence with $\mathcal{V}$-valued entries, $v_{\boldsymbol{\nu}} \in \mathcal{V}$. In particular, when $(\mathcal{V}, \langle \cdot, \cdot \rangle_{\mathcal{V}})$ is a Hilbert space we consider $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ as the inner product with corresponding induced norm $\|\boldsymbol{v}\|_{\mathcal{V}}^2 := \langle \boldsymbol{v}, \boldsymbol{v} \rangle_{\mathcal{V}}$, where

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathcal{V}} = \sum_{\boldsymbol{\nu} \in \Lambda} \langle u_{\boldsymbol{\nu}}, v_{\boldsymbol{\nu}} \rangle_{\mathcal{V}}.$$

For $0 < p \leq \infty$, we define the $\ell^p(\Lambda; \mathcal{V})$ space as the set of those sequences $\boldsymbol{v} = (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ for which $\|\boldsymbol{v}\|_{p;\mathcal{V}} < \infty$, where

$$\|\boldsymbol{v}\|_{p;\mathcal{V}} := \begin{cases} \left(\sum_{\boldsymbol{\nu} \in \Lambda} \|v_{\boldsymbol{\nu}}\|_{\mathcal{V}}^p\right)^{1/p} & 0 \leq p < \infty, \\ \sup_{\boldsymbol{\nu} \in \Lambda} \|v_{\boldsymbol{\nu}}\|_{\mathcal{V}} & p = \infty. \end{cases}$$

Given a vector of positive weights $\boldsymbol{w} = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$, we define the weighted $\ell_w^q(\Lambda; \mathcal{V})$ space, $0 < q \leq 2$, as the set of $\mathcal{V}$-valued sequences $\boldsymbol{v} = (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ for which the weighted (quasi-) norm

$$\|\boldsymbol{v}\|_{q,\boldsymbol{w};\mathcal{V}} := \left(\sum_{\boldsymbol{\nu} \in \Lambda} w_{\boldsymbol{\nu}}^{2-q} \|v_{\boldsymbol{\nu}}\|_{\mathcal{V}}^q\right)^{1/q},$$

is finite. Notice that $\ell_{\boldsymbol{w}}^2(\Lambda; \mathcal{V})$ coincides with the unweighted space $\ell^2(\Lambda; \mathcal{V})$.

2.2.3 Lebesgue-Bochner and Sobolev spaces

Now, given a probability measure ϱ and $1 \leq p \leq \infty$, in either finite or infinite dimensions, we define the weighted Lebesgue-Bochner space $L_\varrho^p(\mathcal{U}; \mathcal{V})$ as the space consisting of (equivalence

classes of) strongly ϱ-measurable functions $f : \mathcal{U} \to \mathcal{V}$ for which $\|f\|_{L^p_\varrho(\mathcal{U};\mathcal{V})} < \infty$, where

$$\|f\|_{L^p_\varrho(\mathcal{U};\mathcal{V})} := \begin{cases} \left(\int_{\mathcal{U}} \|f(\boldsymbol{y})\|_{\mathcal{V}}^p \,\mathrm{d}\varrho(\boldsymbol{y})\right)^{1/p} & 1 \le p < \infty, \\ \operatorname{ess\,sup}_{\boldsymbol{y}\in\mathcal{U}} \|f(\boldsymbol{y})\|_{\mathcal{V}} & p = \infty. \end{cases} \tag{2.2.6}$$

For further details we refer to [151, Chp. 1]. For simplicity if $\mathcal{V} = \mathbb{C}$ we write $L^p_\varrho(\mathcal{U};\mathbb{C}) = L^p_\varrho(\mathcal{U})$. However, typically in this thesis, $\mathcal{V}$ will be a function space.

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with a polyhedral boundary $\partial\Omega$ with $n \in \{2, 3\}$. We frequently consider $\mathcal{V}$ as the the Lebesgue space $\mathrm{L}^p(\Omega;\mathbb{C})$ of μ-measurable functions $v : \Omega \to \mathbb{C}$ for which $\|v\|_{\mathrm{L}^p(\Omega;\mathbb{C})} < \infty$, where

$$\|v\|_{\mathrm{L}^p(\Omega;\mathbb{C})} := \begin{cases} \left(\int_\Omega |v(\boldsymbol{x})|^p \,\mathrm{d}\mu\right)^{1/p} & 1 \le p < \infty, \\ \operatorname{ess\,sup}_{\boldsymbol{x}\in\Omega} |v(\boldsymbol{x})| & p = \infty. \end{cases} \tag{2.2.7}$$

For further details we refer to [50, §4.2]. Similarly, for simplicity we write $\mathrm{L}^p(\Omega;\mathbb{R}) = \mathrm{L}^p(\Omega)$. Note that $L^p_\varrho(\mathcal{U};\mathcal{V})$ is a space of functions from the parametric variable $\boldsymbol{y}$ to $\mathcal{V}$, and the space $\mathrm{L}^p(\Omega)$ is only a space of functions from the physical variable $\boldsymbol{x}$ to $\mathbb{C}$. We use L and L to distinguish the former from the latter. We now define the main examples considered in this thesis.

We introduce the notation $\mathbf{L}^p(\Omega)$ and $\mathbb{L}^p(\Omega)$ to represent the vectorial and tensorial counterparts of $\mathrm{L}^p(\Omega)$. We write $W^{s,p}(\Omega)$ for the standard Sobolev space with $s \in \mathbb{R}$ and $p > 1$, and we write $\mathrm{H}^1(\Omega)$ when $p = 2$ and $s = 1$. We also define the following closed subspace of $\mathrm{H}^1(\Omega)$ given by

$$\mathrm{H}^1_0(\Omega) := \overline{C^\infty_0(\Omega)}^{\|\cdot\|_{\mathrm{H}^1(\Omega)}}.$$

Here $\overline{C^\infty_0(\Omega)}^{\|\cdot\|_{\mathrm{H}^1(\Omega)}}$ denotes the closure of $C^\infty_0(\Omega)$ (i.e., the space of $C^\infty(\Omega)$ functions with compact support) with respect to the norm $\|\cdot\|_{\mathrm{H}^1(\Omega)}$, which is given by

$$\|v\|_{\mathrm{H}^1(\Omega)} := \left(\|\nabla v\|^2_{\mathbf{L}^2(\Omega)} + \|v\|^2_{\mathrm{L}^2(\Omega)}\right)^{1/2} \qquad \forall v \in \mathrm{H}^1(\Omega).$$

Additionally, we consider the space of traces of functions in $\mathrm{H}^1(\Omega)$, denoted by $\mathrm{H}^{1/2}(\partial\Omega)$, and its dual, $\mathrm{H}^{-1/2}(\partial\Omega)$ (see, e.g., [46, §1.2] for further details).

For scalar functions u and vector fields $\boldsymbol{v}$, we use ∇u and $\operatorname{div}(\boldsymbol{v})$ to denote their gradient and divergence, respectively. For tensor fields $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, represented by $(\sigma_{i,j})_{i,j\in[n]}$ and $(\tau_{i,j})_{i,j\in[n]}$ respectively, we define $\mathbf{div}(\boldsymbol{\sigma})$ as the divergence operator div acting along the rows of $\boldsymbol{\sigma}$, and we define the trace and the tensor inner product, respectively, as

$$\operatorname{tr}(\boldsymbol{\sigma}) = \sum_{i\in[n]} \sigma_{i,i}, \text{ and } \boldsymbol{\tau} : \boldsymbol{\sigma} = \sum_{i,j\in[n]} \tau_{i,j}\sigma_{i,j}.$$

Furthermore, we write $\mathbf{H}^1(\Omega)$ and $\mathbb{H}^1(\Omega)$ for the vectorial and tensorial counterparts of $\mathrm{H}^1(\Omega)$, respectively. Keeping this in mind, we introduce the Banach spaces

$$\begin{aligned} \mathbf{H}(\mathrm{div}_q;\Omega) &:= \Big\{ \boldsymbol{v} \in \mathbf{L}^2(\Omega) : \quad \mathrm{div}(\boldsymbol{v}) \in \mathrm{L}^q(\Omega) \Big\}, \\ \mathbb{H}(\mathbf{div}_q;\Omega) &:= \Big\{ \boldsymbol{\tau} \in \mathbb{L}^2(\Omega) : \quad \mathbf{div}(\boldsymbol{\tau}) \in \mathbf{L}^q(\Omega) \Big\} \end{aligned} \tag{2.2.8}$$

provided with the natural norms

$$\begin{aligned} \|\boldsymbol{v}\|_{\mathbf{H}(\mathrm{div}_q;\Omega)} &:= \|\boldsymbol{v}\|_{\mathbf{L}^2(\Omega)} + \|\mathrm{div}(\boldsymbol{v})\|_{\mathrm{L}^q(\Omega)}, \\ \|\boldsymbol{\tau}\|_{\mathbb{H}(\mathbf{div}_q;\Omega)} &:= \|\boldsymbol{\tau}\|_{\mathbb{L}^2(\Omega)} + \|\mathbf{div}(\boldsymbol{\tau})\|_{\mathbf{L}^q(\Omega)}. \end{aligned}$$

In Chapter 7 we use these spaces with $q = 4/3$ and $q = 2$ in the mixed variational formulations of the considered PDEs. For the latter we simply write $\mathbf{H}(\mathrm{div};\Omega)$.

Often, under certain conditions, such as incompressibility conditions [116, eq.(2.4)], it is convenient to define variations of these spaces. For example, we define

$$\mathbb{L}^2_{\mathrm{tr}}(\Omega) := \Big\{ \boldsymbol{\tau} \in \mathbb{L}^2(\Omega) : \mathrm{tr}(\boldsymbol{\tau}) = 0 \Big\}, \tag{2.2.9}$$

which represents the space of integrable functions with zero trace over Ω. Furthermore, given the decomposition (see, e.g., [114])

$$\mathbb{H}(\mathbf{div}_{4/3};\Omega) = \mathbb{H}_0(\mathbf{div}_{4/3};\Omega) \oplus \mathrm{R}\,\mathbb{I}, \tag{2.2.10}$$

we may also consider

$$\mathbb{H}_0(\mathbf{div}_{4/3};\Omega) := \Big\{ \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}_{4/3};\Omega) : \int_\Omega \mathrm{tr}(\boldsymbol{\tau}) = 0 \Big\}, \tag{2.2.11}$$

as the space of elements in $\mathbb{H}(\mathbf{div}_{4/3};\Omega)$ with zero mean trace. Finally, we define

$$\mathbb{L}^2_{\mathrm{skew}}(\Omega) = \{ \boldsymbol{\eta} \in \mathbb{L}^2(\Omega) : \ \boldsymbol{\eta} + \boldsymbol{\eta}^t = 0 \},$$

and the space of $\mathrm{L}^2(\Omega)$ functions with zero integral over Ω as

$$\mathrm{L}^2_0(\Omega) = \Big\{ \nu \in \mathrm{L}^2(\Omega) : \ \int_\Omega \nu = 0 \Big\}.$$

These spaces will be crucial to define the main examples considered in §7.4.

Finite-dimensional discretizations of $\mathcal{V}$

In most cases, we are unable to work directly in the space $\mathcal{V}$ because it is typically infinite-dimensional. Therefore, we also consider a finite-dimensional subspace

$$\mathcal{V}_K \subseteq \mathcal{V}. \tag{2.2.12}$$

Often, $\mathcal{V}$ represents the solution space of a parametric DE. In this case, $\mathcal{V}_K$ would typically correspond to a FE discretization of $\mathcal{V}$. Assuming (2.2.12) corresponds to considering so-called *conforming discretizations* in the context of FEMs [108, Def. 1.68]. We let $\{\varphi_k\}_{k=1}^K$ be a (not necessarily orthonormal) basis of $\mathcal{V}_K$, where $K = \dim(\mathcal{V}_K) \leq \dim(\mathcal{V})$. We also assume that there is a bounded linear operator

$$\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K. \tag{2.2.13}$$

To simplify several of the subsequent bounds, we define

$$\pi_K = \max\{\|\mathcal{P}_K\|_{\mathcal{V}\to\mathcal{V}}, 1\} \tag{2.2.14}$$

(the assumption $\pi_K \geq 1$ is convenient and is of arguably of little consequence for practical purposes). Note that we do not specify a particular form for this operator except when $\mathcal{V}$ is a Hilbert space – see Remark 2.2.1 for some further discussion.

For convenience, if $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$ is defined everywhere, then we write $\mathcal{P}_K(f)$ for the function defined pointwise as

$$\mathcal{P}_K(f)(\boldsymbol{y}) = \mathcal{P}_K(f(\boldsymbol{y})), \quad \boldsymbol{y} \in \mathcal{U}. \tag{2.2.15}$$

Later, our various error bounds involve $f - \mathcal{P}_K(f)$ measured in suitable Lebesgue-Bochner norms.

Remark 2.2.1 When $\mathcal{V}$ is a Hilbert space, it is natural to choose $\mathcal{P}_K$ as the orthogonal projection onto $\mathcal{V}_K$. Then $\mathcal{P}_K(v)$ is the best approximation in $\mathcal{V}_K$ of $v \in \mathcal{V}$ and (2.2.14) holds with $c_K = 1$. The case of Banach spaces is more delicate. First, the best approximation problem

$$\inf_{z\in\mathcal{V}_K} \|v - z\|_{\mathcal{V}}$$

may not have a unique solution (a solution always exists since $\mathcal{V}_K$ is a finite-dimensional subspace). Thus the best approximation map

$$\mathcal{P}_{\mathcal{V}_K} : v \mapsto \{v_K \in \mathcal{V}_K : \|v - v_K\|_{\mathcal{V}} = \inf_{z\in\mathcal{V}_K} \|v - z\|\} \tag{2.2.16}$$

is set-valued. Furthermore there does not generally exist a linear operator $\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K$ with $\mathcal{P}_K(v) \in \mathcal{P}_{\mathcal{V}_K}(v)$, $\forall v \in \mathcal{V}$. The work in [92, 148] establishes conditions on the set $\mathcal{P}_{\mathcal{V}_K}(v)$ to show the linearity of such operators in a general normed linear space $\mathcal{V}$. From [92, Lem. 2.1], such an operator is bounded with $\|\mathcal{P}_K(v)\|_{\mathcal{V}} \leq 2\|v\|_{\mathcal{V}}$. From [92, Thm. 2.2], a sufficient and necessary condition for the existence a linear operator in $\mathcal{P}_{\mathcal{V}_K}(v)$ is that $\ker(\mathcal{P}_{\mathcal{V}_K})$ contains a closed subspace $\mathcal{W}$ such that $\mathcal{V} = \mathcal{V}_K + \mathcal{W}$, where

$$\ker(\mathcal{P}_{\mathcal{V}_K}) := \{v \in \mathcal{V} : 0 \in \mathcal{P}_{\mathcal{V}_K}(v)\}. \tag{2.2.17}$$

Note further that if $\mathcal{V}$ is strictly convex and $\ker(\mathcal{P}_{\mathcal{V}_K})$ is a subspace of $\mathcal{V}$, then $\mathcal{P}_{\mathcal{V}_K}$ is linear [92, Cor. 2.5]. Moreover, if $\mathcal{V} = \mathrm{L}^p(\Omega)$ with $1 < p < \infty$, the operator $\mathcal{P}_K$ is linear if and only if the quotient space $\mathrm{L}^p(\Omega)/\mathcal{V}_K$ is isometrically isomorphic to some other $\mathrm{L}^q(\Omega)$ space [26, Thm. 5].

2.3 Holomorphy

Due to their relevance to our motivating problem, we are interested in smooth functions with respect to their variables. There is a large body of literature [27, 43, 56, 63, 71–73, 77, 122, 136, 137, 140, 144, 146, 147, 157, 170, 258, 265] that has established that solution maps of a wide range of different parametric DEs are *holomorphic* (i.e., *analytic*) functions of their parameters. Here, we assume that the parameter-to-solution map $\boldsymbol{y} \mapsto u(\boldsymbol{y})$ admits a holomorphic extension to an open neighbourhood of a suitable complex region.

We now recall the definition of holomorphy and holomorphic extension for Banach-valued functions. We note that equivalent definitions are possible (see, e.g., [142, Chp. 2]) and that the definition employed in this thesis is based on the notion of the Gateaux partial derivative. For other details on differentiability we refer to [34, Chp. 17], and the references therein. Note the following definitions apply in both the finite- ($d \in \mathbb{N}$) and infinite- ($d = \infty$) dimensional settings, where we recall that $[d] = \mathbb{N}$ and $\mathbb{C}^d = \mathbb{C}^{\mathbb{N}}$ when $d = \infty$.

Definition 2.3.1 (holomorphy; finite- or infinite-dimensional case)**.** Let $d \in \mathbb{N} \cup \{\infty\}$, $\mathcal{O} \subseteq \mathbb{C}^d$ be an open set and $\mathcal{V}$ be a separable Banach space. A function $f : \mathcal{O} \to \mathcal{V}$ is *holomorphic in* $\mathcal{O}$ if and only if it is holomorphic with respect to each variable in $\mathcal{O}$. That is to say, for any $z \in \mathcal{O}$ and any $j \in [d]$, the following limit exists in $\mathcal{V}$:

$$\lim_{\substack{h \in \mathbb{C} \\ h \to 0}} \frac{f(z + h\boldsymbol{e}_j) - f(z)}{h} \in \mathcal{V}.$$

We now give the definition of holomorphic extensions to open sets $\mathcal{O} \subset \mathbb{C}^d$.

Definition 2.3.2 (holomorphic extension)**.** Let $\mathcal{V}$ be a Banach space. A function $f : \mathcal{U} \to \mathcal{V}$ is *holomorphic* in $\mathcal{U} \subseteq \mathcal{O} \subseteq \mathbb{C}^d$ if it has a holomorphic extension to $\mathcal{O}$, i.e., there is a

$\widetilde{f} : \mathcal{O} \to \mathcal{V}$ that is holomorphic in $\mathcal{O}$ with $\widetilde{f}|_{\mathcal{U}} = f$. In this case, we also define $\|f\|_{L^\infty(\mathcal{O};\mathcal{V})} := \|\widetilde{f}\|_{L^\infty(\mathcal{O};\mathcal{V})}$ or, when $\mathcal{V} = \mathbb{C}$, simply $\|f\|_{L^\infty(\mathcal{O})}$.

Observe that if $\mathcal{O}$ is a closed set, then we say that f is holomorphic in $\mathcal{O}$ if it has a holomorphic extension to some open neighbourhood of $\mathcal{O}$.

We now introduce the precise class of functions considered in this work

2.3.1 The class of $(\boldsymbol{b}, \varepsilon)$-holomorphic functions

In classical polynomial approximation theory, convergence rates can be shown by considering functions that are holomorphic in certain polyellipses, whose size then stipulates the rate of decay of the approximation error. We will clarify this concept later. First, we present some key definitions of classical polynomial approximation theory.

In one dimension, for a given $\rho > 1$, we define the filled-in *Bernstein ellipse* of parameter ρ as the complex region defined by

$$\mathcal{E}_\rho = \left\{ \tfrac{1}{2}(z + z^{-1}) : z \in \mathbb{C}, 1 \leq |z| \leq \rho \right\} \subset \mathbb{C}.$$

Note that this defines an ellipse with ± 1 as its foci and major and minor semi-axis lengths given by $\frac{1}{2}(\rho \pm \rho^{-1})$. In addition, by convention $\mathcal{E}_\rho = [-1, 1]$ when $\rho = 1$. As mentioned before, in classical polynomial approximation theory, any f that is holomorphic in $\mathcal{E}_\rho$ can be approximated by a polynomial of degree n with an error depending on ρ^{-n}. See, e.g., [261, Thm. 8.1].

Now, let $d \in \mathbb{N} \cup \{\infty\}$. Given $\boldsymbol{\rho} = (\rho_j)_{j=1}^d \in \mathbb{R}^d$ with $\boldsymbol{\rho} \geq \mathbf{1}$, we define the filled-in Bernstein polyellipse of paramter $\boldsymbol{\rho}$ as the region in the complex plane defined by the Cartesian product

$$\mathcal{E}_{\boldsymbol{\rho}} = \mathcal{E}_{\rho_1} \times \mathcal{E}_{\rho_2} \times \cdots \subset \mathbb{C}^d.$$

We denote the class of Banach-valued functions that are holomorphic in $\mathcal{E}_{\boldsymbol{\rho}}$ with norm at most one as

$$\mathcal{B}(\boldsymbol{\rho}) = \left\{ f : \mathcal{U} \to \mathcal{V}, f \text{ holomorphic in } \mathcal{E}_{\boldsymbol{\rho}}, \|f\|_{L^\infty(\mathcal{E}_{\boldsymbol{\rho}};\mathcal{V})} \leq 1 \right\}. \tag{2.3.1}$$

In infinite dimensions, we also consider a class of functions that are holomorphic in a certain union of Bernstein polyellipses. Let $\boldsymbol{b} = (b_j)_{j \in \mathbb{N}} \in [0, \infty)^{\mathbb{N}}$, $\varepsilon > 0$ and the complex region defined by

$$\mathcal{R}(\boldsymbol{b}, \varepsilon) = \bigcup \left\{ \mathcal{E}_{\boldsymbol{\rho}} : \boldsymbol{\rho} \geq \mathbf{1}, \sum_{j=1}^{\infty} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) b_j \leq \varepsilon \right\} \subset \mathbb{C}^{\mathbb{N}}. \tag{2.3.2}$$

Keeping this in mind, we now define the class of $(\boldsymbol{b},\varepsilon)$-*holomorphic* functions [12,63,77,235]. Since the seminal work [73], this class of functions has played a crucial role in the context of parametric PDEs (see also [71] and references therein).

Definition 2.3.3. A function $f:\mathcal{U}\to\mathcal{V}$, where $\mathcal{U}=[-1,1]^{\mathbb{N}}$, is $(\boldsymbol{b},\varepsilon)$-*holomorphic* if it is holomorphic in the complex region $\mathcal{R}(\boldsymbol{b},\varepsilon)$.

In analogy with $\mathcal{B}_{\boldsymbol{\rho}}$, we define the class $\mathcal{H}(\boldsymbol{b},\varepsilon)$ of $(\boldsymbol{b},\varepsilon)$-*holomorphic* functions with norm at most one over the domain of holomorphy as

$$\mathcal{H}(\boldsymbol{b},\varepsilon)=\left\{f:\mathcal{U}\to\mathcal{V}, f \text{ holomorphic in } \mathcal{R}(\boldsymbol{b},\varepsilon),\ \|f\|_{L^\infty(\mathcal{R}(\boldsymbol{b},\varepsilon);\mathcal{V})}\leq 1\right\}. \tag{2.3.3}$$

The class of $(\boldsymbol{b},\varepsilon)$-*holomorphic* functions was developed in context of parametric DEs. Over the last decade many works have shown that common parametric DEs (1.2.1) possess solution maps $\boldsymbol{y}\mapsto u(\cdot,\boldsymbol{y})$ that are $(\boldsymbol{b},\varepsilon)$-holomorphic functions of their parameters for suitable $\boldsymbol{b}$ depending on the DE [27,56,63,71,72,77,122,136,137,146,234,234,259]. See [12, Ch. 4] and [71] for overviews.

Note that, for simplicity, one could remove the parameter $\varepsilon>0$ in (2.3.2) by rescaling $\boldsymbol{b}$. In some cases ε is redundant and we simply denote $\mathcal{R}(\boldsymbol{b},1)$ as $\mathcal{R}(\boldsymbol{b})$ and $\mathcal{H}(\boldsymbol{b},1)$ as $\mathcal{H}(\boldsymbol{b})$. However, in other cases, the term $\boldsymbol{b}$ may be fixed and related to the smoothness of a parametric PDE. In those cases, it is convenient to treat $\boldsymbol{b}$ and ε separately. See [12, §3.8] for further detals.

Given this, we now introduce the main assumption in this thesis.

Assumption 2.3.4 (holomorphic extension)**.** Let $d\in\mathbb{N}\cup\{\infty\}$. The unknown target function (see §1.1) $f:\mathcal{U}\to\mathcal{V}$ satisfies $f\in\mathcal{B}(\boldsymbol{\rho})$ as in (2.3.1) for some $\boldsymbol{\rho}\geq 1$ (when $d<\infty$) or $f\in\mathcal{H}(\boldsymbol{b},\varepsilon)$ as in (2.3.3) for some $\boldsymbol{b}\geq\boldsymbol{0}$ (when $d=\infty$).

Affine representations and the class $\mathcal{H}(\boldsymbol{b},\varepsilon)$ of functions

Now, we introduce the concept of affine representations and their importance in the analysis of parametric DEs. These representations provide a simple way to identify the parameter $\boldsymbol{b}\in[0,\infty)^{\mathbb{N}}$ of $\mathcal{H}(\boldsymbol{b},\varepsilon)$-holomorphic functions with a sequence $\{\|\psi_j\|_{L^\infty(\Omega)}\}_{j\in\mathbb{N}}$. The general aspects of this sections are taken from [12,71].

Consider the general parametric DEs problem in (1.2.1). In practice, the parameters $\boldsymbol{y}\in[-1,1]^{\mathbb{N}}$ are used to model a term (or terms) in the DEs to quantitatively characterize the effect of uncertainty on the output $u(\boldsymbol{y})$. For example, they may appear in the definition of the diffusion coefficient $a(\cdot,\boldsymbol{y})$ in the elliptic diffusion equation (1.2.2) or the viscosity of a fluid in the NSB problem (1.2.3). Often, we consider affine parametrizations of the coefficient a.

Keeping this in mind, for a parameter a in a suitable set, we can define the solution map to the problem in (1.2.1) by

$$u : a \mapsto u(a) \in \mathcal{V},$$

where a is a random variable with predescribed probability measures. Thus, we can study the holomorphic extension of $\boldsymbol{y} \mapsto u(\boldsymbol{y})$ by looking at the composition of $\boldsymbol{y} \mapsto a(\boldsymbol{y})$ with the solution map $a \mapsto u(a)$ [12, §4.2.2].

As mentioned above, we are interested in affine representations of the coefficient a. More precisely, let Ω be a physical domain and $\{\psi_j\}_{j\in\mathbb{N}}$ be a sequence of functions in $L^\infty(\Omega)$. We say that an *affine representer* is a function of the form

$$a(\boldsymbol{x}, \boldsymbol{y}) = \overline{a}(\boldsymbol{x}) + \sum_{j\in\mathbb{N}} y_j \psi_j(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \Omega, \boldsymbol{y} \in \mathcal{U}, \tag{2.3.4}$$

where $\overline{a}$ is a fixed function in $\mathrm{L}^\infty(\Omega)$, and the series convergences in the $\mathrm{L}^\infty(\Omega)$-norm. For instance, for the unit square $\Omega = (0,1)^2$, consider the following affine coefficients

$$a(\boldsymbol{x}, \boldsymbol{y}) = 2.62 + \sum_{j\in\mathbb{N}} y_j \frac{\sin(\pi x_1 j)}{j^{3/2}}, \quad \forall \boldsymbol{x} \in \Omega, \ \forall \boldsymbol{y} \in [-1,1]^{\mathbb{N}}. \tag{2.3.5}$$

Affine representations arise naturally in parametric formulations, e.g., where a is a piecewise constant over fixed partitions of the physical domain, describing the shape of a boundary of the physical domain or through the eigenfunctions of a covariance operator in a KL expansion (in an infinite-dimensional case) [71, §1.2].

The Karhunen-Loève expansion is an important tool for describing certain infinite-dimensional models. More precisely, consider a probability space $(S, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is a probability measure over a sample space S and $\mathcal{F}$ is a σ-algebra on S. As a consequence of Mercer's theorem [246, Thm 11.3], any second-order stochastic process $a : \Omega \times S \to$ with continuous covariance function can be represented as an infinite sum of random variables [246, Thm 11.4]. For instance, consider $\overline{a} \in \mathrm{L}^\infty(\Omega)$ and the covariance operator $\mathcal{C} : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$

$$v \mapsto \mathcal{C}(v) = \int_\Omega C_a(\cdot, \boldsymbol{x}) v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \quad C_a(\boldsymbol{z}, \boldsymbol{x}) = \mathbb{E}((a(\boldsymbol{z}, \cdot) - \overline{a}(\boldsymbol{z}))(a(\boldsymbol{x}, \cdot) - \overline{a}(\boldsymbol{x}))),$$

for all $\boldsymbol{x}, \boldsymbol{z} \in \Omega$. Then, the Karhunen-Loève expansion has the form (2.3.4) with $\psi_j = \sqrt{\lambda_j}\phi_j$ and

$$y_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \int_\Omega (a(\boldsymbol{x}, \omega) - \overline{a}(\boldsymbol{x}))\phi_j(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

for all $\omega \in S$ and $j \in \mathbb{N}$, where $\{\lambda_j\}_{j\in\mathbb{N}}$ are the real nonnegative eigenvalues and $\{\phi_j\}_{j\in\mathbb{N}}$ are the corresponding eigenfunctions [12, Ex. 4.6]. Later, in Chapter 7 (see (7.4.2)), we will use the coefficient [209, Eq. (5.2)], which represents the truncation of a one-dimensional

random field with stationary covariance

$$C[\log(a_N) - 0.5](x_1, x_2) = \exp\left(\frac{-(x_1 - x_2)^2}{L_c^2}\right).$$

Affine parametric diffusion equation and the class $\mathcal{H}(\boldsymbol{b}, \varepsilon)$ of functions

Here we give a precise example of a parametric PDE in infinite dimensions whose solution admits a well-defined and holomorphic extension to a complex region containing the filled-in Bernstein polyellipse. Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain, $\partial\Omega$ be the boundary of Ω and $F \in \mathrm{L}^2(\Omega)$. Consider the stationary diffusion equation with parametrized diffusion coefficient $a : \Omega \times \mathcal{U} \to \mathbb{R}$ and homogeneous Dirichlet boundary conditions in (1.2.2).

Suppose that a is as in (2.3.4) with $\bar{a} \in \mathrm{L}^\infty(\Omega)$ and assume that for $\{\psi_j\}_{j\in\mathbb{N}}$ we have that

$$\sum_{j\in\mathbb{N}} |\psi_j(\boldsymbol{x})| \leq \bar{a}(\boldsymbol{x}) - r$$

for some $r > 0$. For instance, take a as in (2.3.5) and $r < 0.00762$. Consider the map $\boldsymbol{y} \mapsto u(\boldsymbol{y})$, where $u(\boldsymbol{y}) \in \mathrm{H}_0^1(\Omega)$ is the unique weak solution of the standard formulation: given $\boldsymbol{y} \in [-1, 1]^{\mathbb{N}}$ find $u(\boldsymbol{y})$ such that

$$\int_\Omega a(\boldsymbol{y})\nabla u(\boldsymbol{y})\nabla v = \int_\Omega Fv, \quad \forall v \in \mathrm{H}_0^1(\Omega). \tag{2.3.6}$$

Then it can be shown [12, Prop. 4.9] that this map is $(\boldsymbol{b}, \varepsilon)$-holomorphic for any $\varepsilon < r$ and $\boldsymbol{b} = (b_j)_{j\in\mathbb{N}}$ such that $b_j \geq \|\psi_j\|_{\mathrm{L}^\infty(\Omega)}$. In the case of (2.3.5) we can take $b_j = \|\sin(\pi j \cdot)/j^{3/2}\|_{\mathrm{L}^\infty(\Omega)} = j^{-3/2}$ and $0 < \varepsilon < 0.00762$. In other words, as claimed, for an affine diffusion coefficient, the terms in the expansion (2.3.5) directly relate to the holomorphy parameter $\boldsymbol{b}$.

The diffusion equation represents a classical problem in parametric PDEs. However, most studies addressing this problem focus on homogeneous Dirichlet boundary conditions (see, e.g., [42,83,85,139,167,172]). Naturally, one may encounter problems with nonhomogeneous Dirichlet boundary conditions and seek to establish conditions under which the parameter to solution map $\boldsymbol{y} \mapsto u(\boldsymbol{y})$ has a holomorphic extension to a certain complex region. To achieve this, we can use mixed formulations, addressing the nonhomogeneous case by introducing an additional unknown to the formulation (see [114] for further details).

To illustrate this, in addition to the above formulation, consider a nonparametric term $g \in \mathrm{H}^{1/2}(\partial\Omega)$. Here $\mathrm{H}^{1/2}(\partial\Omega)$ is the trace space on the boundary $\partial\Omega$ as defined in §2.2. Now, consider the linear elliptic equation with Dirichlet boundary conditions

$$\begin{aligned} -\mathrm{div}(a(\boldsymbol{x}, \boldsymbol{y})\nabla u(\boldsymbol{x}, \boldsymbol{y})) &= F(\boldsymbol{x}), \quad \text{in } \Omega \\ u(\boldsymbol{x}, \boldsymbol{y}) &= g(\boldsymbol{x}), \quad \text{on } \partial\Omega. \end{aligned}$$

Then, we can prove that the map $\boldsymbol{y} \in \mathcal{U} \mapsto (u, \boldsymbol{\sigma})(\cdot, \boldsymbol{y}) \in \mathrm{L}^2(\Omega) \times \mathbf{H}(\mathrm{div}; \Omega)$ is also $(\boldsymbol{b}, \varepsilon)$-holomorphic for the same $\boldsymbol{b} \geq 0$ and $\varepsilon > 0$. Here $\boldsymbol{\sigma}$ is an additional variable given by $\boldsymbol{\sigma}(\boldsymbol{y}) = a(\boldsymbol{y})\nabla u(\boldsymbol{y})$ in Ω. The space $\mathrm{L}^2(\Omega) \times \mathbf{H}(\mathrm{div}; \Omega)$ comes from using mixed formulations as in (A.2.3). In particular, $\mathbf{H}(\mathrm{div}; \Omega)$ comes from the need of $\mathrm{div}(\boldsymbol{\sigma}) \in \mathrm{L}^2(\Omega)$. We will return to this example in Chapter 7. See Appendix A for a proof.

2.3.2 Parametric dependence: known and unknown cases

Now return to the general setting, where $f : \mathcal{U} \to \mathcal{V}$ is a $(\boldsymbol{b}, \varepsilon)$-holomorphic function of infinitely many variables. It is worth noting that the $\boldsymbol{b} \in [0, \infty)^{\mathbb{N}}$ is a nonnegative sequence that controls the *anisotropy* of functions in the class $\mathcal{H}(\boldsymbol{b}, \varepsilon)$. Specifically, large b_j means that the condition

$$\sum_{j=1}^{\infty} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) b_j \leq \varepsilon$$

in (2.3.2) holds only for smaller values of the parameter ρ_j, meaning that functions in $\mathcal{H}(\boldsymbol{b}, \varepsilon)$ are less smooth with respect to the variable y_j. That is, they have analytic continuations only to small Bernstein ellipses in this variable. Conversely, if b_j is small (or in the extreme, $b_j = 0$), then functions in $\mathcal{H}(\boldsymbol{b}, \varepsilon)$ possess analytic continuations to larger Bernstein ellipses, and are therefore smoother with respect to the variable y_j.

With this in mind, we now distinguish two important cases for holomorphic function approximation in infinite dimensions.

Known anisotropy

In some settings, the parameter $\boldsymbol{b}$ may be known. In other words, we have prior understanding about the behaviour of the target function with respect to its variables. We refer to this as the *known anisotropy* case. This setting has its advantages. For instance, one can strive to use this information to design an approximation scheme based on $\boldsymbol{b}$. To be more precise, having information about $\boldsymbol{b}$, can be useful when choosing a suitable index set for constructing a polynomial approximation to the target function.

Unknown anisotropy

As discussed, for some particular types of DEs, such as those described in §2.3.1, we can establish bounds for $\boldsymbol{b}$. However, in the more practical UQ setting, where f is considered a black box (i.e., the underlying DE model, if one exists, is hidden) we usually do not have such information. Moreover, even if we can find a sufficient value $\boldsymbol{b} = (b_j)_{j \in \mathbb{N}} = (\|\psi_j\|_{\mathrm{L}^\infty(\Omega)})_{j \in \mathbb{N}}$, such as in the case of the Poisson problem with parametric coefficient (2.3.5), this value may not be sharp. This comes from the fact that $\mathcal{R}(\boldsymbol{b}, \varepsilon) \subseteq \mathcal{R}(\boldsymbol{b}', \varepsilon)$ (see (2.3.2) and (2.4.19)) for

any $\mathbf{0} \leq \boldsymbol{b}' \leq \boldsymbol{b}$. Which makes difficult to know an optimal value for $\boldsymbol{b}$. We refer to *unknown anisotropy* to the case where we do not have information about $\boldsymbol{b}$.

Motivated by this discussion, in this work we consider both the known and unknown anisotropy settings.

2.3.3 Holomorphy and polynomial approximation

Holomorphy motivates the use of polynomial approximation. A particularly important concept in this area is the best *s-term polynomial approximation*, which serves as a key theoretical benchmark. Here, the function f is approximated by an s-term expansion corresponding to its largest s coefficients (measured in the $\mathcal{V}$-norm) with respect to a polynomial basis. Common choices include Taylor polynomials, tensor-product Legendre and Chebyshev polynomials on bounded hypercubes or tensor-product Hermite and Laguerre polynomials on $\mathbb{R}^d$ or $[0,\infty)^d$. Over the last fifteen years, there has been a significant effort in developing the theory of best s-term polynomial approximation (see the aforementioned references, plus those in §2.3). Signature results have established *exponential* and *algebraic* convergence rates for the best s-term approximation. The former assert that the error decays at least exponentially fast in $s^{1/d}$ in finite dimensions for any holomorphic function. The latter assert that the error decays algebraically fast; specifically, like $s^{1/2-1/p}$ for some $0 < p < 1$. These algebraic rates also hold in infinite dimensions, thus establishing best s-term approximation as a (theoretical) means to approximate holomorphic functions of infinitely many variables. We review several such results in §2.4.3.

2.4 Best s-term polynomial approximation

In this section we introduce one of the most important tools used in this work, orthogonal polynomials and polynomials expansions.

2.4.1 Orthogonal polynomial expansions

From [203, §2.1] (or [248, §2.2]) under mild assumptions that are always fulfilled in the context of this work (ϱ is a Lebesgue-Stiltjes probability measure on $\mathbb{R}$ and has finite polynomial moments of all orders) there exists a unique orthonormal polynomial basis $\{\Psi_\nu\}_{\nu\in\mathbb{N}_0}$ of $L^2_\varrho([-1,1])$, where $\Psi_\nu = \Psi^{(1)}_\nu$ is a polynomial of degree ν. In particular, when $\varrho^{(1)}$ is the measure in (2.2.1), we obtain the Legendre and Chebyshev polynomials, respectively.

Consider the multi-index set $\mathcal{F} \subseteq \mathbb{N}_0^d$ and the corresponding tensor-product measure ϱ on $\mathcal{U} = [-1,1]^d$, defined as in (2.1.1) and (2.2.2) when $d < \infty$, or as defined in (2.1.2) and

(2.2.4) when $d = \infty$. Then, the set of functions $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}} \subset L^2_\varrho(\mathcal{U})$ given by the tensorization

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{k\in[d]} \Psi_{\nu_k}(y_k), \quad \boldsymbol{y} \in \mathcal{U},\ \boldsymbol{\nu} \in \mathcal{F}, \tag{2.4.1}$$

form an orthonormal basis of $L^2_\varrho(\mathcal{U})$. Note that, by the definition in (2.1.2), in the infinite dimensional case ($d = \infty$) (2.4.1) is well-defined since any $\boldsymbol{\nu} \in \mathcal{F}$ has only finitely-many nonzero terms. Therefore, when $d = \infty$ (2.4.1) is equivalent to

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{k:\nu_k\neq 0} \Psi_{\nu_k}(y_k),$$

which is a product of finitely-many terms. Note that $\Psi_0^{(1)} = 1$ since $\varrho^{(1)}$ is a probability measure.

Let f be a function in the Lebesgue-Bochner space $L^2_\varrho(\mathcal{U};\mathcal{V})$ defined in §2.2. Then it has a convergent expansion (in $L^2_\varrho(\mathcal{U};\mathcal{V})$) given by

$$f = \sum_{\boldsymbol{\nu}\in\mathcal{F}} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}, \qquad \text{where } c_{\boldsymbol{\nu}} := \int_{\mathcal{U}} f(\boldsymbol{y})\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})\,\mathrm{d}\varrho(\boldsymbol{y}) \in \mathcal{V}, \tag{2.4.2}$$

and the *coefficients* $c_{\boldsymbol{\nu}}$ are elements of $\mathcal{V}$. Now let $S \subset \mathcal{F}$ be a multi-index set and

$$\mathcal{P}_{S;\mathcal{V}} = \left\{ \sum_{\boldsymbol{\nu}\in S} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}} : c_{\boldsymbol{\nu}} \in \mathcal{V} \right\} \subset L^2_\varrho(\mathcal{U};\mathcal{V}). \tag{2.4.3}$$

Then, the $L^2(\mathcal{U};\mathcal{V})$-norm *best s-term polynomial approximation* f_s of f is defined as

$$f_s \in \operatorname{argmin}\left\{ \|f - g\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} : g \in \mathcal{P}_{S,\mathcal{V}},\ S \subset \mathcal{F},\ |S| = s \right\}. \tag{2.4.4}$$

Intuitively, the best s-term polynomial approximation is a theoretical benchmark that aims to measure how well one can approximate a function f using polynomials with indices from an arbitrary index set S of a given size s.

Suppose that $\mathcal{V}$ is a Hilbert space. Then, a short exercise with Parseval's identity gives that

$$\inf\{\|f - g\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} : g \in P_{S;\mathcal{V}}\}$$

is achieved by the expansion

$$f_s = \sum_{\boldsymbol{\nu}\in S^*} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}, \tag{2.4.5}$$

where $S^* \subset \mathcal{F}$, $|S^*| = s$, is the set of consisting of the multi-indices of the largest s values of the coefficient norms $(\|c_{\boldsymbol{\nu}}\|_{\mathcal{V}})_{\boldsymbol{\nu}\in\mathcal{F}}$. In this case, when $\mathcal{V}$ is a Hilbert space, using Parseval's

identity we obtain that the error satisfies

$$\|f - f_s\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \sqrt{\sum_{\boldsymbol{\nu}\notin S^*} \|c_{\boldsymbol{\nu}}\|^2_{\mathcal{V}}}. \tag{2.4.6}$$

We note in passing another important property of the best s-term approximation, the map $f \mapsto f_s$ is nonlinear. Since given two functions $f, g \in L^2(\mathcal{U};\mathcal{V})$ we have that $(f+g)_s \neq f_s + g_s$.

2.4.2 Sparsity and best s-term approximation error

The equivalence (2.4.6) motivates studying s-term approximation of the sequences of polynomial coefficients. We now provide a series of definitions pertaining to this study. We recall that here and elsewhere, for a sequence $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ and a set $S \subseteq \Lambda$, we define $\boldsymbol{c}_S$ as the sequence with $\boldsymbol{\nu}$th entry equal to $c_{\boldsymbol{\nu}}$ if $\boldsymbol{\nu} \in S$ and zero otherwise.

Definition 2.4.1 (Sparsity)**.** Let $\Lambda \subseteq \mathcal{F}$ denote a (possibly infinite) multi-index set, and $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ be a $\mathcal{V}$-valued sequence. The *support* of $\boldsymbol{c}$ is the set

$$\operatorname{supp}(\boldsymbol{c}) = \{\boldsymbol{\nu} \in \Lambda : \|c_{\boldsymbol{\nu}}\|_{\mathcal{V}} \neq 0\}. \tag{2.4.7}$$

A sequence is *s-sparse* for some $s \in \mathbb{N}_0$ satisfying $s \leq |\Lambda|$ if it has at most s nonzero entries, i.e.,

$$|\operatorname{supp}(\boldsymbol{c})| \leq s.$$

The set of such vectors is denoted by Σ_s.

Definition 2.4.2 (Best s-term approximation error)**.** Let $\Lambda \subseteq \mathcal{F}$ denote a (possibly infinite) multi-index set, $0 < p \leq \infty$, $\boldsymbol{c} \in \ell^p(\Lambda;\mathcal{V})$ and $s \in \mathbb{N}_0$ with $s \leq |\Lambda|$. The *ℓ^p-norm best s-term approximation error* of $\boldsymbol{c}$ is

$$\sigma_s(\boldsymbol{c})_{p;\mathcal{V}} = \min\left\{\|\boldsymbol{c} - \boldsymbol{z}\|_{p;\mathcal{V}} : \boldsymbol{z} \in \ell^p(\Lambda;\mathcal{V}),\ |\operatorname{supp}(\boldsymbol{z})| \leq s\right\}, \tag{2.4.8}$$

where this norm and the space $\ell^p(\Lambda;\mathcal{V})$ are defined in §2.2.

Notice that this is equivalent to

$$\sigma_s(\boldsymbol{c})_{p;\mathcal{V}} = \inf\left\{\|\boldsymbol{c} - \boldsymbol{c}_S\|_{p;\mathcal{V}} : S \subseteq \Lambda,\ |S| \leq s\right\}.$$

Recall that the space $\ell^p(\Lambda;\mathcal{V})$ and $\|\cdot\|_{p;\mathcal{V}}$ are defined in §2.2.

Let $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ be the coefficients of some function $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$, as defined in (2.4.2). Then, when $p = 2$ and $\mathcal{V}$ is a Hilbert space, we have the following:

$$\sigma_s(\boldsymbol{c})_{2;\mathcal{V}} = \|f - f_s\|_{L^2_\varrho(\mathcal{U};\mathcal{V})},$$

where f_s is its best s-term polynomial approximation (2.4.4). Thus, we can study the error of f_s by studying the quantity $\sigma_s(\boldsymbol{c})_{2;\mathcal{V}}$. For notational purposes, we denote this quantity in terms of the coefficients $\boldsymbol{c}$. However, in other works, this term is sometimes referred to as $\sigma_s(f)_{2;\mathcal{V}}$.

In later chapters, it is also useful to consider approximations to sequences involving weights that penalize large coefficients at certain indices. Now, let $\Lambda \subseteq \mathcal{F}$ and $\boldsymbol{w} = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda} > \boldsymbol{0}$ be positive weights. Given a set $S \subseteq \Lambda$, we define its weighted cardinality as

$$|S|_{\boldsymbol{w}} := \sum_{i\in S} w_i^2.$$

Note that $|S|_{\boldsymbol{w}}$ may take values in $\mathbb{R} \cup \{+\infty\}$. The following two definitions extend Definitions 2.4.1 and 2.4.2 to the weighted setting:

Definition 2.4.3 (Weighted sparsity)**.** Let $\Lambda \subseteq \mathcal{F}$. A $\mathcal{V}$-valued sequence $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ is *weighted* $(k, \boldsymbol{w})$*-sparse* for some $k \geq 0$ and weights $\boldsymbol{w} = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda} > \boldsymbol{0}$ if

$$|\operatorname{supp}(\boldsymbol{c})|_{\boldsymbol{w}} \leq k,$$

where $\operatorname{supp}(\boldsymbol{z}) = \{\boldsymbol{\nu} : \|z_{\boldsymbol{\nu}}\|_{\mathcal{V}} \neq 0\}$ is the *support* of $\boldsymbol{z}$. The set of such vectors is denoted by $\Sigma_{k,\boldsymbol{w}}$.

Definition 2.4.4 (Weighted best $(k, \boldsymbol{w})$-term approximation error)**.** Let $\Lambda \subseteq \mathcal{F}$, $0 < p \leq 2$, $\boldsymbol{w} > \boldsymbol{0}$, $\boldsymbol{c} \in \ell^p_{\boldsymbol{w}}(\Lambda; \mathcal{V})$ and $k \geq 0$. The $\ell^p_{\boldsymbol{w}}$*-norm weighted best* $(k, \boldsymbol{w})$*-term approximation error* of $\boldsymbol{c}$ is

$$\sigma_k(\boldsymbol{c})_{p,\boldsymbol{w};\mathcal{V}} = \min\left\{\|\boldsymbol{c} - \boldsymbol{z}\|_{p,\boldsymbol{w};\mathcal{V}} : \boldsymbol{z} \in \Sigma_{k,\boldsymbol{w}}\right\}. \tag{2.4.9}$$

Recall that the space $\ell^p_{\boldsymbol{w}}(\Lambda; \mathcal{V})$ and $\|\cdot\|_{p,\boldsymbol{w};\mathcal{V}}$ are defined in §2.2. Notice that this is equivalent to

$$\sigma_k(\boldsymbol{c})_{p,\boldsymbol{w};\mathcal{V}} = \inf\left\{\|\boldsymbol{c} - \boldsymbol{c}_S\|_{p,\boldsymbol{w};\mathcal{V}} : S \subseteq \Lambda,\ |S|_{\boldsymbol{w}} \leq k\right\}. \tag{2.4.10}$$

2.4.3 Rates of best s-term polynomial approximation

As noted, best s-term polynomial approximation of holomorphic functions is a well-studied subject, especially in the context of solutions of parametric DEs. See, e.g., [35, 36, 43, 47, 63, 72, 73, 136, 215, 255, 258] and, in particular, [71] and [12, Chpt. 3]. Here, we recap two standard types of error decay rates for this approximation, those of *algebraic* and *exponential* type, respectively. Note that these results are for Chebyshev and Legendre polynomial approximations – the main focus of this work. The latter type of decay rate holds in finite dimensions, while the former holds in both finite and infinite dimensions. In this work, these error decay rates serve as the optimal benchmark against which to compare the approximations computed from sample values.

The following results are standard in the Hilbert-valued case, and have appeared in various different guises in the aforementioned works. See for instance [12, Chpt. 3].

Theorem 2.4.5 (Algebraic rates of convergence; finite-dimensional case)**.** *Let $\mathcal{V}$ be a Hilbert space, $0 < p < \infty$ and $f \in \mathcal{B}(\boldsymbol{\rho})$ for some $\boldsymbol{\rho} \geq \mathbf{1}$. Let $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathbb{N}_0^d}$ be as in (2.4.2). Then, for every $s \geq 1$ there are sets $S_1, S_2 \subset \mathcal{F}$, $|S_1|, |S_2| \leq s$, such that*

$$\|f - f_{S_1}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1/2-1/p}, \qquad \|f - f_{S_2}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1-1/p}, \tag{2.4.11}$$

where $f_{S_i} = \sum_{\boldsymbol{\nu} \in S_i} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ for $i = 1, 2$ and $C = C(d, p, \boldsymbol{\rho}) > 0$ depends on d, p and $\boldsymbol{\rho}$ only.

Theorem 2.4.6 (Algebraic rates of convergence; infinite-dimensional case)**.** *Let $\mathcal{V}$ be a Hilbert space, $0 < p < 1$, $\varepsilon > 0$, $\boldsymbol{b} = (b_j)_{j \in \mathbb{N}} \in \ell^p(\mathbb{N})$ with $\boldsymbol{b} \geq \mathbf{0}$ and $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$, where $\mathcal{H}(\boldsymbol{b}, \varepsilon)$ is as in (2.3.3). Then, for every $s \geq 1$ there are sets $S_1, S_2 \subset \mathcal{F}$, $|S_1|, |S_2| \leq s$, such that*

$$\|f - f_{S_1}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1/2-1/p}, \qquad \|f - f_{S_2}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq C \cdot s^{1-1/p}, \tag{2.4.12}$$

where $f_{S_i} = \sum_{\boldsymbol{\nu} \in S_i} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ for $i = 1, 2$ and $C = C(\boldsymbol{b}, \varepsilon, p) > 0$ depends on $\boldsymbol{b}$, ε and p only.

We next state a result on exponential convergence in finite dimensions. Such rates have been established in various different works (see, e.g., [35, 36, 71, 215, 258]). The following result is a minor modification of [12, Thm. 3.15], in which we allow arbitrary $s \geq 1$ at the expense of a constant C in the error bound.

Theorem 2.4.7 (Exponential rates of convergence; finite-dimensional case)**.** *Let $\mathcal{V}$ be a Hilbert space, $0 < p \leq 2$ and $f \in \mathcal{B}(\boldsymbol{\rho})$ for some $\boldsymbol{\rho} \geq \mathbf{1}$. Let $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathbb{N}_0^d}$ be as in (2.4.2). Then, for every $s \geq 1$ there is a set $S \subset \mathcal{F}$, $|S| \leq s$, such that*

$$\|f - f_S\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f - f_S\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq C \cdot \exp(-\gamma s^{1/d}), \tag{2.4.13}$$

for all

$$0 < \gamma < (d+1)^{-1} \left(d! \prod_{j=1}^{d} \ln(\rho_j) \right)^{1/d}, \tag{2.4.14}$$

where $f_S = \sum_{\boldsymbol{\nu} \in S} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ and $C = C(d, \gamma, p, \boldsymbol{\rho}) > 0$ is a constant depending on d, γ, p and $\boldsymbol{\rho}$ only.

Remark 2.4.8 It is possible to improve the rate (2.4.13) by removing the $(d+1)^{-1}$ factor in (2.4.14) [258]. The difficulty in doing this is that such rates are not necessarily attained in lower sets (this is, however, true if $\boldsymbol{\rho}$ is sufficiently large – see [12, Lem. 7.20]). As we discuss next, lower sets are a crucial ingredient in our analysis. Conversely, the rates described in Theorem 2.4.7 can always be attained in lower sets.

2.4.4 Lower and anchored sets

A common thread throughout this thesis is the construction of polynomial approximations that satisfy similar error bounds to those of the best s-term approximation f_s, for any holomorphic function f. Hence, ideally, we would have access to the multi-index set S corresponding to the largest s coefficients of f (measured in the $\mathcal{V}$-norm). However, this is not possible in general, since the only information we have about f is its values at a finite number of sample points. Another problem is that its largest coefficients could occur at arbitrarily-large multi-indices. Fortunately, it is well known that near-best s-term polynomial approximations can be constructed using sets of multi-indices with additional structure. These are *lower* sets (used in the finite-dimensional case) and *anchored* sets (used in the infinite-dimensional case). Classical references for lower and anchored sets include [89, 171, 186, 254].

Definition 2.4.9. A set $\Lambda \subseteq \mathcal{F}$ is *lower* if the following holds for every $\boldsymbol{\nu}, \boldsymbol{\mu} \in \mathcal{F}$:

$$(\boldsymbol{\nu} \in \Lambda \text{ and } \boldsymbol{\mu} \leq \boldsymbol{\nu}) \implies \boldsymbol{\mu} \in \Lambda.$$

A set $\Lambda \subseteq \mathcal{F}$ is *anchored* if it is lower and if the following holds for every $j \in \mathbb{N}$:

$$\boldsymbol{e}_j \in \Lambda \implies \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{j-1}\} \subseteq \Lambda.$$

More recently, these structures have been used extensively in the construction of interpolation, least-squares and compressed sensing schemes for polynomial approximation with desirable sample complexity bounds (see, e.g., [12] and references therein).

Minimal monotone majorant

In the infinite-dimensional case, we need an additional assumption on $\boldsymbol{b}$ in order to establish convergence of the various approximation methods. Let $\boldsymbol{z} = (z_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be a sequence. We define its *minimal monotone majorant* as

$$\tilde{\boldsymbol{z}} = (\tilde{z}_i)_{i\in\mathbb{N}}, \quad \text{where } \tilde{z}_i = \sup_{j\geq i} |z_j|, \ \forall i \in \mathbb{N}. \tag{2.4.15}$$

Then, given $0 < p < \infty$, we define the *monotone* ℓ^p space $\ell^p_{\mathsf{M}}(\mathbb{N})$ as

$$\ell^p_{\mathsf{M}}(\mathbb{N}) = \{\boldsymbol{z} \in \ell^\infty(\mathbb{N}) : \|\boldsymbol{z}\|_{p,\mathsf{M}} := \|\tilde{\boldsymbol{z}}\|_p < \infty\}. \tag{2.4.16}$$

In subsequent chapters, we will often assume that $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$.

2.4.5 Weights and error bounds for Banach-valued functions

Theorems 2.4.5-2.4.7 provide error bounds for polynomial approximations. However, they have the limitations that they only apply to Hilbert-valued functions, and the sets whose existence they assert may not be structured in the above sense (see §2.4.4). We now recap a series of results that are valid for Banach-valued functions, and considered structured sets.

In the previous section, we introduced the concept of weighted sparsity (Definition 2.4.3) and the definition of weighted best $(k, \boldsymbol{w})$-term approximation error (Definition 2.4.4) with reference to general weights. However, a specific type of weights $\boldsymbol{u} \geq \boldsymbol{1}$, known as *intrinsic weights*, offer useful properties that we will extensively utilize in this thesis. We commence by specifying these weights.

$$\boldsymbol{w} = \boldsymbol{u} = (u_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}, \quad u_{\boldsymbol{\nu}} = \|\Psi_{\boldsymbol{\nu}}\|_{L^\infty(\mathcal{U})}, \ \boldsymbol{\nu} \in \Lambda. \tag{2.4.17}$$

These weights have the property that if the coefficients $\boldsymbol{c} \in \ell^1_{\boldsymbol{u}}(\Lambda; \mathcal{V})$ of a polynomial expansion of u given by (2.4.2) then $u \in L^\infty(\mathcal{U}; \mathcal{V})$ and the expansion converges in the $L^\infty(\mathcal{U}; \mathcal{V})$-norm. Moreover, applying triangle inequality, the definition of the $L^\infty(\mathcal{U}; \mathcal{V})$-norm in (2.2.6) and the fact that $\boldsymbol{u} \geq \boldsymbol{1}$, we have the bound

$$\Big\| \sum_{\boldsymbol{\nu}\in\Lambda} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}} \Big\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq \sum_{\boldsymbol{\nu}\in\Lambda} \|c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq \sum_{\boldsymbol{\nu}\in\Lambda} \|c_{\boldsymbol{\nu}}\|_{\mathcal{V}}\|\Psi_{\boldsymbol{\nu}}\|_{L^\infty(\mathcal{U})} = \sum_{\boldsymbol{\nu}\in\Lambda} \|c_{\boldsymbol{\nu}}\|_{\mathcal{V}} u_{\boldsymbol{\nu}} = \|\boldsymbol{c}\|_{1,\boldsymbol{u};\mathcal{V}}. \tag{2.4.18}$$

In particular, for Chebyshev and Legendre polynomials these are given explicitly by

$$u_{\boldsymbol{\nu}} = \|\Psi_{\boldsymbol{\nu}}\|_{L^\infty(\mathcal{U})} = \begin{cases} \prod_{j=1}^d \sqrt{2\nu_j + 1}, & \text{Legendre} \\ 2^{\|\boldsymbol{\nu}\|_0/2}, & \text{Chebyshev} \end{cases}$$

where $\|\boldsymbol{\nu}\|_0 := |\mathrm{supp}(\nu)|$ for all $\boldsymbol{\nu} \in \Lambda$.

The finite-dimensional case

Consider the finite-dimensional case, where $\mathcal{U} = [-1, 1]^d$ for $d < \infty$ and $f : \mathcal{U} \to \mathcal{V}$ is a Banach-valued function. We now summarize the various approximation error bounds in the following theorem. This result combines various well-known results in the literature. It is essentially the same as [12, Thm. 3.25]. However, we have made a number of minor edits to fit the notation and setup of this thesis (see Remark 2.4.11 below).

Theorem 2.4.10 (Best s-term decay rates; finite dimensions)**.** *Let $d \in \mathbb{N}$, $f \in \mathcal{B}(\boldsymbol{\rho})$ for some $\boldsymbol{\rho} \geq \boldsymbol{1}$, where $\mathcal{B}(\boldsymbol{\rho})$ is as in (2.3.1), and $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathbb{N}_0^d}$ be its Chebyshev or Legendre coefficients. Then the following best s-term decay rates hold:*

(i) for any $0 < p \leq q \leq 2$ *and* $s \in \mathbb{N}$, *there exists a lower set* $S \subset \mathbb{N}_0^d$ *of size* $|S| \leq s$ *such that*

$$\sigma_s(\boldsymbol{c})_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}} \leq C \cdot s^{1/q-1/p},$$

where $\sigma_s(\boldsymbol{c})_{q;\mathcal{V}}$ *is as in Definition 2.4.2 (with* $\Lambda = \mathbb{N}_0^d$*),* $\boldsymbol{u}$ *is as in (2.4.17) and* $C = C(d, p, \boldsymbol{\rho}) > 0$ *depends on* d, p *and* $\boldsymbol{\rho}$ *only;*

(ii) for any $0 < p \leq q \leq 2$ *and* $k > 0$,

$$\sigma_k(\boldsymbol{c})_{q,\boldsymbol{u};\mathcal{V}} \leq C \cdot k^{1/q-1/p},$$

where $\sigma_k(\boldsymbol{c})_{q,\boldsymbol{u};\mathcal{V}}$ *is as in Definition 2.4.4,* $\boldsymbol{u}$ *is as in (2.4.17) (with* $\Lambda = \mathbb{N}_0^d$*) and* $C = C(d, p, \boldsymbol{\rho}) > 0$ *depends on* d, p *and* $\boldsymbol{\rho}$ *only;*

(iii) for any $0 < p \leq 2$,

$$0 < \gamma < (d+1)^{-1} \left(d! \prod_{j=1}^{d} \log(\rho_j) \right)^{1/d},$$

and $s \in \mathbb{N}$, *there exists a lower set* $S \subset \mathbb{N}_0^d$ *of size* $|S| \leq s$ *such that*

$$\sigma_s(\boldsymbol{c})_{p;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{p;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{p,\boldsymbol{u};\mathcal{V}} \leq C \cdot \exp(-\gamma s^{1/d}),$$

where $\sigma_s(\boldsymbol{c})_{p;\mathcal{V}}$ *is as in Definition 2.4.2 (with* $\Lambda = \mathbb{N}_0^d$*),* $\boldsymbol{u}$ *is as in (2.4.17) and* $C = C(d, \gamma, p, \boldsymbol{\rho}) > 0$ *depends on* d, γ, p *and* $\boldsymbol{\rho}$ *only.*

Remark 2.4.11 There are several differences between Theorem 2.4.10 and [12, Thm. 3.25]. A minor difference is that we do not specify the various constants C appearing in the result. Another difference is in the presentation of (iii). Here we allow arbitrary $s \geq 1$ (instead of $s \geq \bar{s}$) at the expense of a larger (and unspecified) constant C. The main difference, however, is the additional term $\|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}}$ appearing in (i). This can be shown as follows. First, one defines the sequence $\bar{\boldsymbol{c}} = (u_{\boldsymbol{\nu}}^{2/q-1} c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathbb{N}_0^d}$ so that $\|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}} = \|\bar{\boldsymbol{c}} - \bar{\boldsymbol{c}}_S\|_{q;\mathcal{V}}$ and then uses Stechkin's inequality in lower sets (see, e.g., [12, Lem. 3.9]) to show that $\|\bar{\boldsymbol{c}} - \bar{\boldsymbol{c}}_S\|_{q;\mathcal{V}} \leq s^{1/q-1/p} \|\bar{\boldsymbol{c}}\|_{p,\mathsf{M};\mathcal{V}}$, where $\|\cdot\|_{p,\mathsf{M};\mathcal{V}}$ is the norm on the *majorant* ℓ^p *space* $\ell^p_{\mathsf{M}}(\mathbb{N}_0^d; \mathcal{V})$ (see, e.g., [12, Defn. 3.8]). Finally, it can be shown that $\|\bar{\boldsymbol{c}}\|_{p,\mathsf{M};\mathcal{V}} \leq C(d, p, \boldsymbol{\rho})$ using standard arguments. See, e.g., [12, Lem. 7.19] (this lemma only considers the scalar-valued case; however the extension to the Banach-valued case is straightforward).

Remark 2.4.12 In the Hilbert-valued case, Theorem 2.4.10 implies Theorems 2.4.5 and 2.4.7. For the former, we note that $\|f - f_{S_1}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \|\boldsymbol{c} - \boldsymbol{c}_{S_1}\|_{2;\mathcal{V}}$ and from (2.4.18) we have that $\|f - f_{S_2}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq \|\boldsymbol{c} - \boldsymbol{c}_{S_2}\|_{1,\boldsymbol{u};\mathcal{V}}$. We then apply (i) with $q = 2$ or $q = 1$. For the latter, we use (iii) with $p = 1$. Also note that proving Theorem 2.4.10 and Theorem 2.4.13 (see also [12, Thm. 3.25 and Thm. 3.15]) involves establishing the summability of a bounding

sequence for the $\mathcal{V}$-norms of the polynomial coefficients in (2.4.2). This bound is equal to $\|f\|_{L^\infty(\mathcal{E}_{\boldsymbol{\rho}};\mathcal{V})}$ multiplied by a factor that is independent of f and depending on $\boldsymbol{\rho}$ only. Consequently, the index set in Theorem 2.4.10 and Theorem 2.4.13 are independent of f.

The infinite-dimensional case

We now consider the infinite-dimensional case, where $d = \infty$ and $\mathcal{U} = [-1,1]^{\mathbb{N}}$. The following theorem is based on [12, Thms. 3.29 and 3.33].

Theorem 2.4.13 (Best s-term decay rates; infinite-dimensional case)**.** *Let ϱ be the tensor-product uniform or Chebyshev measure on $\mathcal{U} = [-1,1]^{\mathbb{N}}$ and $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$ be the corresponding tensor-product orthonormal Legendre or Chebyshev polynomial basis of $L^2_\varrho(\mathcal{U})$. Let $0 < p < 1$, $\varepsilon > 0$, $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with $\boldsymbol{b} > \boldsymbol{0}$ and $f \in \mathcal{H}(\boldsymbol{b},\varepsilon)$, where $\mathcal{H}(\boldsymbol{b},\varepsilon)$ is as in (2.3.3). Let $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ be the Chebyshev or Legendre coefficients of f, as in (2.4.2). Then the following best s-term decay rates hold:*

(i) For any $p \leq q < \infty$ and $s \in \mathbb{N}$, there exists a lower set $S \subset \mathcal{F}$ of size $|S| \leq s$ such that

$$\sigma_s(\boldsymbol{c})_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}} \leq C \cdot s^{1/q-1/p},$$

where $\sigma_s(\boldsymbol{c})_{q;\mathcal{V}}$ is as in (2.4.8) (with $\Lambda = \mathcal{F}$), $\boldsymbol{u}$ is as in (2.4.17) and $C = C(\boldsymbol{b},\varepsilon,p) > 0$ depends on $\boldsymbol{b}$, ε and p only.

(ii) For any $p \leq q \leq 2$ and $k > 0$, there exists a set $S \subset \mathcal{F}$ with $|S|_{\boldsymbol{u}} \leq k$ such that

$$\sigma_k(\boldsymbol{c})_{q,\boldsymbol{u};\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}} \leq C \cdot k^{1/q-1/p},$$

where $\sigma_k(\boldsymbol{c})_{q,\boldsymbol{u};\mathcal{V}}$ is as in (2.4.10) (with $\Lambda = \mathcal{F}$), $\boldsymbol{u}$ is as in (2.4.17) and $C = C(\boldsymbol{b},\varepsilon,p) > 0$ depends on $\boldsymbol{b}$, ε and p only.

(iii) Suppose that $\boldsymbol{b}$ is monotonically nonincreasing. Then, for any $p \leq q < \infty$ and $s \in \mathbb{N}$, there exists an anchored set $S \subset \mathcal{F}$ of size $|S| \leq s$ such that

$$\sigma_s(\boldsymbol{c})_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}} \leq C \cdot s^{1/q-1/p},$$

where $\sigma_s(\boldsymbol{c})_{q;\mathcal{V}}$ is as in (2.4.8) (with $\Lambda = \mathcal{F}$), $\boldsymbol{u}$ is as in (2.4.17) and $C = C(\boldsymbol{b},\varepsilon,p) > 0$ depends on $\boldsymbol{b}$, ε and p only.

Note that Theorem 2.4.13 implies Theorem 2.4.6. This follows from (i) with $q = 2$ or $q = 1$.

Remark 2.4.14 Besides the term $\|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}}$, parts (i) and (iii) can be found in [12, Thm. 3.28] and [12, Thm. 3.33], respectively. As in the finite-dimensional case (see Remark 2.4.11), the main difference is the assertion of the bound on $\|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}}$. This can be

established through similar arguments, using either the majorant ℓ^p space $\ell^p_{\mathsf{M}}(\mathcal{F};\mathcal{V})$ or the anchored ℓ^p space $\ell^p_{\mathsf{A}}(\mathcal{F};\mathcal{V})$ (see, e.g., [12, Defn. 3.31]) and then Stechkin's inequality in lower or anchored sets (see, e.g., [12, Lem. 3.32]). See also [12, Lem. 7.23] (this lemma only considers the scalar-valued case; however the extension to the Banach-valued case is straightforward).

Note that neither [12, Thm. 3.28] nor [12, Thm. 3.33] asserts part (ii) of Theorem 2.4.13. This can be shown via the weighted Stechkin's inequality (see, e.g., [12, Lem. 3.12]), which gives the bound $\sigma_k(\boldsymbol{c})_{q,\boldsymbol{u};\mathcal{V}} \leq \|\boldsymbol{c}\|_{p,\boldsymbol{u};\mathcal{V}} \cdot k^{1/q-1/p}$, and then by showing that $\|\boldsymbol{c}\|_{p,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b},\varepsilon,p)$. This latter fact can be obtained by the straightforward extension of [12, Lem. 7.23] to the Banach-valued setting.

Note that part (iii) of Theorem 2.4.13 assume that $\boldsymbol{b} \in \ell^p(\mathbb{N})$ is monotonically nonincreasing. In our main theorems we consider the weaker assumption that $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\boldsymbol{b} \geq \boldsymbol{0}$. For this we require the following corollary of Theorem 2.4.13.

Corollary 2.4.15. *Let $0 < p < 1$, $\varepsilon > 0$, $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\boldsymbol{b} \geq \boldsymbol{0}$ and $f \in \mathcal{H}(\boldsymbol{b},\varepsilon)$, where $\mathcal{H}(\boldsymbol{b},\varepsilon)$ is as in (2.3.3). Let $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ be the Chebyshev or Legendre coefficients of f, as in (2.4.2). Thenfor any $p \leq q < \infty$ and $s \in \mathbb{N}$, there exists an anchored set $S \subset \mathcal{F}$ of size $|S| \leq s$ such that*

$$\sigma_s(\boldsymbol{c})_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q,\boldsymbol{u};\mathcal{V}} \leq C \cdot s^{1/q-1/p},$$

where $\sigma_s(\boldsymbol{c})_{q;\mathcal{V}}$ is as in (2.4.8) (with $\Lambda = \mathcal{F}$), $\boldsymbol{u}$ is as in (2.4.17) and $C = (\boldsymbol{b},\varepsilon,p) > 0$ depends on $\boldsymbol{b}$, ε and p only.

Proof. Let $\tilde{\boldsymbol{b}}$ be the minimal monotone majorant of $\boldsymbol{b}$, defined in (2.4.15). We first claim that

$$\mathcal{R}(\tilde{\boldsymbol{b}},\varepsilon) \subseteq \mathcal{R}(\boldsymbol{b},\varepsilon), \tag{2.4.19}$$

where $\mathcal{R}$ is as in (2.3.2) with $\varepsilon = 1$. Let $\boldsymbol{\rho} \geq \boldsymbol{1}$ be such that

$$\sum_{j=1}^{\infty} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) \tilde{b}_j \leq \varepsilon.$$

Then, since $\tilde{b}_j \geq b_j$ for all j, we have

$$\sum_{j=1}^{\infty} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) b_j \leq \varepsilon.$$

Thus,

$$\mathcal{R}(\tilde{\boldsymbol{b}},\varepsilon) = \bigcup \left\{ \mathcal{E}(\boldsymbol{\rho}) : \boldsymbol{\rho} \geq \mathbf{1},\ \sum_{j=1}^{\infty} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) \tilde{b}_j \leq \varepsilon \right\}$$
$$\subseteq \bigcup \left\{ \mathcal{E}(\boldsymbol{\rho}) : \boldsymbol{\rho} \geq \mathbf{1},\ \sum_{j=1}^{\infty} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) b_j \leq \varepsilon \right\} = \mathcal{R}(\boldsymbol{b},\varepsilon),$$

as required.

Now let $f \in \mathcal{H}(\boldsymbol{b},\varepsilon)$, i.e., f is holomorphic in $\mathcal{R}(\boldsymbol{b},\varepsilon)$ and satisfies $\|f\|_{L^{\infty}_{\varrho}(\mathcal{R}(\boldsymbol{b},\varepsilon))} \leq 1$. It follows from (2.4.19) that $f \in \mathcal{H}(\tilde{\boldsymbol{b}},\varepsilon)$. Since $\tilde{\boldsymbol{b}}$ is monotonically nonincreasing and ℓ^p-summable, part (iii) of Theorem 2.4.13 now immediately implies the result. □

Remark 2.4.16 (Dependence of the set S on $\boldsymbol{b}$, ε and p only) As stated, the various sets S described in these two results in infinite dimensions depend on the function f being approximated. In fact, an inspection of the proofs of these results (see the references [12, Thm. 3.28] and [12, Thm. 3.33]) reveals that they only depend on $\boldsymbol{b}$ and ε. This holds because the proofs rely on bounds for the polynomial coefficients $c_{\boldsymbol{\nu}}$ that depend on $\boldsymbol{\nu}$, $\boldsymbol{b}$ and ε only. To be more precise, [12, Thm. 3.28] involves establishing the summability of a bounding sequence for the $\mathcal{V}$-norms of the polynomial coefficients in (2.4.2). This bound is equal to $\|f\|_{L^{\infty}(\mathcal{R}(\boldsymbol{b},\varepsilon);\mathcal{V})}$ multiplied by a factor that is independent of f and depending on $\boldsymbol{\rho}$ and ε only. Consequently, the index set in Theorem 2.4.13 part (ii) is independent of f. Likewise, for the index sets in Theorem 2.4.13 part (i) and (iii), the coefficients in [12, Thm. 3.33]) are bounded by a monotonically nonincreasing sequence (see [12, Eq. (3.55)]) that only depends on $\boldsymbol{b}$ and ε. We will use this observation later in the proofs of the main results of Chapter 5 (Theorems 5.3.3 and 5.3.4).

2.4.6 Hyperbolic cross index sets

Consider a smooth function f with expansion (2.4.2) and a polynomial approximation of the form (2.5.2) with a target index set S. As mentioned in §2.3.2, we consider both the known and unknown anisotropy cases. For the former, we may aim to use the knowledge of $\boldsymbol{b}$ to choose a suitable index set S that attains the desired error bounds in §2.4.5. When $\boldsymbol{b}$ is unknown, we do not have knowledge of a suitable index set. Later, we will tackle this issue with compressed sensing techniques. However, to do so, we need to restrict the infinite index set $\mathcal{F}$ to a finite, but large index set Λ. We shall use *hyperbolic cross index sets* for this task.

$$\Lambda = \Lambda^{\mathsf{HC}}_{n,d} = \left\{ \boldsymbol{\nu} = (\nu_k)_{k=1}^{d} \in \mathbb{N}_0^d : \prod_{k=1}^{d} (\nu_k + 1) \leq n \right\} \subset \mathbb{N}_0^d. \tag{2.4.20}$$

We term n the *order* of the hyperbolic cross. Note that it is common to consider (2.4.20) as the hyperbolic cross of order $n-1$. We use n here as it is slightly more convenient for

this work. When defined this way, $\Lambda_{n,d}^{\mathsf{HC}}$ is in fact the union of all lower sets (see Definition 2.4.9) in d dimensions of size at most n (see, e.g., [12, Prop. 2.5]). Therefore, $\Lambda_{n,d}^{\mathsf{HC}}$ is a good candidate for the set Λ, since, by Theorems 2.4.10 and 2.4.13 we know that it contains a lower set S of size n that achieves the stipulated algebraic or exponential error bounds.

In infinite dimensions, we define the following index set

$$\Lambda = \Lambda_n^{\mathsf{HCI}} = \left\{ \boldsymbol{\nu} = (\nu_k)_{k=1}^{\infty} \in \mathcal{F} : \prod_{j=1}^{n} (\nu_k + 1) \leq n,\ \nu_k = 0,\ k > n \right\} \subset \mathcal{F}. \tag{2.4.21}$$

In this case, the union of all anchored sets (Definition 2.4.9) of size at most n in infinite dimensions is a subset of Λ_n^{HCI} (see, e.g., [12, Prop. 2.18]). Note that Λ_n^{HCI} is isomorphic to $\Lambda_{n,n}^{\mathsf{HC}}$ under the restriction map $\boldsymbol{\nu} = (\nu_k)_{k=1}^{\infty} \in \mathcal{F} \mapsto (\nu_k)_{k=1}^{n} \in \mathbb{N}_0^d$. For convenience, we now also define

$$N = \Theta(n, d) = \begin{cases} |\Lambda_{n,d}^{\mathsf{HC}}| & d < \infty, \\ |\Lambda_n^{\mathsf{HCI}}| = |\Lambda_{n,n}^{\mathsf{HC}}| & d = \infty, \end{cases} \tag{2.4.22}$$

as the cardinality of the index set employed. In general, the exact behaviour of $\Theta(n, d)$ is unknown. However, it admits a variety of different bounds. These are summarized as follows for $d < \infty$:

$$N = |\Lambda_{n,d}^{\mathsf{HC}}| \leq \min \left\{ 2n^3 4^d, \mathrm{e} n^{2+\log(d)/\log(2)}, \frac{n(\log(n) + d\log(2))^{d-1}}{(d-1)!} \right\}. \tag{2.4.23}$$

The bounds are based on [60, 169]. See also [12, Lem. B.3–B.5].

2.5 Recovery of orthogonal polynomial coefficients

As noted in §1.5, a main contribution of this thesis is methods and algorithms for computing polynomial approximations to holomorphic, Banach-valued functions from the samples (1.1.1). Having introduced the necessary components, we now describe how this is done. In particular, we explain how the approximation of f via orthogonal polynomials can be reformulated as a problem of recovering a finite vector consisting of its (unknown) coefficients in some suitable multi-index set.

Let $f \in L_{\varrho}^2(\mathcal{U}; \mathcal{V})$ be a function defined everywhere with convergent expansion (2.4.2), $N \in \mathbb{N}$ and $\mathcal{V}^N$ be the vector space of Banach-valued vectors of length N, i.e., $\boldsymbol{v} = (v_i)_{i=1}^N$ where $v_i \in \mathcal{V}$, $i = 1, \ldots, N$. Let $\Lambda \subset \mathcal{F}$ be a finite multi-index set of size $|\Lambda| = N$ with the ordering $\Lambda = \{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_N\}$ and let $\boldsymbol{w} = (w_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda} \in \mathbb{R}^N$, with $\boldsymbol{w} > \boldsymbol{0}$ be a vector of positive weights. In practice, in this thesis, the weights will be chosen as (2.4.17) and Λ as the set (2.4.20) in finite dimensions or (2.4.21) in infinite dimensions, for a suitable choice of n.

Consider $m \in \mathbb{N}$ and let $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in \mathcal{U}$ be sample points. Define the normalized measurement matrix and the measurement and error vectors by

$$\boldsymbol{A} = \left(\frac{\Psi_{\boldsymbol{\nu}_j}(\boldsymbol{y}_i)}{\sqrt{m}} \right)_{i,j=1}^{m,N} \in \mathbb{C}^{m \times N}, \quad \boldsymbol{f} = \frac{1}{\sqrt{m}} \left(f(\boldsymbol{y}_i) + n_i \right)_{i=1}^{m} \quad \text{and} \quad \boldsymbol{e} = \frac{1}{\sqrt{m}} (n_i)_{i=1}^{m} \in \mathcal{V}^m, \tag{2.5.1}$$

where n_i is a measurement error term as in (1.3.1).

We also define the truncated expansion of f based on the index set Λ and its corresponding vector of coefficients as

$$f_\Lambda = \sum_{\boldsymbol{\nu} \in \Lambda} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}, \quad \boldsymbol{c}_\Lambda = (c_{\boldsymbol{\nu}_j})_{j=1}^{N} \in \mathcal{V}^N. \tag{2.5.2}$$

Notice that the matrix $\boldsymbol{A} = (a_{i,j})_{i,j=1}^{m,N}$ immediately extends to a bounded linear operator $\boldsymbol{A} : \mathcal{V}^N \to \mathcal{V}^m$. Specifically, $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ is given by

$$\boldsymbol{x} = (x_i)_{i=1}^{N} \in \mathcal{V}^N \mapsto \boldsymbol{A}\boldsymbol{x} = \left(\sum_{j=1}^{N} a_{i,j} x_j \right)_{i=1}^{m} \in \mathcal{V}^m. \tag{2.5.3}$$

For ease of notation, we make no distinction henceforth between a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times N}$ and the corresponding linear operator in $\mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ (or $\mathcal{B}(\mathcal{V}_K^N, \mathcal{V}_K^m)$). With this in hand, notice that

$$\boldsymbol{A}\boldsymbol{c}_\Lambda = \frac{1}{\sqrt{m}} \left(f_\Lambda(\boldsymbol{y}_i) \right)_{i=1}^{m} = \frac{1}{\sqrt{m}} (f(\boldsymbol{y}_i))_{i=1}^{m} - \frac{1}{\sqrt{m}} \left(f(\boldsymbol{y}_i) - f_\Lambda(\boldsymbol{y}_i) \right)_{i=1}^{m},$$

and therefore

$$\boldsymbol{A}\boldsymbol{c}_\Lambda + \boldsymbol{e} + \widetilde{\boldsymbol{e}} = \boldsymbol{f}, \quad \text{where} \quad \widetilde{\boldsymbol{e}} = \frac{1}{\sqrt{m}} \left(f(\boldsymbol{y}_i) - f_\Lambda(\boldsymbol{y}_i) \right)_{i=1}^{m}. \tag{2.5.4}$$

Therefore, we have formulated the recovery of $\boldsymbol{c}_\Lambda$ (and consequently $f_\Lambda \approx f$) as the solution of a noisy linear system (2.5.4), where the noise term $\boldsymbol{e} + \widetilde{\boldsymbol{e}}$ encompasses both the noise $\boldsymbol{e} = (n_i)_{i=1}^{m}/\sqrt{m}$ in the sample values and the error $\widetilde{\boldsymbol{e}}$ due to the truncation (2.5.2) of the infinite expansion (2.4.2) via the index set Λ. Thus, a key goal in Chapters 3–4 is to develop methods and algorithms that take m sample values $(f(\boldsymbol{y}_i) + n_i)_{i=1}^{m}$ and construct approximations to $\boldsymbol{c}_\Lambda \in \mathcal{V}^N$ from the linear system (2.5.4). Note that, in this thesis, (2.5.4) is an underdetermined linear system with m equations and $N > m$ unknowns. Which means that, in general, the exact recovery of the coefficients $\boldsymbol{c}_\Lambda \in \mathcal{V}^N$ is not feasible since it has infinitely many solutions. This motivates the use of sparsity and techniques from compressed sensing, where we exploit the fact that c_Λ is an (approximately) sparse vector.

2.5.1 Finite-dimensional approximation

In this thesis, we aim to construct approximations using methods and algorithms that produce outputs in finite computational time. Note that the coefficients $\boldsymbol{c}_\Lambda$ belong to the space $\mathcal{V}^N$, which is generally a Hilbert or Banach space. In order to compute an approximation, we consider $\{\varphi_K\}_{k=1}^K$ as a basis for $\mathcal{V}_K$. Consequently, the approximation to f must belong to a finite-dimensional subspace $\mathcal{V}_K \subset \mathcal{V}$. The technical aspects of this process are discussed in more detail in Chapter 4. In summary, based on the discussion in §1.4, the approximation to f we aim to construct is given by

$$\hat{f}_\Lambda = \sum_{\boldsymbol{\nu}\in\Lambda} \hat{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}, \quad \hat{\boldsymbol{c}}_\Lambda = (\hat{c}_{\boldsymbol{\nu}_j})_{j=1}^N \in \mathcal{V}_K^N, \tag{2.5.5}$$

where

$$\hat{c}_{\boldsymbol{\nu}_i} = \sum_{k=1}^K \hat{c}_{i,k} \varphi_k,$$

with $(\hat{c}_{i,k})_{k,j=1}^{N,K} \in \mathbb{C}^{N\times K}$. Therefore, the goal is to construct approximations $\hat{\boldsymbol{c}}_\Lambda \in \mathcal{V}_K^N$ to $\boldsymbol{c}_\Lambda \in \mathcal{V}^N$ from the linear system (2.5.4).

2.5.2 Unknown anisotropy recovery

We now consider the unknown anisotropy setting as mentioned in §2.3.2.

In view of the best s-term approximation error bounds shown in §2.4.2–2.4.4, we expect the vector of coefficients $\boldsymbol{c}_\Lambda \in \mathcal{V}^N$ to be *approximately sparse*. In other words, $\boldsymbol{c}_\Lambda$, should be well approximated by its s largest coefficients. However, in view of the theoretical approximation results shown in §2.4.5, we also expect $\boldsymbol{c}_\Lambda$ to be well approximated by a subset of s coefficients whose indices define a lower or anchored set.

In classical compressed sensing, one exploits sparse structure via minimizing an ℓ^1-norm. To exploit sparse and lower or anchored structure, we follow ideas of [2, 4, 12, 64] and use a weighted ℓ^1-norm penalty. To be more precise, based on the discussion in §2.5.1, we consider minimizing an objective function recovering a vector $\boldsymbol{z} \in \mathcal{V}_K^N$ with a data fidelity term $\|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}$ and a sparsity-promoting term $\|\boldsymbol{z}\|_{1,\boldsymbol{w};\mathcal{V}}$. Minimizing the weighted ℓ^1-norm has some geometrical properties to promote sparse solutions. In addition, it can be viewed as a convex relaxation of minimizing an ℓ^0 type of problem, which is nonconvex and generally NP-hard to solve [112, Chpt. 2]. See [19, §5.4] for further details. Specifically, we compute an approximate solution via the Banach-valued, *weighted Square-Root LASSO (SR-LASSO)* optimization problem [3, 40, 247]

$$\min_{\boldsymbol{z}\in\mathcal{V}_K^N} \mathcal{G}(\boldsymbol{z}), \qquad \mathcal{G}(\boldsymbol{z}) := \lambda\|\boldsymbol{z}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}. \tag{2.5.6}$$

Here $\lambda > 0$ is a tuning parameter and $\boldsymbol{A}$ and $\boldsymbol{f}$ are as in (2.5.1). This parameter balances the trade-off between the fidelity term and the sparsity promoting term [12, §6.2.4]. Note that if $\hat{\boldsymbol{c}}_\Lambda$ is a solution of (2.5.6) then we define the approximation to f as

$$\hat{f}_\Lambda = \sum_{\boldsymbol{\nu}\in\Lambda} \hat{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}.$$

Note that $\hat{f}_\Lambda$ is a solution to the problem

$$\min_{p\in P_{S;\mathcal{V}_K}} \lambda \mathcal{J}(p) + \sqrt{\frac{1}{m}\sum_{i=1}^{m} \|d_i - p(\boldsymbol{y}_i)\|_{\mathcal{V}}^2}, \tag{2.5.7}$$

where $d_i = f(\boldsymbol{y}_i) + n_i$ is as in (1.3.1), $\mathcal{P}_{\Lambda;\mathcal{V}_K}$ is as in (2.4.3) and $\mathcal{J} : P_{S;\mathcal{V}_K} \to \mathbb{R}_+$ is a norm over the coefficients of p given by

$$\mathcal{J}(p) = \|\boldsymbol{c}\|_{1,\boldsymbol{w};\mathcal{V}}.$$

Moreover, the coefficients $\hat{\boldsymbol{c}}_\Lambda$ of any solution $\hat{f}$ to (2.5.7) also solve (2.5.6). Thus, the two problems are equivalent.

Remark 2.5.1 As an alternative to solving (2.5.6), we could use a formulation based on a constrained basis pursuit or unconstrained LASSO problem. However, we consider the SR-LASSO problem (2.5.6) instead. While other approaches are arguably more common, based on [3] the SR-LASSO has the desirable property that the optimal values of its hyper-parameter λ is independent of the noise term (in this case $\boldsymbol{e} + \boldsymbol{e}'$). This is not the case for other formulations, whose hyperparameters need to be chosen in terms of the (unknown) magnitude of the noise in order to ensure good theoretical and practical performance (see, e.g., [19, Chpt. 6]). This is particularly problematic in the setting of function approximation, where such terms are function dependent (for instance, the term $\boldsymbol{e}'$ depends on the expansion tail $f - f_\Lambda$) and therefore generally unknown. See [3] and [12, §6.6] for further discussion.

2.5.3 Known anisotropy recovery

In the previous case, we assumed that the coefficients were approximately sparse, but, due to the unknown anisotropy, we do not know which coefficients are the most significant. In the results for the best s-term approximation §2.4.3 the set S is independent of f (see Remark 2.4.12 and Remark 2.4.16). Thus we choose the set $\Lambda \subset \mathcal{F}$ as a large set in which we expect these significant coefficients to lie. Conversely, in the known anisotropy setting where $\boldsymbol{b}$ is known, we also know a suitable set $S \subset \mathcal{F}$ of size $|S| = s$, due to the results in Theorems 2.4.13.

Analogously as before, we define the normalized measurement matrix by

$$\boldsymbol{A} = \left(\frac{\Psi_{\boldsymbol{\nu}_j}(\boldsymbol{y}_i)}{\sqrt{m}} \right)_{i,j=1}^{m,s} \in \mathbb{C}^{m \times s}, \tag{2.5.8}$$

where $\{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_s\}$ is an ordering of S. Likewise, we truncate the expansion of f and its vector coefficients based on (2.5.2) for the index set S. In contrast to the previous case, where $m < N$, we now assume that $m \geq s$. Hence, there is no need for compressed sensing techniques. Instead, we formulate the vector recovery problem as the Banach-valued minimization problem

$$\min_{\boldsymbol{z} \in \mathcal{V}_K^s} \mathcal{G}(\boldsymbol{z}), \qquad \mathcal{G}(\boldsymbol{z}) := \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}. \tag{2.5.9}$$

As before, if $\hat{\boldsymbol{c}}_S$ is a solution of (2.5.9) then we define the approximation to f as

$$\hat{f}_S = \sum_{\boldsymbol{\nu} \in S} \hat{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}.$$

Note that $\hat{f}_S$ is a solution to the problem

$$\min_{p \in P_{S;\mathcal{V}_K}} \sqrt{\frac{1}{m} \sum_{i=1}^{m} \|d_i - p(\boldsymbol{y}_i)\|_{\mathcal{V}}^2}, \tag{2.5.10}$$

where $d_i = f(\boldsymbol{y}_i) + n_i$ is as in (1.3.1), $\mathcal{P}_{S;\mathcal{V}_K}$ is as in (2.4.3). As before, the coefficients $\hat{\boldsymbol{c}}_S$ of any solution $\hat{f}$ to (2.5.10) also solve (2.5.9). Thus, the two problems are equivalent.

To end this section, it is worth mentioning here that mathematically, the known anisotropy setting is just a particular case of the unknown anisotropy setting with $\lambda = 0$ and Λ and N replaced by S and s, respectively.

2.6 Deep learning

Similar to past works in approximation theory [7, 16, 82, 84, 88, 141, 213, 215, 236], the main theoretical results about DL in this thesis are proved by drawing a connection between DNNs and the recovery of polynomials coefficients. In Chapter 5 use DNNs to emulate polynomial approximation via least squares (in the known anisotropy case) as in (2.5.9) and compressed sensing (in the unknown anisotropy case) as in (2.5.6).

In this section, we define the DL setup considered in this work. It is important to note that machine learning, particularly DL, is a broad subject that we will not attempt to review beyond what is necessary for this thesis.

2.6.1 Deep neural networks

In general terms, the objective of the DL framework used in this thesis is to approximate a certain function $g : \mathbb{R}^n \to \mathbb{R}^K$ by constructing a mapping that uses the available data $(\boldsymbol{z}_i, g(\boldsymbol{z}_i))_{i=1}^m \subset \mathbb{R}^n \times \mathbb{R}^K$ to find a function $\Phi : \mathbb{R}^n \to \mathbb{R}^K$ that is able to generalize on new points $\boldsymbol{z} \in \mathbb{R}^n$ and to obtain good approximations on the training data $g(\boldsymbol{z}_i) \in \mathbb{R}^K$ for $i = 1, \dots, m$. To train the DNN it is standard to use a loss function $\mathcal{L} : \mathcal{N} \to \mathbb{R}$, where $\mathcal{N}$ is a certain family of DNNs.

Before introducing DNNs, we need further notation and setup.

Definition 2.6.1 (affine maps)**.** An *affine linear map* $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^p$ is an operator that can be written as $\mathcal{A}(\boldsymbol{z}) = \boldsymbol{W}\boldsymbol{z} + \boldsymbol{b}$, where $\boldsymbol{W} \in \mathbb{R}^{p \times n}$ is the weight matrix and $\boldsymbol{b} \in \mathbb{R}^p$ is the bias vector.

In the following, $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ denotes an activation function on vectors $\boldsymbol{z} \in \mathbb{R}^n$. In this work, we consider either the Rectified Linear Unit (ReLU)

$$\sigma_1(z) := \max\{0, z\},$$

Exponential Linear Unit (ELU)

$$\sigma(z) = \begin{cases} z & z > 0, \\ e^z - 1 & z \leq 0, \end{cases}$$

Rectified Polynomial Unit (RePU)

$$\sigma_\ell(z) := \max\{0, z\}^\ell, \qquad \ell = 2, 3, \dots$$

or hyperbolic tangent (tanh)

$$\sigma_0(z) = \frac{\mathrm{e}^z - \mathrm{e}^{-z}}{\mathrm{e}^z + \mathrm{e}^{-z}}$$

activation function. Keeping these concepts in mind, we define a DNN.

Definition 2.6.2 (feedforward DNNs)**.** Let $D \in \mathbb{N}_0$ and $N_0, \dots N_{D+2} \in \mathbb{N}$. Consider affine maps $\mathcal{A}_\ell : \mathbb{R}^{N_\ell} \to \mathbb{R}^{N_{\ell+1}}$ given by $\mathcal{A}_\ell(\boldsymbol{z}) = \boldsymbol{W}_\ell \boldsymbol{z} + \boldsymbol{b}_\ell$ and an activation function σ which we assume acts componentwise, i.e., $\sigma(\boldsymbol{z}) := (\sigma(z_i))_{i=1}^n$ for $\boldsymbol{z} = (z_i)_{i=1}^n$. Then a *DNN is a map* $\Phi : \mathbb{R}^{N_0} \to \mathbb{R}^{N_{D+2}}$ *given by*

$$\Phi : \mathbb{R}^{N_0} \to \mathbb{R}^{N_{D+2}},\ \boldsymbol{z} \mapsto \Phi(\boldsymbol{z}) = \mathcal{A}_{D+1}(\sigma(\mathcal{A}_D(\sigma(\cdots \sigma(\mathcal{A}_0(\boldsymbol{z})) \cdots)))). \tag{2.6.1}$$

The values $\{N_l\}_{l=1}^{D+1}$ are the widths of the hidden layers. For convenience, we define

$$\mathrm{width}(\mathcal{N}) = \max\{N_1, \dots, N_{D+1}\}, \qquad \mathrm{depth}(\mathcal{N}) = D,$$

where $\mathcal{N}$ is a class of DNNs with a fixed architecture (i.e., fixed activation function, depth and widths).

2.6.2 Recovery of coefficients via DNNs

Recall that we consider functions of the form $f : \mathcal{U} \to \mathcal{V}$, where $\mathcal{U} = [-1,1]^d$, $d \in \mathbb{N}$ (or $d = \mathbb{N}$) and noisy samples $d_i = f(\boldsymbol{y}_i) + n_i$ is as in (1.3.1). As in the previous case §2.5.1, we must consider approximations of $f(\boldsymbol{y})$ in the finite dimensional space $\mathcal{V}_K$. Using the basis $\{\varphi_k\}_{k=1}^K$ we can rewrite this as

$$f(\boldsymbol{y}) \approx \sum_{k=1}^{K} c_k(\boldsymbol{y})\varphi_k.$$

Here we focus on DNNs approximating these coefficient functions $c_k : \mathcal{U} \to \mathbb{R}$. Note that a DNN, as defined in Definition 2.6.2, has finite domain $n \in \mathbb{N}$ whereas the coefficient may have $d = \mathbb{N}$. For now, to keep notation simple we assume $d < \infty$. In §5.1.1 we introduce a variable restriction operator (see (5.1.3)) to solve this issue.

In this thesis we aim to construct a DNN $\Phi : \mathbb{R}^d \to \mathbb{R}^K$ as in Definition 2.6.2 that approximates f from the data $(d_i)_{i=1}^m \in \mathcal{V}^m$ by minimizing an objective function $\mathcal{G} : \mathcal{N} \to \mathbb{R}$

$$\min_{\Phi \in \mathcal{N}} \mathcal{G}(\Phi).$$

We primarily choose $\mathcal{G}$ as

$$\mathcal{G}(\Phi) := \sqrt{\frac{1}{m}\sum_{i=1}^{m} \|f_\Phi(\boldsymbol{y}_i) - d_i\|_{\mathcal{V}}^2} + \mathcal{J}(\Phi), \tag{2.6.2}$$

where $\mathcal{J} : \mathcal{N} \to \mathbb{R}$ is a function promoting sparsity or some other desirable feature and

$$f_\Phi(\boldsymbol{y}) = \sum_{k=1}^{K} \Phi(\boldsymbol{y})_k \varphi_k, \quad \forall \boldsymbol{y} \in \mathcal{U}. \tag{2.6.3}$$

Note that, if $\hat{\Phi} : \mathbb{R}^d \to \mathbb{R}^K$ is a solution to the problem (2.6.2), then we define the approximation to f as $f_{\hat{\Phi}}$ as in (2.6.3).

2.7 Main sources of error

The remainder of this thesis focuses on the approximation to holomorphic functions based on either polynomials, as discussed in §2.5, or DNNs as in §2.6. A key objective in the subsequent chapters is to establish error bounds for the various methods introduced. We measure this error in the $L^2_\varrho(\mathcal{U};\mathcal{V})$- and $L^\infty(\mathcal{U};\mathcal{V})$-Bochner norms. Specifically, we shall

derive bounds of the form

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim E_{\mathsf{app}} + m^{\theta_1} \left(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right),$$

in the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm and

$$\|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim m^{\theta_2} \left(E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right),$$

in the $L^\infty(\mathcal{U};\mathcal{V})$-norm. Here $\theta_1, \theta_2 \geq 0$ are a small algebraic factor (often $\theta_1 = 0$) and the other terms are the main sources of error in the problem. We now discuss these error sources.

(i) E_{app} is the *approximation error*. Depending on the specific setup, it decays algebraically or exponentially with respect to the samples m (up to some log terms). It is equivalent to the corresponding decay rate for the best s-term approximation error, which depends on the smoothness of f (see §2.3.3).

(ii) E_{disc} is a *physical discretization error*. It accounts for the fact that we cannot typically perform computations in $\mathcal{V}$, since it is an infinite-dimensional function space (see §1.4). Instead, we perform computations in some reduced dimension finite-dimensional subspace $\mathcal{V}_K \subset \mathcal{V}$, e.g., a FE space in the case of parametric PDEs (see §2.2).

(iii) E_{samp} is the *sampling error*. It is equal to $\sqrt{\frac{1}{m}\sum_{i=1}^m \|n_i\|_{\mathcal{V}}^2}$, and it accounts for the measurement errors $(n_i)_i$. In other words, this means that the approximation based on polynomials or DNNs are robust to noise in the samples (1.3.1).

(iv) E_{opt} is the *optimization error*. It accounts for the fact that the problems (2.5.6) and (2.5.9) are never solved exactly, but only within some tolerance. This error depends on how close we are to minimizing the objective function $\mathcal{G}$ in §2.5.

In the following chapters, we will delve into a detailed analysis of function approximation, encompassing approximation errors, physical discretization errors, sampling errors, and optimization errors. We will study how these errors manifest in different contexts and discuss strategies for mitigating them to improve the overall performance and reliability of function approximation methods answering Questions 1–9 of §1.6. Specifically, the term E_{app} plays a crucial role in the answers to Question 1 for methods using orthogonal polynomials, Question 3 in the context of algorithms approximating Hilbert-valued functions of infiniterly many variables and Question 5 for DNNs approximating high-dimensional Banach-valued functions from limited samples. Furthermore, we investigate whether the rates obtained for this approximation error are optimal, addressing Question 8. We also investigate how other errors, including E_{disc}, E_{samp} and E_{opt} affect the accuracy of the approximation, while answering Question 9.

Chapter 3

Compressed sensing for near-best polynomial approximation from limited samples

This chapter focuses on methods used to compute approximations of Hilbert-valued functions based on a finite set of sample values. We begin in §3.1 with various preliminaries. We recall key notation and present the problem statement in §3.1.2. In §3.2 we describe the main constributions of this chapter. Next, in §3.3, we state our main results on the existence of methods approximating smooth Hilbert-valued functions. We provide a discussion on these results in §3.4. In §3.5 we recap the main setup and describe the methods from the main results. Next, in §3.6 we present extensions of relevant results of compressed sensing that will be use later in the proofs. In §3.7 we use these to present three general Theorems from which we derive the main results in this chapter. In §3.8 we preent the proofs of the main results. Finally, in §3.9 we write our conclusions and address Question 1 of §1.6, before outlining some future work in §3.10.

The content of this chapter is primarily derived from [10].

3.1 Preliminaries

Broadly speaking, in the contex of Chapter 3 a method is a map that takes a finite input and computes a polynomial approximation to a certain function of interest f.

Our main results in this chapter consider $(\mathcal{V}, \langle \cdot, \cdot \rangle_{\mathcal{V}})$ to be a Hilbert space and assume holomorphy (see §2.3) of the underlying function with respect to the parameter space $\mathcal{U}$ in order to attain the desired rates in §2.4.3. Extensions to Banach spaces will be considered in Chpt 5 in the context of DL. We assume no *a priori* knowledge of the region of holomorphy (see §2.3.2), concentrating solely on the unknown anisotropy case. If such information is available, least-squares methods can be applied as in §2.5.3 in a straightforward manner to compute an approximation. Note that we assume holomorphy in order to provide concrete

algebraic and exponential rates of approximation. The methods introduced in this chapter exist independently of the smoothness assumption. Specifically, they can be applied to nonholomorphic functions, such as those with finite smoothness, although without the same theoretical guarantees.

3.1.1 Setup

We now describe the setup considered in this chapter. Let $\mathcal{U} = [-1,1]^d$, where $d \in \mathbb{N}$ or $d = \infty$. Let ϱ be either the uniform or Chebyshev (arcsine) measure and consider the associated tensor-product Legendre or Chebyshev polynomials. Now let $f : \mathcal{U} \to \mathcal{V}$ be the function we seek to approximate, which we now assume is continuous. Draw m sample points $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ i.i.d. from ϱ and let

$$d_i = f(\boldsymbol{y}_i) + n_i, \quad i = 1, \dots, m, \tag{3.1.1}$$

be m noisy samples of f, where $\boldsymbol{n} = (n_i)_{i=1}^m \in \mathcal{V}^m$ accounts for errors in the measurements.

The discrete space $\mathcal{V}_K$

Recall the discussion of finite-dimensional discretizations from §2.2. We now make the assumption that the given data (3.1.1) belongs to a finite-dimensional space $\mathcal{V}_K \subseteq \mathcal{V}$. Consider a basis $\{\varphi_k\}_{k=1}^K$ for $\mathcal{V}_K$. We assume that the computation that evaluates $f(\boldsymbol{y}_i)$ produces the coefficients of the sample values d_i in this basis. This is a natural assumption to make. For example, for a function that arises as the solution of a DE as in (1.2.1), typically these are the coefficients associated with a basis derived from a FEM. Therefore, we now write the sample values as

$$d_i = \ f(\boldsymbol{y}_i) + n_i = \sum_{k=1}^K d_{i,k}\varphi_k, \quad i = 1, \dots, m, \tag{3.1.2}$$

and consider the values $d_{i,k} \in \mathbb{C}$ as the *data* we obtain by sampling f (See (iii) in §2.7). Recall from §2.2 that we assume the existence of bounded linear operator $\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K$. Note that, in this chapter $\mathcal{V}$ is a Hilbert space and, by the discussion in Remark 2.2.1, $\mathcal{P}_K$ is the orthogonal projector from $\mathcal{V}$ onto $\mathcal{V}_K$, where $\mathcal{P}_K(f)(\boldsymbol{y}) = \mathcal{P}_K(f(\boldsymbol{y}))$ (see (2.2.15)) when f is defined everywhere.

3.1.2 Problem statement

We now formally define the input, the output of the method and the problem statement.

Definition 3.1.1 (Input)**.** The *input* of the mapping is the collection of sample points $(\boldsymbol{y}_i)_{i=1}^m$ and the array of mK values $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ defined by (3.1.2).

We now define the output. To this end, we first fix a multi-index set $\Lambda \subset \mathcal{F}$ of size $|\Lambda| = N$ for some $N \geq 1$. This set defines a polynomial space (see (2.4.3))

$$\mathcal{P}_{\Lambda;\mathcal{V}} = \left\{ \sum_{\boldsymbol{\nu}\in\Lambda} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}} : c_{\boldsymbol{\nu}} \in \mathcal{V} \right\} \subset L^2_\varrho(\mathcal{U};\mathcal{V}),$$

within which we shall construct the resulting polynomial approximation, where $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$ form an orthonormal basis of $L^2_\varrho(\mathcal{U})$ as in (2.4.1).

Definition 3.1.2 (Output)**.** The *output* of the method are the coefficients $(\hat{c}_{j,k})_{j,k=1}^{N,K} \in \mathbb{C}^{N\times K}$ such that the approximation $\hat{f} \in \mathcal{P}_{\Lambda;\mathcal{V}_K}$ is given by

$$\hat{f} : \boldsymbol{y} \mapsto \sum_{j=1}^{N} \left(\sum_{k=1}^{K} \hat{c}_{j,k} \varphi_k \right) \Psi_{\boldsymbol{\nu}_j}(\boldsymbol{y}), \tag{3.1.3}$$

where $\hat{c}_{j,k} \in \mathbb{C}$ for $j \in [N], k \in [K]$ and $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N$ is some indexing of the multi-indices in Λ, and $\Psi_{\boldsymbol{\nu}_j}$ is as in (2.4.1) for $j \in [N]$.

Definition 3.1.3 (Methods for polynomial approximation of Hilbert-valued functions)**.** Let $\Lambda \subset \mathcal{F}$ of size $|\Lambda| = N$ be given, along with an indexing $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N$ of the multi-indices in Λ. A *method for polynomial approximation of Hilbert-valued functions from sample values* is a map of the form

$$\mathcal{M} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K}, \ \left((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K} \right) \mapsto (\hat{c}_{j,k})_{j,k=1}^{N,K}. \tag{3.1.4}$$

With this in hand, the formal problem we study in this chapter is: *provided m sample values as in (3.1.2), represented as mK input values (as in Definition 3.1.1), construct methods as in Definition 3.1.3 that compute NK coefficients (as defined in Definition 3.1.2) of a polynomial approximation $\hat{f}$ to f, while providing guarantees on the error $f - \hat{f}$ in the $L^2_\varrho(\mathcal{U};\mathcal{V})$- and $L^\infty(\mathcal{U};\mathcal{V})$-norms.*

Remark 3.1.4 As formulated above, it is up to the user to choose a suitable multi-index set Λ in Definition 3.1.3. Fortunately, as we see in our main results below, this multi-index set is given simply and explicitly in terms of m and another parameter ϵ (a failure probability). In particular, no 'oracle' knowledge of the function being approximated is required. Thus, one can also make the stronger assertion in what follows in which the mapping takes the same input, but outputs both the desired index set Λ and the polynomial coefficients. For ease of presentation, we shall not do this.

When $d = \infty$ each sample point $\boldsymbol{y}_i$ is an infinite sequence of real numbers. It is implicit in Definition 3.1.3 that the mapping only accesses finitely-many entries of this sequence. This does not cause any problems. As noted, the polynomial approximation is obtained in

the index set Λ, which is a finite subset of $\mathcal{F}$. Hence, the multi-indices in Λ are nonzero only in their first n entries, for some n. Therefore, it is only necessary to access the first n entries of each sequence $\boldsymbol{y}_i$. More concretely, in our main results below, the polynomial approximation in infinite dimensions is obtained in a multi-index set Λ in which only the first n terms can be nonzero, where n is an integer given explicitly in terms of m and ϵ.

3.2 Contributions

Our main contribution is three theorems providing algebraic decay rates (finite- and infinite-dimensional case) and exponential decay rates (finite dimensional case) for methods, in the sense of Definition 3.1.3, approximating holomorphic functions f, where $\mathcal{V}$ is a Hilbert space. In each theorem, we construct methods approximating f to within an explicit error bound with high probability. Specifically, our error bounds take the form

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}},$$

and in the $L^\infty(\mathcal{U};\mathcal{V})$-norm

$$\|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \sqrt{\frac{m}{L}} \left(E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}\right),$$

where $\hat{f}$ is an approximation to f, $L = L(m,\varepsilon)$ is a (poly)logarithmic factor in m (see (3.3.1)), and the terms E_{app}, E_{disc} and E_{samp} are as in §2.7.

There are several distinguishing features of our analysis that we now highlight:

1. We overcome the curse of dimensionality in the approximation error. The term E_{app} decays algebraically fast in m/L. Specifically, when $f \in \mathcal{B}(\boldsymbol{\rho})$ with $\mathcal{B}(\boldsymbol{\rho})$ as in (2.3.2), is holomorphic in $\mathcal{E}_{\boldsymbol{\rho}}$ for an arbitrary $\boldsymbol{\rho}$ in finite dimensions or $f \in \mathcal{H}(\boldsymbol{b},\varepsilon)$ with $\mathcal{H}(\boldsymbol{b},\varepsilon)$ as in (2.3.3) is $(\boldsymbol{b},\varepsilon)$-holomorphic in infinite dimensions, for some $0 < p < 1$, we have

$$E_{\mathsf{app}} \lesssim C \cdot \left(\frac{m}{L}\right)^{1/2-1/p},$$

where C depends on d, p and $\boldsymbol{\rho}$ (finite dimensional case) or $\boldsymbol{b}$, ε and p (infinite-dimensional case).

2. We achieve exponential decay rates in the approximation error in finite dimensions. Let $f \in \mathcal{B}(\boldsymbol{\rho})$ with $\mathcal{B}(\boldsymbol{\rho})$ as in (2.3.2) be holomorphic in $\mathcal{E}_{\boldsymbol{\rho}}$ for an arbitrary $\boldsymbol{\rho}$. The term E_{app} decays exponentially fast in m/L. That is, for some $\gamma > 0$ and a constant

$c_0 > 0$ to be specified,

$$E_{\mathsf{app}} \lesssim C \cdot \begin{cases} \exp\left(-\frac{\gamma}{2}\left(\frac{m}{c_0 L}\right)^{\frac{1}{d}}\right) & \text{Exponential rate, Chebyshev,} \\ \exp\left(-\gamma\left(\frac{m}{c_0 L}\right)^{\frac{1}{2d}}\right) & \text{Exponential rate, Legendre,} \end{cases} \tag{3.2.1}$$

where C depends on d, γ and $\boldsymbol{\rho}$.

3. The exponential decay rates obtained in the approximation error are uniform guarantees (see §1.4). This means that a single draw of the sample points is sufficient for recovering any function in $\mathcal{B}(\boldsymbol{\rho})$ with high probability. In contrast, we only achieve the desired algebraic rate with high probability for each fixed f, which makes our algebraic rate results nonuniform.

3.3 Main results

We now present the main results of this chapter. In addition to the above discussion, our contribution are methods achieving the rates of the best s-term approximation with respect to the number of samples m (recall Theorems 2.4.10–2.4.13). These results employ specific choices of the index set Λ in order to obtain the desired approximation rates. See §2.4.6 for further details. Specifically, in finite dimensions, recall that the *hyperbolic cross* index set is the set $\Lambda_{n,d}^{\mathsf{HC}}$ defined in (2.4.20). In infinite dimensions, we use the index set Λ_n^{HCI} defined in (2.4.21). Let $N = \Theta(n,d)$ (see (2.4.22)) be the cardinality of the index set employed. Recall that the exact behaviour of $\Theta(n,d)$ is unknown, but it admits the bounds in (2.4.23).

Given $m \geq 3$ and $\epsilon \in (0,1)$, we define

$$L = L(m,d,\epsilon) = \begin{cases} \log^2(m) \cdot \min\{\log(m) + d, \log(2d) \cdot \log(m)\} + \log(\epsilon^{-1}) & d < \infty, \\ \log^4(m) + \log(\epsilon^{-1}) & d = \infty. \end{cases} \tag{3.3.1}$$

Algebraic rates of convergence, finite dimensions

Theorem 3.3.1 (Existence of a method; algebraic case, finite dimensions)**.** *Let $d \in \mathbb{N}$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathbb{N}_0^d} \subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis and $\{\varphi_k\}_{k=1}^K$ be a basis for $\mathcal{V}_K$. Then for every $m \geq 3$, $0 < \epsilon < 1$ and $K \geq 1$, there is a mapping*

$$\mathcal{M} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

where $N = \Theta(n,d)$ is as in (2.4.22) with $n = \lceil m/L \rceil$ and $L = L(m,d,\epsilon)$ as in (3.3.1), with the following property. Let $f \in \mathcal{B}(\boldsymbol{\rho})$ for arbitrary $\boldsymbol{\rho} \geq \mathbf{1}$, draw $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ randomly and independently according to ϱ and let $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ be as in (3.1.2) for arbitrary

noise terms $\boldsymbol{n} = (n_i)_{i=1}^m \in \mathcal{V}$. *Let* $(\hat{c}_{j,k})_{j,k=1}^{N,K} = \mathcal{M}((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$ *and define the approximation* $\hat{f}$ *as in* (3.1.3) *based on the index set* $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$. *Then the following holds with probability at least* $1 - \epsilon$. *The error satisfies*

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \zeta, \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \zeta, \tag{3.3.2}$$

for any $0 < p \leq 1$, *where*

$$\zeta := C \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}, \tag{3.3.3}$$

$c_0, c_1, c_2 \geq 1$ *are universal constants and* $C = C(d, p, \boldsymbol{\rho})$ *depends on* d, p *and* $\boldsymbol{\rho}$ *only.*

We now make several remarks about this result. The same remarks apply (with obvious modifications) to all subsequent results as well. First, notice how the index set Λ in which the approximation is constructed is given completely explicitly in terms of m, d and ϵ. Thus, as claimed in Remark 3.1.4, no 'oracle' information about the function being approximated is required. Indeed, notice that the mapping described in this theorem is *universal* in the sense that its applies equally to *any* function $f \in \mathcal{B}(\boldsymbol{\rho})$ and *any* $\boldsymbol{\rho} \geq \mathbf{1}$.

A key aspect of this theorem is the factor ζ, defined in (3.3.3), which determines the error bounds (3.3.2). As claimed in §2.7, this incorporates three main key errors arising in the approximation process:

(i) *The approximation error.* This is the algebraically-decaying term $E_{\mathsf{app}} = C \cdot (m/(c_0 L))^{1/2-1/p}$. It is completely equivalent to the best s-term approximation error bound in Theorem 2.4.5, except with s replaced by $m/(c_0 L)$.

(ii) *The sampling error.* This is the term $E_{\mathsf{samp}} = \|\boldsymbol{n}\|_{2;\mathcal{V}}/\sqrt{m}$, where $\boldsymbol{n} = (n_i)_{i=1}^m$ is as in (3.1.2). In other words, the effect of any errors in computing the sample values $f(\boldsymbol{y}_i)$ enters linearly in the overall error bound.

(iii) *The physical discretization error.* This is the term $E_{\mathsf{disc}} = \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}$. It describes the effect of working in the finite-dimensional subspace $\mathcal{V}_K$, instead of the full space $\mathcal{V}$. Critically, it depends on the orthogonal projection (best approximation) $\mathcal{P}_K(f)$ of f from $\mathcal{V}_K$.

Notice that (i) also describes the *sample complexity* of the scheme. Indeed, Theorem 3.3.1 asserts that there is a polynomial approximation that can be obtained from m samples that attains the best s-term rate $s^{1/2-1/p}$, where $s = m/(c_0 L)$ scales like m up to the polylogarithmic factor L.

Algebraic rates of convergence, infinite dimensions

We now consider algebraic rates of convergence in the infinite-dimensional setting. In this case, we assume that f belongs to the class $\mathcal{H}(\boldsymbol{b}, \varepsilon)$, where $\boldsymbol{b} \in \ell^p(\mathbb{N})$ and $\varepsilon > 0$ (see §2.3.1).

Theorem 3.3.2 (Existence of a method; algebraic case, infinite dimensions)**.** *Let $d = \infty$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}} \subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis and $\{\varphi_k\}_{k=1}^K$ be a basis for $\mathcal{V}_K$. Then for every $m \geq 3$, $0 < \epsilon < 1$ and $K \geq 1$, there is a mapping*

$$\mathcal{M} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

where $N = \Theta(n, d)$ is as in (2.4.22) with $n = \lceil m/L \rceil$, where $L = L(m, d, \epsilon)$ is as in (3.3.1), with the following property. Let $\varepsilon > 0$, $0 < p < 1$ and $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\boldsymbol{b} \geq \boldsymbol{0}$. Let $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$, draw $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ randomly and independently according to ϱ and let $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ be as in (3.1.2) for arbitrary noise terms $\boldsymbol{n} = (n_i)_{i=1}^m \in \mathcal{V}$. Let $(\hat{c}_{j,k})_{j,k=1}^{N,K} = \mathcal{M}((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$ and define the approximation $\hat{f}$ as in (3.1.3) based on the index set $\Lambda = \Lambda_n^{\mathsf{HCI}}$. Then the following holds with probability at least $1 - \epsilon$. The error satisfies

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \zeta, \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \zeta, \tag{3.3.4}$$

where

$$\zeta := C \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}, \tag{3.3.5}$$

$c_0, c_1, c_2 \geq 1$ are universal constants and $C = C(\boldsymbol{b}, \varepsilon, p)$ depends on $\boldsymbol{b}$, ε and p only.

We now make a few remarks about the algebraic rates discussed in this chapter. Theorems 3.3.1 and 3.3.2 are *nonuniform* and achieve the corresponding algebraic rates for a fixed function f. Specifically, a single draw of the sample points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ is sufficient to recover a fixed function f with high probability up to the specified error bound (see §1.4).

Note that in finite dimensions, we attain an algebraic rate that scales like m up to a polylogarithmic factor of the order $\mathcal{O}(\log^3(m))$ in terms of m. Conversely, in the infinite-dimensional case, it is of the order of $\mathcal{O}(\log^4(m))$. This rate is not equivalent to that of the best s-term rates in §2.4.3, specifically (2.4.12). This discrepancy arises from the necessity for accurate and stable recovery methods based on orthonormal polynomials (see §3.6). Specifically, these logarithmic factors arise when asserting the wRIP for the measurement matrix. See Lemma 3.7.1. Additionally, unlike the finite-dimensional case, the polylogarithmic factor $\log^3(m)$ becomes $\log^4(m)$. This change is due to the factor d in L from (3.3.1) becoming $\log(m)$.

As mentioned in Theorem 3.3.1, the method in finite dimensions is more general in the sense that it applies to any function $f \in \boldsymbol{\mathcal{B}}(\boldsymbol{\rho})$ and any $\boldsymbol{\rho} \geq \boldsymbol{1}$. In contrast, in infinite dimensions, it applies to a specific class of anisotropic holomorphic functions in $\mathcal{H}(\boldsymbol{b}, \varepsilon)$. Specifically, Theorem 3.3.2 involve the assumption $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$. Unfortunately, as we will see

in Chapter 6, it is impossible to learn infinite-dimensional functions for which we only know that $\boldsymbol{b} \in \ell^p(\mathbb{N})$.

Exponential rates of convergence, finite dimensions

Finally, we consider exponential rates of convergence in finite dimensions.

Theorem 3.3.3 (Existence of a method; exponential case, finite dimensions)**.** *Let $d \in \mathbb{N}$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathbb{N}_0^d} \subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis and $\{\varphi_k\}_{k=1}^K$ be a basis for $\mathcal{V}_K$. Then for every $m \geq 3$, $0 < \epsilon < 1$ and $K \geq 1$, there is a mapping*

$$\mathcal{M} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

where $N = \Theta(n, d)$ is as in (2.4.22) with

$$n = \begin{cases} \lceil \sqrt{m/L} \rceil & \text{Legendre,} \\ \lceil m/(2^d L) \rceil & \text{Chebyshev,} \end{cases} \tag{3.3.6}$$

and L as in (3.3.1), with the following property. Draw $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ randomly and independently according to ϱ. Then, with probability at least $1 - \epsilon$, the following holds. Let $f \in \mathcal{B}(\boldsymbol{\rho})$ for arbitrary $\boldsymbol{\rho} \geq \mathbf{1}$, $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ be as in (3.1.2) for arbitrary noise terms $\boldsymbol{n} = (n_i)_{i=1}^m \in \mathcal{V}$. Let $(\hat{c}_{j,k})_{j,k=1}^{N,K} = \mathcal{M}((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$ and define the approximation $\hat{f}$ as in (3.1.3) based on the index set $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$. Then the error satisfies

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \zeta, \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \zeta, \tag{3.3.7}$$

for any

$$0 < \gamma < (d+1)^{-1} \left(d! \prod_{j=1}^d \log(\rho_j) \right)^{1/d},$$

where

$$\zeta := C \cdot \begin{cases} \exp\left(-\frac{\gamma}{2}\left(\frac{m}{c_0 L}\right)^{\frac{1}{d}}\right) & \text{Chebyshev} \\ \exp\left(-\gamma\left(\frac{m}{c_0 L}\right)^{\frac{1}{2d}}\right) & \text{Legendre} \end{cases} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}, \tag{3.3.8}$$

$c_0, c_1, c_2 \geq 1$ are universal constants and $C = C(d, \gamma, \boldsymbol{\rho})$ depends on d, γ and $\boldsymbol{\rho}$ only.

A notable difference between Theorem 3.3.3 and Theorems 3.3.1 and 3.3.2 is that it provides different decay rates for the Chebyshev or uniform measures. For points drawn from the Chebyshev measure, the exponential approximation rate is better and takes the form $\exp\left(-\frac{\gamma}{2}\left(\frac{m}{c_0 L}\right)^{\frac{1}{d}}\right)$, having a more favourable exponent of $1/d$ on the $m/(c_0 L)$ term

than the exponent $1/(2d)$ for points drawn from the uniform measure. This difference reflects the fact that uniformly-distributed sample points are relatively poor points for polynomial approximation in finite (and, in particular, low) dimensions, whereas points drawn from the Chebyshev measure are much better. It is notable that this difference vanishes when considering algebraic rates. Thus in higher, in particular, infinite dimensions we expect uniformly-distributed points to perform well. This phenomenon has been investigated in [6]

Theorem 3.3.3 asserts that there are methods in which the approximation error decays exponentially fast in terms of the number of samples m for finite-dimensional function approximation. However, the curse of dimensionality is not avoided in the decay rate, making this result less desirable in high dimensions. This brings a key advantage for using algebraic rates. Here, we pay the price of not having exponential decay rates but avoid the curse of dimensionality on the approximation error. Note that the curse of dimensionality is not avoided in the constant $C(d, p, \boldsymbol{\rho})$ in Theorem 3.3.1. However, by virtue of the infinite-dimensional setting, there is no curse of dimensionality in Theorem 3.3.2

Remark 3.3.4 In the algebraic case, in order to obtain the desired algebraic exponent $1/2 - 1/p$ we bound a term in (3.7.9) with high probability for each fixed f. See Step 4 of the proof of Theorem 3.7.3. This renders the ensuing result nonuniform. Conversely, in the exponential case (where the appearance of small algebraic factors is not a concern, since they can be absorbed into the exponentially-decaying term) we bound this term with probability one for *any* f. See Step 4 of the proof of Theorem 3.7.5. Note that one could also derive uniform guarantees in the algebraic case by considering a fixed value of p and letting $\mathcal{M}$ and $\mathcal{A}$ depend on p, or by considering a restricted range $0 < p \leq p^* < 1$. Both strategies involve a larger value of n, with its size depending on p or p^*. See [12, §7.6.2] for further discussion.

3.4 Discussion

This chapter bridges a gap between the best s-term polynomial approximation theory and the practical scenario of computing such approximations from sample values. In particular, it asserts that algebraic and exponential rates with respect to the number of samples m that are highly similar to those of the best approximation.

Our main results assume holomorphy of the underlying function in order to attain these rates. However, they assume no a priori knowledge of the region of holomorphy. As discussed in §2.5.3, if such information is available, then least-squares methods can be applied in a straightforward manner to compute an approximation. The holomorphy assumption is made in order to have concrete algebraic and exponential rates.

Our analysis is based on compressed sensing theory and involve computing approximate minimizers of certain weighted ℓ^1 -minimization problems. Here we highlight some key contributions of the proofs:

1. We provide precise error rates for polynomial approximation via compressed sensing. Most prior work on compressed sensing involves quantifying the sample complexity to obtain a certain (weighted) best approximation error. Subject to a holomorphy assumption, we use this to obtain specific algebraic and exponential rates.

2. Prior works consider polynomial approximations formed by exact minimizers of nonlinear optimization problems. Our main results in this chapter do not involve minimizers. However, in §3.7.3 we consider error bounds for inexact minimizers, which will be useful in Chapter 4.

3. Most prior works on compressed sensing (with the exception of [95]) focus on scalar-valued functions, e.g., quantities of interest of parametric DEs. We show results about mappings that work in the Hilbert-valued setting, and, crucially, provide error bounds that take into account discretization error.

4. Following [2, 4, 12, 64, 225, 227], we work in a weighted setting in order to promote sparsity in lower or anchored sets (recall §2.4.4).

5. Finally, as discussed below, we consider *noise-blind* decoders (see Remark 2.5.1, and also [3]).

The optimization problem

Our approach first formulates the approximation problem as the recovery of a finite, Hilbert-valued vector (i.e., an element of $\mathcal{V}^N$) via a weighted Square-Root LASSO (SR-LASSO) optimization problem. The use of SR-LASSO, as opposed to the classical LASSO or various constrained formulations, is crucial to this work. SR-LASSO is *noise-blind* (see Remark 2.5.1), meaning it admits an optimal parameter choice that is independent of the noise $\boldsymbol{e}$ and $\tilde{\boldsymbol{e}}$ in (2.5.4). Consequently, it gives rise to methods that do not require any a priori (and generally unavailable) estimates of the measurement error $(n_i)_{i=1}^m$ or the truncation error with respect to the finite polynomial space in which the approximation is constructed. See §2.5.2.

The index set and universality

Notice how the index set Λ in which the approximation is constructed is given completely explicitly in terms of m, d and ϵ. Thus, as claimed in Remark 3.1.4, no 'oracle' information about the function being approximated is required. Indeed, notice that the methods described in this theorem are universal, in the sense that they apply equally to *any* function $f \in \mathcal{B}(\boldsymbol{\rho})$ and *any* $\boldsymbol{\rho} \geq \mathbf{1}$ (in finite dimensions) or any $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$ and any $\boldsymbol{b}, \varepsilon$ (in infinite dimensions).

The sample complexity

A key aspect of this theorem is the factor ζ, which determines the error bounds (3.3.2) (3.3.4) and (3.3.7). As discussed in §2.7, ζ incorporates three main errors arising in the approximation process. Notably, the approximation error also describes the *sample complexity* of the scheme. Specifically, Theorem 3.3.1 asserts that a polynomial approximation can be obtained from m samples, achieving the s-term rate $s^{1/2-1/p}$, where $s = m/(c_0 L)$ scales like m up to the polylogarithmic factor L. This is also true for Theorem 3.3.2. Moreover, the exponential rates in the approximation error in Theorem 3.3.3 are comparable to the rates in (2.4.13) of the form $\exp(-\gamma s^{1/d})$ with γ defined in (2.4.14).

3.5 The methods and proofs setup

Before diving into the details of the proofs, we now recap our main setup and important aspects for the rest of the chapter. Consider a high-dimensional function $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$ with expansion (2.4.2). We follow the setup from §2.5. More specifically §2.5.2. Here the solution to the linear system in (2.5.4) recovers the polynomial coefficients $\boldsymbol{c}_\Lambda \in \mathcal{V}_K^N$ of the truncated expansion (2.5.2) of f from m sample values. Keeping this in mind, our methods are based on the solution to the minimization problem (2.5.6) recovering the coefficients in (2.5.4).

3.5.1 The methods in Theorems 3.3.1– 3.3.3

In the following we describe more about the methods in Theorems 3.3.1– 3.3.3. These are described in Table 3.1. In particular, the output of the method is defined by $\widehat{\boldsymbol{C}} = (\hat{c}_{i,k})_{i,k=1}^{N,K} \in \mathbb{C}^{N\times K}$ where each coefficient $\hat{c}_{i,k}$ is defined by the relation

$$\hat{c}_{\boldsymbol{\nu}_i} = \sum_{k=1}^{K} \hat{c}_{i,k}\varphi_k \in \mathcal{V}_K, \quad i \in [N],$$

where $\{\varphi_k\}_{k=1}^K$ is a basis of $\mathcal{V}_K \subset \mathcal{V}$. Note that these are indeed well-defined methods, since the minimizer of (2.5.6) with smallest ℓ^2-norm is unique (this follows from the facts that (2.5.6) is a convex problem, therefore its set of minimizers is a convex set, and the function $\boldsymbol{z} \mapsto \|\boldsymbol{z}\|_{2,\mathcal{V}}^2$ is strongly convex). This particular choice is arbitrary, and is made solely so as to have a well-defined method. It is of no consequence whatsoever. Indeed, the various error bounds we prove later hold for any minimizer of (2.5.6).

Having defined the methods, the following section provides the main theory to prove Theorems 3.3.1– 3.3.3.

3.6 Compressed sensing

While the construction of the methods asserted in our main results in this chapter is based on classical techniques from compressed sensing [12,19,112], in the following section we extend

- Let m, ϵ and n be as given in the particular theorem and set $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$ (Theorem 3.3.1 and 3.3.3) or $\Lambda = \Lambda_n^{\mathsf{HCI}}$ (Theorem 3.3.2).
- Set $\lambda = (4\sqrt{m/L})^{-1}$, where $L = L(m, d, \epsilon)$ is as in (3.3.1).
- Let $\boldsymbol{D} = (d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ and $\boldsymbol{Y} = (\boldsymbol{y}_i)_{i=1}^m$ be an input, as in (3.1.2), and set $\boldsymbol{F} = \frac{1}{\sqrt{m}}\boldsymbol{D}$.
- Let $\boldsymbol{A}$ and $\boldsymbol{w}$ be as in (2.5.1) and (2.4.17), respectively.
- Define the output $\widehat{\boldsymbol{C}} = \mathcal{M}(\boldsymbol{Y}, \boldsymbol{D})$ as the minimizer of (2.5.6) with smallest ℓ^2-norm.

Table 3.1: The methods $\mathcal{M} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K}$ used in Theorems 3.3.1, 3.3.2 and 3.3.3

relevant aspects of the theory of compressed sensing to the Hilbert- and Banach-valued setting, and use these to obtain various error bounds and sample complexity estimates that lead to the main theorems in not only this chapter, but subsequent results in this thesis.

3.6.1 The weighted robust Null Space Property

Although the results in this chapter pertain to Hilbert-valued functions, the theory presented in this section can be applied to the more general setting where $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ is a Banach space. Henceforth, $\mathcal{V}$ will be assumed to be a Banach space unless stated otherwise.

As described in §2.5.2, we consider weighted ℓ^1-minimization type of problems to recover the polynomial coefficients of an expansion of f of the form (2.4.2). Let $\boldsymbol{A} \in \mathcal{B}(\mathcal{W}^N, \mathcal{W}^m)$ be the measurement matrix as in (2.5.1), where $\mathcal{W} \subseteq \mathcal{V}$ is a closed subspace. The weighted robust Null Space Property (wrNSP) is a property on the matrix $\boldsymbol{A}$ that, if satisfied, ensures accurate and stable recovery. It implies certain distance bounds in the $\ell^1_{\boldsymbol{w}}$- and ℓ^2-norms. We now formally define this property. See, e.g., [12, Defn. 6.22] or [227, §4.1].

Definition 3.6.1. The matrix $\boldsymbol{A}$ has the *weighted robust Null Space Property (rNSP) over* $\mathcal{W}$ *of order* $(k, \boldsymbol{w})$ *with constants* $0 \leq \rho < 1$ *and* $\gamma \geq 0$ if

$$\|\boldsymbol{x}_S\|_{2;\mathcal{V}} \leq \frac{\rho\|\boldsymbol{x}_{S^c}\|_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \gamma\|\boldsymbol{A}\boldsymbol{x}\|_{2;\mathcal{V}}, \quad \forall \boldsymbol{x} \in \mathcal{W}^N,$$

for any $S \subseteq [N]$ with $|S|_{\boldsymbol{w}} \leq k$.

Recall from §2.1 that $\boldsymbol{x}_S$ is the vector with ith entry equal to x_i if $i \in S$ and $|S|_{\boldsymbol{w}}$ is the weighted cardinality of set S.

The following lemma is standard in the scalar case (see, e.g., [12, Lem. 6.24]). We omit the proof of its extension to the Banach-valued case, since it follows almost exactly the same arguments.

Lemma 3.6.2 (Weighted rNSP implies $\ell^1_{\boldsymbol{w}}$ and ℓ^2 distance bounds)**.** *Suppose that $\boldsymbol{A} \in \mathcal{B}(\mathcal{W}^N, \mathcal{W}^m)$ has the weighted rNSP over a closed subspace $\mathcal{W} \subseteq \mathcal{V}$ of order $(k, \boldsymbol{w})$ with constants $0 < \rho < 1$ and $\gamma > 0$. Let $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{W}^N$. Then*

$$\|\boldsymbol{z} - \boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} \leq C_1 \left(2\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{z}\|_{1,\boldsymbol{w};\mathcal{V}} - \|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}}\right) + C_2\sqrt{k}\|\boldsymbol{A}(\boldsymbol{z} - \boldsymbol{x})\|_{2;\mathcal{V}}, \quad (3.6.1)$$

$$\|\boldsymbol{z} - \boldsymbol{x}\|_{2;\mathcal{V}} \leq \frac{C_1'}{\sqrt{k}} \left(2\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{z}\|_{1,\boldsymbol{w};\mathcal{V}} - \|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}}\right) + C_2'\|\boldsymbol{A}(\boldsymbol{z} - \boldsymbol{x})\|_{2;\mathcal{V}}, \quad (3.6.2)$$

where the constants are given by

$$C_1 = \frac{(1+\rho)}{(1-\rho)}, \quad C_2 = \frac{2\gamma}{(1-\rho)}, \quad C_1' = \left(\frac{(1+\rho)^2}{1-\rho}\right) \quad \text{and} \quad C_2' = \left(\frac{(3+\rho)\gamma}{1-\rho}\right).$$

The previous result implies bounds for certain approximate minimizers of the optimization problems introduced in §2.5. In particular, consider the optimization problem in (2.5.6). We say that $\boldsymbol{x} \in \mathcal{V}_K^N$ is an inexact minimizer of (2.5.6), for some $\zeta > 0$, if

$$\mathcal{G}(\boldsymbol{x}) \leq \zeta + \min_{\boldsymbol{z} \in \mathcal{V}_K^N} \mathcal{G}(\boldsymbol{z}).$$

Lemma 3.6.3 (Weighted rNSP implies error bounds for inexact minimizers)**.** *Suppose that $\boldsymbol{A} \in \mathcal{B}(\mathcal{W}^N, \mathcal{W}^m)$ has the weighted rNSP over a closed subspace $\mathcal{W} \subseteq \mathcal{V}$ of order $(k, \boldsymbol{w})$ with constants $0 \leq \rho < 1$ and $\gamma > 0$. Let $\boldsymbol{x} \in \mathcal{W}^N$, $\boldsymbol{f} \in \mathcal{V}^m$ and $\boldsymbol{e} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{f} \in \mathcal{V}^m$, and consider the minimization problem*

$$\min_{\boldsymbol{z} \in \mathcal{W}^N} \mathcal{G}(\boldsymbol{z}), \qquad \mathcal{G}(\boldsymbol{z}) := \lambda\|\boldsymbol{z}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}, \quad (3.6.3)$$

with parameter

$$0 < \lambda \leq \frac{(1+\rho)^2}{(3+\rho)\gamma} k^{-1/2}. \quad (3.6.4)$$

Then

$$\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} \leq C_1 \left(2\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \frac{\mathcal{G}(\tilde{\boldsymbol{x}}) - \mathcal{G}(\boldsymbol{x})}{\lambda}\right) + \left(\frac{C_1}{\lambda} + C_2\sqrt{k}\right)\|\boldsymbol{e}\|_{2;\mathcal{V}},$$
$$\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_{2;\mathcal{V}} \leq \frac{C_1'}{\sqrt{k}} \left(2\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \frac{\mathcal{G}(\tilde{\boldsymbol{x}}) - \mathcal{G}(\boldsymbol{x})}{\lambda}\right) + \left(\frac{C_1'}{\sqrt{k}\lambda} + C_2'\right)\|\boldsymbol{e}\|_{2;\mathcal{V}},$$

for any $\tilde{\boldsymbol{x}} \in \mathcal{W}^N$,where the constants are given by

$$C_1 = \frac{(1+\rho)}{(1-\rho)}, \quad C_2 = \frac{2\gamma}{(1-\rho)}, \quad C_1' = \left(\frac{(1+\rho)^2}{1-\rho}\right) \quad \text{and} \quad C_2' = \left(\frac{(3+\rho)\gamma}{1-\rho}\right).$$

Proof. First notice that $C_1'/C_2' \leq C_1/C_2$ since $0 < \rho < 1$, where C_1, C_2, C_1' and C_2' are as in Lemma 3.6.2. Hence the condition on λ implies that

$$\lambda \leq \min\{C_1/C_2, C_1'/C_2'\}k^{-1/2}, \tag{3.6.5}$$

Using this lemma and this bound, we deduce that

$$\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} \leq 2C_1\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \frac{C_1}{\lambda}\left(\lambda\|\tilde{\boldsymbol{x}}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{f}\|_{2;\mathcal{V}} - \lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}}\right) + C_2\sqrt{K}\|\boldsymbol{e}\|_{2;\mathcal{V}}.$$

The definition of $\mathcal{G}$ in (3.6.3) gives

$$\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} \leq 2C_1\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \frac{C_1}{\lambda}\left(\mathcal{G}(\tilde{\boldsymbol{x}}) - \mathcal{G}(\boldsymbol{x}) + \|\boldsymbol{e}\|_{2;\mathcal{V}}\right) + C_2\sqrt{k}\|\boldsymbol{e}\|_{2;\mathcal{V}},$$

which is the first result. The second follows in an analogous manner. □

This result provides information about the error bound for inexact minimizers in terms of the best k-term approximation error, an optimization error $\mathcal{G}(\tilde{\boldsymbol{x}}) - \mathcal{G}(\boldsymbol{x})$ and error in the measurement errors. Notice that defining the weighted rNSP over a closed subspace $\mathcal{W} \subseteq \mathcal{V}$ allows us to assert error bounds for the minimization problem over, for example, the space $\mathcal{W} = \mathcal{V}_K$. This fact will be useful later in the proofs.

3.6.2 Matrices satisfying the weighted rNSP over Banach spaces

In this section, we give explicit conditions in terms of m under which the measurement matrix (2.5.1) satisfy the weighted rNSP over Banach spaces. It is well known that showing the (weighted) rNSP directly can be difficult. In the classical, scalar setting, this is overcome by showing that the (weighted) rNSP is implied by the so-called (weighted) restricted isometry property (wRIP). Keeping in mind the definition of weighted sparsity and the set of weighted sparse vectors $\Sigma_{k,\boldsymbol{w}}$ from Definition 2.4.3, we now introduced the wRIP and describe its relation to the (weighted) rNSP.

Definition 3.6.4. Let $\boldsymbol{w} > \boldsymbol{0}$ and $k > 0$. A bounded linear operator $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the *weighted Restricted Isometry Property (wRIP)* over $\mathcal{V}$ of order $(k, \boldsymbol{w})$ if there exists a constant $0 < \delta < 1$ such that

$$(1-\delta)\|\boldsymbol{z}\|_{2;\mathcal{V}}^2 \leq \|\boldsymbol{A}\boldsymbol{z}\|_{2;\mathcal{V}}^2 \leq (1+\delta)\|\boldsymbol{z}\|_{2;\mathcal{V}}^2, \quad \forall \boldsymbol{z} \in \Sigma_{k,\boldsymbol{w}} \subseteq \mathcal{V}^N. \tag{3.6.6}$$

The smallest constant such that this property holds is called the $(k, \boldsymbol{w})$th *weighted Restricted Isometry Constant (wRIC)* of $\boldsymbol{A}$, and is denoted as $\delta_{k,\boldsymbol{w}}$. See, e.g., [19, Def. 6.25] and [227, §4.2].

wRIP implies weighted rNSP

The following result shows that the wRIP is a sufficient condition for the weighted rNSP. This result is well known in the scalar-valued case (see, e.g., [12, Theorem 6.26]).

Lemma 3.6.5 (wRIP implies the weighted rNSP)**.** *Let $\boldsymbol{w} > \boldsymbol{0}$, $k > 0$ and suppose that $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ has the wRIP over $\mathbb{C}$ of order $(2k, \boldsymbol{w})$ with constant $\delta_{2k,\boldsymbol{w}} < (2\sqrt{2}-1)/7$. Then $\boldsymbol{A}$ has the weighted rNSP of order $(k, \boldsymbol{w})$ over $\mathcal{V}$ with constants $\rho = 2\sqrt{2}\delta_{2k,\boldsymbol{w}}/(1-\delta_{2k,\boldsymbol{w}})$ and $\gamma = \sqrt{1+\delta_{2k,\boldsymbol{w}}}/(1-\delta_{2k,\boldsymbol{w}})$.*

However, in this thesis, rather than the classical setting of a vector in $\mathbb{C}^N$, one seeks to recover a Hilbert- or Banach-valued vector in $\mathcal{V}^N$. The former, due to their approximation properties and inner product has a simpler analysis. Below we prove an equivalence between the scalar wRIP over $\mathbb{C}$ and the Hilbert-valued wRIP over $\mathcal{V}$. While for the latter, we only consider sufficient conditions to show the weighted rNSP over Banach spaces.

The Hilbert-valued case

In the Hilbert-valued case by showing the wRIP over $\mathbb{C}$ (or $\mathcal{V}$) we are able to show the wrNSP over the Hilbert space $\mathcal{V}$. The following result shows the equivalence between the scalar wRIP over $\mathbb{C}$ and the Hilbert-valued wRIP over $\mathcal{V}$.

Lemma 3.6.6 (wRIP over $\mathbb{C}$ is equivalent to the wRIP over $\mathcal{V}$)**.** *Let $\boldsymbol{w} > \boldsymbol{0}$, $k > 0$ and $\boldsymbol{A} = (a_{i,j})_{i,j=1}^{m,N} \in \mathbb{C}^{m\times N}$ be a matrix. Then $\boldsymbol{A}$ satisfies the wRIP over $\mathbb{C}$ of order $(k, \boldsymbol{w})$ with constant $0 < \delta < 1$ if and only if the corresponding bounded linear operator $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ defined by*

$$\boldsymbol{x} = (x_i)_{i=1}^N \in \mathcal{V}^N \mapsto \boldsymbol{A}\boldsymbol{x} := \left(\sum_{i=1}^N a_{i,j}x_j\right)_{i=1}^m \in \mathcal{V}^m,$$

satisfies the wRIP over $\mathcal{V}$ of order $(k, \boldsymbol{w})$ with the same constant δ.

Proof. We follow similar arguments to [95, Rmk. 3.5]. First, we rewrite the equivalence as follows:

$$(1-\delta)\|\boldsymbol{x}\|_{2;\mathcal{V}}^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_{2;\mathcal{V}}^2 \leq (1+\delta)\|\boldsymbol{x}\|_{2;\mathcal{V}}^2, \qquad \forall \boldsymbol{x} \in \mathcal{V}^N, |\mathrm{supp}(\boldsymbol{x})|_{\boldsymbol{w}} \leq k, \tag{3.6.7}$$

if and only if

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1+\delta)\|\boldsymbol{x}\|_2^2, \qquad \forall \boldsymbol{x} \in \mathbb{C}^N, |\mathrm{supp}(\boldsymbol{x})|_{\boldsymbol{w}} \leq k. \tag{3.6.8}$$

Suppose that (3.6.8) holds. Let $\boldsymbol{x} = (x_j)_{i=1}^N \in \mathcal{V}^N$ be $(k, \boldsymbol{w})$-sparse and $\{\phi_i\}_i$ be an orthonormal basis of $\mathcal{V}$. Then, for each $i \in [N]$, $x_i \in \mathcal{V}$ can be uniquely represented as

$$x_i = \sum_j \alpha_{i,j}\phi_j, \quad \alpha_{i,j} \in \mathbb{C}.$$

Let $\boldsymbol{x}_j = (\alpha_{i,j})_{i=1}^N \in \mathbb{C}^N$. Then $\operatorname{supp}(\boldsymbol{x}_j) \subseteq \operatorname{supp}(\boldsymbol{x})$ and therefore $\boldsymbol{x}_j$ is $(k, \boldsymbol{w})$-sparse. Hence (3.6.8) gives

$$(1-\delta)\|\boldsymbol{x}_j\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}_j\|_2^2 \leq (1+\delta)\|\boldsymbol{x}_j\|_2^2. \tag{3.6.9}$$

Now observe that

$$\sum_j \|\boldsymbol{x}_j\|_2^2 = \sum_{i=1}^N \sum_j |\alpha_{i,j}|^2 = \sum_{i=1}^N \|x_i\|_{\mathcal{V}}^2 = \|\boldsymbol{x}\|_{2;\mathcal{V}}^2,$$

and

$$\sum_j \|\boldsymbol{A}\boldsymbol{x}_j\|_2^2 = \sum_j \sum_{i=1}^m \left| \sum_{k=1}^N a_{i,k}\alpha_{kj} \right|^2 = \sum_{i=1}^m \left\| \sum_{k=1}^N a_{i,k} x_k \right\|_{\mathcal{V}}^2 = \|\boldsymbol{A}\boldsymbol{x}\|_{2;\mathcal{V}}^2.$$

Summing (3.6.9) over j, we deduce that (3.6.7) holds.

Conversely, suppose that (3.6.7) holds and let $\boldsymbol{z} = (z_i)_{i=1}^N \in \mathbb{C}^N$ with $|\operatorname{supp}(\boldsymbol{z})|_{\boldsymbol{w}} \leq k$. Define $\boldsymbol{x} = (z_i \phi_i)_{i=1}^N \in \mathcal{V}^N$ and notice that $\|\boldsymbol{x}\|_{2;\mathcal{V}} = \|\boldsymbol{z}\|_2$ and $\|\boldsymbol{A}\boldsymbol{x}\|_{2;\mathcal{V}} = \|\boldsymbol{A}\boldsymbol{z}\|_2$. Since $\operatorname{supp}(\boldsymbol{x}) = \operatorname{supp}(\boldsymbol{z})$ and $|\operatorname{supp}(\boldsymbol{z})|_{\boldsymbol{w}} \leq k$, we now apply (3.6.7) to deduce that $(1-\delta)\|\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \leq (1+\delta)\|\boldsymbol{z}\|_2^2$. We conclude that (3.6.8) holds. □

wRIP over Banach spaces

Let $\boldsymbol{A} \in \mathbb{C}^{m \times N}$ be a matrix of the form (2.5.1). In the Banach-valued case, by showing that $\boldsymbol{A}$ has the wRIP over $\mathbb{C}$ we are able to show that also has the wrNSP over the Banach space $\mathcal{V}$ at the cost of an extra factor $\sqrt{m}$ in the error bound (see Chp. 5). We now derive explicit conditions for such a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times N}$ to give rise to an associated operator $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ (see (2.5.3)) that satisfies the weighted rNSP over $\mathcal{V}^N$ with $\mathcal{V}$ a Banach space. Note that as an alternative to the equivalence in Lemma 3.6.7 between wRIP over $\mathbb{C}$ and Hilbert spaces, the following lemma shows that the wRIP over $\mathbb{C}$ is a sufficient condition for the wRIP over Banach spaces.

Lemma 3.6.7 (Weighted rNSP over $\mathbb{C}$ implies weighted rNSP over $\mathcal{V}$)**.** *Suppose that a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times N}$ satisfies the weighted rNSP over $\mathbb{C}$ of order $(k, \boldsymbol{w})$ with $0 \leq \rho < 1$ and $\gamma \geq 0$, and let $s^* = s^*(k) := \max\{|S| : |S|_{\boldsymbol{w}} \leq k,\ S \subseteq [N]\}$. Then the corresponding operator $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ defined by (2.5.3) satisfies the weighted rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{w})$ with constants $0 \leq \rho' < 1$ and $\gamma' > 0$ given by $\rho' = \sqrt{s^*}\rho$ and $\gamma' = \sqrt{s^*}\gamma$, respectively.*

Proof. Let $\boldsymbol{v} \in \mathcal{V}^N$ and $|S|_{\boldsymbol{w}} \leq k$. Using (2.2.5) we get

$$\|\boldsymbol{v}_S\|_{2;\mathcal{V}}^2 = \sum_{i \in S} \|v_i\|_{\mathcal{V}}^2 = \sum_{i \in S} \left(\max_{\substack{v^* \in \mathcal{V}^* \\ \|v^*\|_{\mathcal{V}^*}=1}} |v^*(v_i)| \right)^2 \leq |S| \left(\max_{\substack{v^* \in \mathcal{V}^* \\ \|v^*\|_{\mathcal{V}^*}=1}} \|v^*(\boldsymbol{v}_S)\|_2 \right)^2, \tag{3.6.10}$$

where $v^*(\boldsymbol{v}_S) := (v^*(v_i))_{i=1}^N \in \mathbb{C}^N$. Since $\boldsymbol{A}$ has the weighted rNSP over $\mathbb{C}$ we get

$$\|v^*(\boldsymbol{v}_S)\|_2 \leq \frac{\rho\|v^*(\boldsymbol{v}_{S^c})\|_{1,\boldsymbol{w}}}{\sqrt{k}} + \gamma\|\boldsymbol{A}(v^*(\boldsymbol{v}))\|_2, \quad \forall \boldsymbol{v} \in \mathcal{V}^N, \forall v^* \in \mathcal{V}^*. \tag{3.6.11}$$

If $\|v^*\|_{\mathcal{V}^*} = 1$, then we also have

$$\|v^*(\boldsymbol{v}_{S^c})\|_{1,\boldsymbol{w}} \leq \|\boldsymbol{v}_{S^c}\|_{1,\boldsymbol{w};\mathcal{V}} \quad \text{and} \quad \|\boldsymbol{A}(v^*(\boldsymbol{v}))\|_2 = \|v^*(\boldsymbol{A}(\boldsymbol{v}))\|_2 \leq \|\boldsymbol{A}(\boldsymbol{v})\|_{2;\mathcal{V}}.$$

Therefore

$$\|v^*(\boldsymbol{v}_S)\|_2 \leq \frac{\rho\|\boldsymbol{v}_{S^c}\|_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \gamma\|\boldsymbol{A}(\boldsymbol{v})\|_{2;\mathcal{V}}, \quad \forall \boldsymbol{v} \in \mathcal{V}^N,\ \forall v^* \in \mathcal{V}^*, \|v^*\|_{\mathcal{V}^*} = 1.$$

Combining this with (3.6.10) and noting that $|S| \leq s^*(k)$, we deduce that

$$\|\boldsymbol{v}_S\|_{2;\mathcal{V}} \leq \sqrt{s^*(k)}\left(\frac{\rho\|\boldsymbol{v}_{S^c}\|_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \gamma\|\boldsymbol{A}(\boldsymbol{v})\|_{2;\mathcal{V}}\right), \quad \forall \boldsymbol{v} \in \mathcal{V}^N,$$

as required. □

As mentioned before Lemma 3.6.7, the extra factor $\sqrt{s^*}$ is one of the causes of the extra m-dependent factors–the other being the absence of Parseval's identity in Banach spaces–in the final error bound for the Banach-valued case as opposed to the Hilbert-valued case. See for instance, the discussion about Theorem 5.3.1 and 5.3.2 in §5.4.

3.7 Error bounds for polynomial approximation via weighted SR-LASSO

Having developed the necessary tools for compressed sensing in the Hilbert-valued setting, we now specialize to the case of polynomial approximation via the Hilbert-valued, weighted SR-LASSO problem (2.5.6). Our main results in this section, Theorems 3.7.3–3.7.5, yield error bounds for (inexact) minimizers of this problem in terms of the best polynomial approximation error, the Hilbert space discretization error and the noise.

3.7.1 Overview

This section involve a number of technical steps, we now give a brief overview of how these proofs proceed.

We begin by focusing on the polynomial approximation problem. We first give a sufficient condition in terms of m for the measurement matrix (2.5.1) to satisfy the wRIP with high probability (Lemma 3.7.1). Next, we state and prove three general results (Theorems 3.7.3–3.7.5) that give error bounds for polynomial approximations obtained as inexact minimizers

of the Hilbert-valued, weighted SR-LASSO problem. These results are then ready to be divided into the three cases considered in our main results(Theorems 3.3.1– 3.3.3), i.e., the algebraic and finite-dimensional case, the algebraic and infinite-dimensional case, and the exponential case. We finally prove the main results in §3.8.

3.7.2 The wRIP for the polynomial approximation problem

Let $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}} \subset L^2_\varrho(\mathcal{U})$ be either the tensor Chebyshev or Legendre polynomial basis,

$$\Lambda = \begin{cases} \Lambda^{\mathsf{HC}}_{n,d} & d < \infty, \\ \Lambda^{\mathsf{HCI}}_{n} & d = \infty, \end{cases} \tag{3.7.1}$$

be the hyperbolic cross index set and draw $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ independently and identically from the measure ϱ. Then we define the measurement matrix $\boldsymbol{A}$ exactly as in (2.5.1).

We now assert conditions on m under which the measurement matrix (2.5.1) satisfies the wRIP. For this, we use the following result, which is an adaptation of [51, Thm. 2.14] (we make several small notational changes herein for consistency with the notation used in this thesis; moreover, we replace the logarithmic factor $\log^2(k/\delta^2)$ with $\log^2(k/\delta)$, as revealed by an inspection of the proof).

Lemma 3.7.1 (wRIP for Chebyshev and Legendre polynomials)**.** *Let ϱ be the tensor-product uniform or Chebyshev measure on $\mathcal{U} = [-1,1]^d$ with $d \in \mathbb{N}$ or $d = \infty$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$ be the corresponding tensor-product orthonormal Legendre or Chebyshev polynomial basis of $L^2_\varrho(\mathcal{U})$, Λ be as in (3.7.1) for some $n \geq 1$ and $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ be drawn independently and identically from the measure ϱ. Let c_0 be a universal constant, $0 < \delta, \epsilon < 1$ and $k \geq 1$,*

$$L' = L'(k,n,d,\epsilon) := \begin{cases} \log^2(k/\delta) \cdot \min\{\log(n) + d, \log(2d) \cdot \log(2n)\} + \log(2/\epsilon) & d < \infty, \\ \log^2(k/\delta) \cdot \log^2(en) + \log(2/\epsilon) & d = \infty, \end{cases}$$

and suppose that

$$m \geq c_0 \cdot \delta^{-2} \cdot k \cdot L'(k,n,d,\epsilon), \tag{3.7.2}$$

then, with probability at least $1-\epsilon$, the matrix $\boldsymbol{A}$ defined in (2.5.1) satisfies the wRIP over $\mathbb{C}$ of order $(k, \boldsymbol{u})$ with constant $\delta_{k,\boldsymbol{u}}$, where $\boldsymbol{u}$ are the intrinsic weights (2.4.17).

Proof. We let $N = |\Lambda|$. Then

$$\left\| \Psi_{\boldsymbol{\nu}_j} \right\|_{L^\infty_\varrho(\mathcal{U})} = u_{\boldsymbol{\nu}_j},$$

and therefore the condition $\|\Psi_{\boldsymbol{\nu}_j}\|_{L^\infty(\mathcal{U})} \leq u_{\boldsymbol{\nu}}$ required by [51, Thm. 2.14] holds. Now, using (2.4.23) (and recall that $|\Lambda_n^{\mathsf{HCI}}| = |\Lambda_{n,n}^{\mathsf{HC}}|$) we can get the estimate

$$\log(\mathrm{e}N) \leq c \begin{cases} \min\{d + \log(n), \log(2d) \cdot \log(2n)\} & d < \infty, \\ \log^2(en) & d = \infty, \end{cases}$$

for a potentially different universal constant. Here, in the last inequality, we used the estimate $\log(\mathrm{e}N) \leq 4\log^2(en)$, which comes from (2.4.23) and some basic algebra,

$$\log(\mathrm{e}N) \leq \log\left(\mathrm{e}^2 n^{2+\log(n)/\log(2)}\right) \leq \left(2 + \frac{\log(n)}{\log(2)}\right)\log(en) \leq 4\log^2(en).$$

Now, condition (3.7.2) gives

$$\begin{aligned} c_0 \cdot \delta^{-2} \cdot k \cdot (\log(\mathrm{e}N) \cdot \log^2(k/\delta)) &\leq c_0 \cdot \delta^{-2} \cdot k \cdot (\log(\mathrm{e}N) \cdot \log^2(k/\delta) + \log(2/\epsilon)) \\ &\leq c_0 \cdot \delta^{-2} \cdot k \cdot L' \leq m \end{aligned}$$

after relabelling the constant c_0 in (3.7.2) as $c \cdot c_0$. Therefore, from [51, Thm. 2.14] we obtain that with probability at least $1 - 2\exp(-c_1\delta^{-2}m/k)$ the matrix $\boldsymbol{A}$ defined in (2.5.1) satisfies the wRIP over $\mathbb{C}$ of order $(k, \boldsymbol{u})$ with constant $\delta_{k,\boldsymbol{u}}$ for some $c_1 > 0$. To conclude the result, we notice that

$$m \geq c_1\delta^{-2} \cdot k \cdot \log(2/\epsilon) \quad \Rightarrow \quad 2\exp(-c_1\delta^{-2}m/k) \leq \epsilon.$$

Hence, replacing c_0 by $\max\{c_0, c_1\}$ in (3.7.2), we deduce that the result holds with probability at least $1 - \epsilon$, as required. □

Remark 3.7.2 As formulated above, Lemma 3.7.1 only considers the case $k \geq 1$. However, for $k < 1 \leq \min_{\boldsymbol{\nu}\in\mathcal{F}} u_{\boldsymbol{\nu}}^2$, the set of weighted $(k, \boldsymbol{u})$-sparse vectors is empty. Therefore the RIP of order $(k, \boldsymbol{u})$ is trivially satisfied in this case.

3.7.3 Bounds for polynomial approximations obtained as inexact minimizers

We now present the main results of this section. These three results provide error bounds for polynomial approximations obtained as (inexact) minimizers to the weighted SR-LASSO problem (2.5.6). Each theorem corresponds to one of the three scenarios in our main results in §3.3. Hence, we label them accordingly as algebraic and finite dimensional, algebraic and infinite dimensional, and exponential. In order to state these results, we now define some additional notation. Given $f \in L^2_\varrho(\mathcal{U}; \mathcal{V})$ and $\Lambda \subseteq \mathcal{F}$, where $\mathcal{F}$ is as in (2.1.1)–(2.1.2), we let

$$E_{\Lambda,2}(f) = \|f - f_\Lambda\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}, \qquad E_{\Lambda,\infty}(f) = \|f - f_\Lambda\|_{L^\infty(\mathcal{U};\mathcal{V})},$$

where f_Λ is as in (2.5.2), and, given a subspace $\mathcal{V}_K \subseteq L^2_\varrho(\mathcal{U};\mathcal{V})$, we let

$$E_{\mathsf{disc}}(f) = \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})},$$

where $\mathcal{P}_K(f)$ is as in (2.2.15).

Theorem 3.7.3 (Error bounds for inexact minimizers, algebraic and finite-dim. case)**.** *Let $d \in \mathbb{N}$, $m \geq 3$, $0 < \epsilon < 1$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathbb{N}_0^d} \subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis, $\mathcal{V}_K$ be a subspace of $\mathcal{V}$ and $\Lambda = \Lambda^{\mathsf{HC}}_{n,d}$ be the hyperbolic cross index set with $n = \lceil m/L \rceil$ where $L = L(m,d,\epsilon)$ is as in (3.3.1). Let $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$ be a function defined everywhere, draw $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m$ randomly and independently according to ϱ and suppose that $\boldsymbol{A}$, $\boldsymbol{f}$ and $\boldsymbol{e}$ are as in (2.5.1). Consider the Hilbert-valued, weighted SR-LASSO problem (2.5.6) with weights $\boldsymbol{w} = \boldsymbol{u}$ as in (2.4.17) and $\lambda = (4\sqrt{m/L})^{-1}$. Then there exists universal constants $c_0, c_1, c_2 \geq 1$ such that the following holds with probability at least $1-\epsilon$. Any $\tilde{\boldsymbol{c}} = (\tilde{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda} \in \mathbb{C}^N$ satisfies*

$$\|f - \tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \xi, \quad \|f - \tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{k} \cdot \xi, \qquad \tilde{f} := \sum_{\boldsymbol{\nu}\in\Lambda} \tilde{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}},$$

where

$$\xi = \frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f) + E_{\mathsf{disc}}(f) + \mathcal{G}(\tilde{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}},$$

$\boldsymbol{c}_\Lambda$ is as in (2.5.2), $\mathcal{P}_K(\boldsymbol{c}_\Lambda) = (\mathcal{P}_K(c_{\boldsymbol{\nu}}))_{\boldsymbol{\nu}\in\Lambda}$, $k = m/(c_0 L)$ for $L = L(m,d,\epsilon)$ as in (3.3.1), and $\boldsymbol{n}$ is as in (2.5.1).

Proof. We divide the proof into several steps.

Step 1: Splitting the error into separate terms. Consider the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm error first. By the triangle inequality and the fact that $\mathcal{P}_K$ is a projection, we have

$$\begin{aligned}
\|f - \tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} &\leq \|f - \mathcal{P}_K(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f) - \mathcal{P}_K(f_\Lambda)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f_\Lambda) - \tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \\
&\leq \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|f - f_\Lambda\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f_\Lambda) - \tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \\
&= E_{\mathsf{disc}}(f) + E_{\Lambda,2}(f) + \|\mathcal{P}_K(f_\Lambda) - \tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}.
\end{aligned}$$

Then, by orthonormality, we have

$$\|f - \tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq E_{\mathsf{disc}}(f) + E_{\Lambda,2}(f) + \|\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \tilde{\boldsymbol{c}}\|_{2;\mathcal{V}}.$$

Similarly, for the $L^\infty(\mathcal{U};\mathcal{V})$-norm error, we have

$$\begin{aligned}\|f-\tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} &\leq \|f-\mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f)-\mathcal{P}_K(f_\Lambda)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f_\Lambda)-\tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \\ &\leq \|f-\mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|f-f_\Lambda\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f_\Lambda)-\tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \\ &= E_{\mathsf{disc}}(f) + E_{\Lambda,\infty}(f) + \|\mathcal{P}_K(f_\Lambda)-\tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})}.\end{aligned}$$

Using the definition (2.4.17) of the weights $\boldsymbol{u}$, we deduce that

$$\|f-\tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq E_{\mathsf{disc}}(f) + E_{\Lambda,\infty}(f) + \|\mathcal{P}_K(\boldsymbol{c}_\Lambda)-\tilde{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}}.$$

Therefore, the rest of the proof is devoted to showing the following bounds:

$$\|\mathcal{P}_K(\boldsymbol{c}_\Lambda)-\tilde{\boldsymbol{c}}\|_{2;\mathcal{V}} \leq c_1\cdot\xi, \quad \|\mathcal{P}_K(\boldsymbol{c}_\Lambda)-\tilde{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \leq c_2\cdot\sqrt{k}\cdot\xi. \tag{3.7.3}$$

We do this in the next two steps by first asserting that $\boldsymbol{A}$ has the weighted rNSP (Step 2) and then by applying the error bounds of Lemma 3.6.3 (Steps 3 and 4).

Step 2: Asserting the weighted rNSP. We now show that $\boldsymbol{A}$ has the weighted rNSP over $\mathcal{V}_K$ of order $(k,\boldsymbol{u})$ with probability at least $1-\epsilon/2$. This is based on Lemma 3.7.1. Let $\bar{c}_0$ be the constant in Lemma 3.7.1, set $\bar{\delta}=1/4$ and observe that

$$L = L(m,d,\epsilon) \geq \log^2(3)\cdot\min\{\log(3)+1, \log(3)\cdot\log(2)\} \geq 1,$$

since $m\geq 3$. This implies that $m \geq m/L \geq m/(c_0L) = k$ since $c_0\geq 1$ as well. Since $n=\lceil m/L\rceil \leq m/L+1\leq 2m$, we deduce that

$$\begin{aligned}&\log^2(2k/\bar{\delta})\min\{\log(n)+d, \log(2d)\cdot\log(2n)\}+\log(4/\epsilon) \\ &\qquad\leq \log^2(2m/\bar{\delta})\cdot\min\{\log(2m)+d, \log(2d)\cdot\log(4m)\}+\log(4/\epsilon) \\ &\qquad\lesssim L(m,d,\epsilon).\end{aligned}$$

In particular, this implies that

$$m = c_0\cdot k\cdot L(m,d,\epsilon) \geq \bar{c}_0\cdot\bar{\delta}^{-2}\cdot 2k\cdot L'(2k,n,d,\epsilon/2),$$

where L' is defined as in Lemma 3.7.1, and therefore (again assuming a suitably-large choice of c_0) (3.7.2) holds with k replaced by $2k$. We deduce that $\boldsymbol{A}$ satisfies the wRIP of order $(2k,\boldsymbol{u})$ with constant $\delta_{2k,\boldsymbol{u}}\leq 1/4$, with probability at least $1-\epsilon/2$. Then, we deduce from Lemmas 3.6.6 and 3.6.5 that $\boldsymbol{A}$ has (with the same probability) the weighted rNSP of order $(k,\boldsymbol{u})$ over $\mathcal{V}_K$ with constants $\rho=2\sqrt{2}/3$ and $\gamma=2\sqrt{5}/3$.

Step 3: Bounding $\mathcal{P}_K(\boldsymbol{c}_\Lambda)-\tilde{\boldsymbol{c}}$ using the weighted rNSP. We use Lemma 3.6.3. First, consider the value of λ. Since $c_0 \geq 1$ we have $m/L \geq m/(c_0 L) = k$. Hence, recalling the values for ρ and γ obtained in the previous step, we have

$$\frac{1}{4\sqrt{c_0}}\frac{1}{\sqrt{k}} = \frac{1}{4\sqrt{m/L}} = \lambda \leq \frac{1}{4\sqrt{k}} < \frac{(1+\rho)^2}{(3+\rho)\gamma}\frac{1}{\sqrt{k}}. \tag{3.7.4}$$

Therefore (3.6.4) holds. We now apply this lemma with $\mathcal{V} = \mathcal{V}_K$, $\boldsymbol{x} = \mathcal{P}_K(\boldsymbol{c}_\Lambda)$, $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{c}}$ and $\boldsymbol{e} = \boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}$. Notice first that the best $(k, \boldsymbol{u})$-approximation error (2.4.10) satisfies

$$\sigma_k(\mathcal{P}_K(\boldsymbol{c}_\Lambda))_{1,\boldsymbol{u};\mathcal{V}} = \inf\left\{ \sum_{\boldsymbol{\nu}\in\Lambda\setminus S} u_{\boldsymbol{\nu}} \|\mathcal{P}_K(c_{\boldsymbol{\nu}})\|_{\mathcal{V}} : S \subseteq \Lambda,\ |S|_{\boldsymbol{u}} \leq k \right\} \leq \sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}, \tag{3.7.5}$$

since $\mathcal{P}_K$ is a projection. Hence, applying Lemma 3.6.3 and using the lower bound in (3.7.4), we get

$$\begin{aligned}
\|\tilde{\boldsymbol{c}} - \mathcal{P}_K(\boldsymbol{c}_\Lambda)\|_{2;\mathcal{V}} &\leq c_1 \left[\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \mathcal{G}(\tilde{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) + \|\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}\|_{2;\mathcal{V}}, \right], \\
\|\tilde{\boldsymbol{c}} - \mathcal{P}_K(\boldsymbol{c}_\Lambda)\|_{1,\boldsymbol{u};\mathcal{V}} &\leq c_2 \left[\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{w};\mathcal{V}} + \sqrt{k}\left(\mathcal{G}(\tilde{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda))\right) + \sqrt{k}\|\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}\|_{2;\mathcal{V}} \right],
\end{aligned} \tag{3.7.6}$$

with probability at least $1 - \epsilon/2$. Therefore, to show (3.7.3) and therefore complete the proof, it suffices to show that the following holds with probability at least $1 - \epsilon/2$:

$$\|\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}\|_{2;\mathcal{V}} \leq \sqrt{2}\left(\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f) \right) + E_{\mathsf{disc}}(f) + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}}. \tag{3.7.7}$$

The overall result then follows by the union bound.

Step 4: Showing that (3.7.7) holds. Observe that

$$\begin{aligned}
\sqrt{m}\|(\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f})_i\|_{\mathcal{V}} &\leq \|\mathcal{P}_K(f_\Lambda)(\boldsymbol{y}_i) - f(\boldsymbol{y}_i)\|_{\mathcal{V}} + \|n_i\|_{\mathcal{V}} \\
&\leq \|\mathcal{P}_K(f_\Lambda)(\boldsymbol{y}_i) - \mathcal{P}_K(f)(\boldsymbol{y}_i)\|_{\mathcal{V}} + \|f(\boldsymbol{y}_i) - \mathcal{P}_K(f)(\boldsymbol{y}_i)\|_{\mathcal{V}} + \|n_i\|_{\mathcal{V}} \\
&\leq \|f(\boldsymbol{y}_i) - f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}} + E_{\mathsf{disc}}(f) + \|n_i\|_{\mathcal{V}}.
\end{aligned}$$

Therefore

$$\|\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}\|_{\mathcal{V};2} \leq \sqrt{\frac{1}{m}\sum_{i=1}^m \|f(\boldsymbol{y}_i) - f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}}^2} + E_{\mathsf{disc}} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}}. \tag{3.7.8}$$

For this final step, we follow near-identical arguments to those found in [12, Lem. 7.11]. This shows that

$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}\|f(\boldsymbol{y}_i)-f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}}^2} \leq \sqrt{2}\left(\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}}+E_{\Lambda,2}(f)\right), \tag{3.7.9}$$

with probability at least $1-\epsilon/2$, provided $m \geq 2k\log(4/\epsilon)$. However, the bound on this *discrete* error term follows due to the assumptions on m and the arguments given in Step 2. Thus we obtain (3.7.7) and the proof is complete. □

Theorem 3.7.4 (Error bounds for inexact minimizers, algebraic and infinite-dim. case)**.** *Let $d=\infty$, $m\geq 3$, $0<\epsilon<1$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}\subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis, $\mathcal{V}_K$ be a subspace of $\mathcal{V}$ and $\Lambda=\Lambda_n^{\mathsf{HCI}}$ be the hyperbolic cross index set with $n=\lceil m/L\rceil$ where $L=L(m,d,\epsilon)$ is as in (3.3.1). Let $f\in L^2_\varrho(\mathcal{U};\mathcal{V})$ be a function defined everywhere, draw $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m$ randomly and independently according to ϱ and suppose that $\boldsymbol{A}$, $\boldsymbol{f}$ and $\boldsymbol{e}$ are as in (2.5.1). Consider the Hilbert-valued, weighted SR-LASSO problem (2.5.6) with weights $\boldsymbol{w}=\boldsymbol{u}$ as in (2.4.17) and $\lambda=(4\sqrt{m/L})^{-1}$. Then there exists universal constants $c_0,c_1,c_2\geq 1$ such that the following holds with probability at least $1-\epsilon$. Any $\tilde{\boldsymbol{c}}=(\tilde{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}\in\mathbb{C}^N$ satisfies*

$$\|f-\tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}\leq c_1\cdot\xi,\quad \|f-\tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})}\leq c_2\cdot\sqrt{k}\cdot\xi,\qquad \tilde{f}:=\sum_{\boldsymbol{\nu}\in\Lambda}\tilde{c}_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}},$$

where

$$\xi=\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}}+\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}}+E_{\Lambda,2}(f)+E_{\mathsf{disc}}(f)+\mathcal{G}(\tilde{\boldsymbol{c}})-\mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda))+\frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}},$$

$\boldsymbol{c}_\Lambda$ is as in (2.5.2), $\mathcal{P}_K(\boldsymbol{c}_\Lambda)=(\mathcal{P}_K(c_{\boldsymbol{\nu}}))_{\boldsymbol{\nu}\in\Lambda}$, $k=m/(c_0L)$ for $L=L(m,d,\epsilon)$ as in (3.3.1), and $\boldsymbol{n}$ is as in (2.5.1).

Proof. The proof has the same structure as that of the previous theorem. Steps 1, 3 and 4 are identical. The only differences occur in Step 2. We now describe these changes. Once more we observe that $L=L(m,\infty,\epsilon)\geq 1$ since $m\geq 3$. Hence $m\geq m/L\geq m/(c_0L)=k$ since $c_0\geq 1$. We also have $n=\lceil m/L\rceil\leq 2m$. Using this, we deduce that

$$\bar{c}_0\cdot\bar{\delta}^{-2}\cdot 2k\cdot(\log^2(2k/\bar{\delta})\cdot\log^2(en)+\log(4/\epsilon))\leq c_0\cdot k\cdot L(m,\infty,\epsilon)=m,$$

for a suitably-large choice of c_0. An application of Lemma 3.7.1 now shows that $\boldsymbol{A}$ has the wRIP of order $(2k,\boldsymbol{u})$ with constant $\delta_{2k,\boldsymbol{u}}\leq 1/4$, as required. □

Theorem 3.7.5 (Error bounds for inexact minimizers, exponential case)**.** *Let $d\in\mathbb{N}$, $m\geq 3$, $0<\epsilon<1$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathbb{N}_0^d}\subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis, $\mathcal{V}_K$ be a subspace of $\mathcal{V}$ and $\Lambda=\Lambda_{n,d}^{\mathsf{HC}}$ be the hyperbolic cross index set with n as in (3.3.6). Draw*

$\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ randomly and independently according to ϱ. Then, with probability at least $1-\epsilon$, the following holds. Let $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$ be a function defined everywhere and suppose that $\boldsymbol{A}$, $\boldsymbol{f}$ and $\boldsymbol{e}$ are as in (2.5.1). Consider the Hilbert-valued, weighted SR-LASSO problem (2.5.6) with weights $\boldsymbol{w} = \boldsymbol{u}$ as in (2.4.17) and $\lambda = (4\sqrt{m/L})^{-1}$. Then there exists universal constants $c_0, c_1, c_2 \geq 1$ such that any $\tilde{\boldsymbol{c}} = (\tilde{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda} \in \mathbb{C}^N$ satisfies

$$\|f - \tilde{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \xi, \quad \|f - \tilde{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{k} \cdot \xi, \qquad \tilde{f} := \sum_{\boldsymbol{\nu}\in\Lambda} \tilde{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}},$$

where

$$\xi = \frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + E_{\Lambda,\infty}(f) + E_{\mathsf{disc}}(f) + \mathcal{G}(\tilde{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}},$$

$\boldsymbol{c}_\Lambda$ is as in (2.5.2), $\mathcal{P}_K(\boldsymbol{c}_\Lambda) = (\mathcal{P}_K(c_{\boldsymbol{\nu}}))_{\boldsymbol{\nu}\in\Lambda}$, $k = m/(c_0 L)$ for $L = L(m,d,\epsilon)$ as in (3.3.1), and $\boldsymbol{n}$ is as in (2.5.1).

Proof. The proof has the same structure as that of Theorem 3.7.3. Step 1 is identical, and reduces the proof to showing that (3.7.3) holds. We now describe the modifications needed in Steps 2–4:

Step 2: Asserting the weighted rNSP. We now show that $\boldsymbol{A}$ has the weighted rNSP over $\mathcal{V}_K$ of order $(k, \boldsymbol{u})$ with probability at least $1-\epsilon$. This step is essentially the same, except for the choice of n and the probability $1-\epsilon$ instead of $1-\epsilon/2$.

Step 3: Bounding $\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \tilde{\boldsymbol{c}}$ using the weighted rNSP. Since λ and k are the same as in Theorem 3.7.3, the bound (3.7.4) also holds in this case. We then follow the same arguments, leading to (3.7.6) holding with probability at least $1-\epsilon$. Finally, rather than (3.7.7), we ask for the slightly modified bound

$$\|\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}\|_{2;\mathcal{V}} \leq E_{\Lambda,\infty}(f) + E_{\mathsf{disc}}(f) + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}}, \tag{3.7.10}$$

to hold with probability one.

Step 4: Showing (3.7.10) holds. By the same argument, we see that (3.7.8) holds. Instead of the probabilistic bound (3.7.9), we now simply bound it as

$$\sqrt{\frac{1}{m}\sum_{i=1}^m \|f(\boldsymbol{y}_i) - f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}}^2} \leq \|f - f_\Lambda\|_{L^\infty(\mathcal{U};\mathcal{V})} = E_{\Lambda,\infty}(f).$$

This immediately implies (3.7.10).

Finally, we observe that we can simplify the previous estimates in this case using the bound $E_{\Lambda,2}(f) \leq E_{\Lambda,\infty}(f)$. □

3.8 Proofs of the main results: Theorems 3.3.1– 3.3.3

We are now ready to prove the main results of this chapter. In several of these proofs, we require the following definition. Let $s \in \mathbb{N}$ and define

$$k(s) := \max\{|S|_{\boldsymbol{u}} : S \subset \mathbb{N}_0^d, |S| \leq s, \, S \, \text{lower}\}, \tag{3.8.1}$$

where $\boldsymbol{u}$ are the intrinsic weights (2.4.17) (recall the definition of a lower set from Definition 2.4.9). It can be shown that

$$k(s) = s^2, \quad \text{(Legendre)}, \qquad k(s) \leq \min\{2^d s, s^{\log(3)/\log(2)}\}, \quad \text{(Chebyshev)}. \tag{3.8.2}$$

See, e.g., [12, Eqn. (7.42) and Props. 5.13 & 5.17]. We will use this property several times in what follows.

3.8.1 Theorem 3.3.1: algebraic rates of convergence, finite dimensions

Proof. The mapping was described in Table 3.1. As shown therein, we can write the corresponding approximation as $\hat{f} = \sum_{\boldsymbol{\nu} \in \Lambda} \hat{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$, where $\hat{\boldsymbol{c}} = (\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ is a minimizer of (2.5.6). Next, due to the various assumptions made, we may apply Theorem 3.7.3. Setting $\tilde{f} = \hat{f}$ and $\tilde{\boldsymbol{c}} = \hat{\boldsymbol{c}}$, we deduce that

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \xi, \quad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{k} \cdot \xi, \tag{3.8.3}$$

where (after writing out the term $E_{\mathsf{disc}}(f)$ explicitly)

$$\xi = \frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f) + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}}, \tag{3.8.4}$$

and $k = m/(c_0 L)$ with $c_0 \geq 1$ a universal constant. We now bound each term separately.

Step 1. The terms $\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}/\sqrt{k}$, $E_{\Lambda,\infty}(f)/\sqrt{k}$ and $E_{\Lambda,2}(f)$. The term $\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}/\sqrt{k}$ is estimated via (ii) of Theorem 2.4.10 with $q = 1$. This gives

$$\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} \leq C(d,p,\boldsymbol{\rho}) \cdot k^{1/2-1/p} = C(d,p,\boldsymbol{\rho}) \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p}. \tag{3.8.5}$$

We estimate the term $E_{\Lambda,2}(f)$ by first recalling that $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$ is the union of all lower sets (see Definition 2.4.9) of size at most $n = \lceil m/L \rceil$ (see §2.4.6). Hence, using (i) of Theorem

2.4.10 with $s = n$ and $q = 2$, we get

$$E_{\Lambda,2}(f) = \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{2;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{2;\mathcal{V}} \leq C(d,p,\boldsymbol{\rho}) \cdot n^{1/2-1/p} \leq C(d,p,\boldsymbol{\rho}) \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p}. \tag{3.8.6}$$

Here, in the last step we recall that $n = \lceil m/L \rceil$ and $c_0 \geq 1$.

It remains to consider $E_{\Lambda,\infty}(f)/\sqrt{k}$. Due to the choice of weights, we have $E_{\Lambda,\infty}(f) \leq \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}}$. We now apply (i) of Theorem 2.4.10 once more, with $s = n$ and $q = 1$, to get

$$E_{\Lambda,\infty}(f) \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(d,p,\boldsymbol{\rho}) \cdot n^{1-1/p}.$$

Since $n = \lceil m/L \rceil \geq m/(c_0 L) = k$, we obtain

$$\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} \leq C(d,p,\boldsymbol{\rho}) \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p}. \tag{3.8.7}$$

Step 2. The term $\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda))$. Since $\hat{\boldsymbol{c}}$ is a minimizer of (2.5.6) and $\mathcal{P}_K(\boldsymbol{c}_\Lambda) \in \mathcal{V}_K^N$ is feasible for (2.5.6), this term satisfies

$$\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) \leq 0. \tag{3.8.8}$$

Step 3. Conclusion. We now substitute the bounds (3.8.5)–(3.8.8) into (3.8.4). Since $k \leq m/L$, we deduce that $\xi \leq \zeta$, where ζ is given by (3.3.3). This completes the proof. □

3.8.2 Theorem 3.3.2: algebraic rates of convergence, infinite dimensions

Proof. The proof is similar to that of Theorem 3.3.1, except that it uses Theorem 3.7.4 in place of Theorem 3.7.3. In particular, we see that (3.8.3) also holds in this case with ξ as in (3.8.4) and $k = m/(c_0 L)$.

Step 2 is identical. The only differences occur in Step 1. We now describe the changes needed in this case. First consider the term $\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}/\sqrt{k}$. To bound this, we use Theorem 2.4.13 with $q = 1 > p$. This gives

$$\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} \leq C(\boldsymbol{b},\varepsilon,p) \cdot k^{1/2-1/p} = C(\boldsymbol{b},\varepsilon,p) \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p}.$$

To estimate $E_{\Lambda,2}(f)$, recall that $\Lambda = \Lambda_n^{\mathsf{HCI}}$ contains all anchored sets (see Definition 2.4.9) of size at most $n = \lceil m/L \rceil$ (§2.4.6). Hence, using Corollary 2.4.15 with $s = n$ and $q = 2 > p$, we get

$$E_{\Lambda,2}(f) = \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{2;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{2;\mathcal{V}} \leq C(\boldsymbol{b},\varepsilon,p) \cdot n^{1/2-1/p} \leq C(\boldsymbol{b},\varepsilon,p) \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p}.$$

Finally, for $E_{\Lambda,\infty}(f)$, we use Corollary 2.4.15 once more (with $q = 1 > p$) to get

$$\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} \leq \frac{\|\boldsymbol{c} - \boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} \leq C(\boldsymbol{b},\varepsilon,p)\cdot k^{1/2-1/p} = C(\boldsymbol{b},\varepsilon,p)\cdot\left(\frac{m}{c_0 L}\right)^{1/2-1/p}.$$

Having done this, we also observe that $\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) \leq 0$ in this case, since $\hat{\boldsymbol{c}}$ is once more an exact minimizer. Using this and the previously-derived bounds, we conclude that $\xi \leq \zeta$, where ζ is as in (3.3.5). This gives the result. □

3.8.3 Theorem 3.3.3: exponential rates of convergence, finite dimensions

Proof . The proof has the same structure to that of Theorem 3.3.1, the only differences being the use of Theorem 3.7.5 instead of Theorem 3.7.3 and the estimation of the various terms in Step 1. Suppose first that $m \geq c_0 2^{d+2} L$ and define the following:

$$s = \begin{cases} \lceil \sqrt{m/(4c_0 L)} \rceil & \text{Legendre,} \\ \lceil m/(4c_0 2^d L) \rceil & \text{Chebyshev.} \end{cases} \tag{3.8.9}$$

Observe that

$$s \leq \begin{cases} \sqrt{m/(c_0 L)} & \text{Legendre,} \\ m/(c_0 2^d L) & \text{Chebyshev,} \end{cases}$$

and therefore the quantity $k(s) \in \mathbb{N}$ defined in (3.8.1) satisfies

$$1 \leq k(s) \leq \frac{m}{c_0 L} = k.$$

Now consider the term $\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}/\sqrt{k}$. Notice that $\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}} \leq \sigma_{k(s)}(\boldsymbol{c})_{1,\boldsymbol{u};\mathcal{V}}$. Using this and (iii) of Theorem 2.4.10 with $p = 1$ we have

$$\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} \leq \frac{\sigma_{k(s)}(\boldsymbol{c})_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} \leq \frac{C(d,\gamma,\boldsymbol{\rho})\cdot\exp(-\gamma s^{1/d})}{\sqrt{k}} \leq C(d,\gamma,\boldsymbol{\rho})\cdot\exp(-\gamma s^{1/d}).$$

Note that this is possible since any lower set S of size at most s satisfies $|S| \leq |S|_{\boldsymbol{u}} \leq k(s)$ by definition. In the last inequality we used that $k \geq 1$.

Now consider $E_{\Lambda,\infty}(f)$. Recall that $\Lambda = \Lambda^{\mathsf{HC}}_{n,d}$, where n is as in (3.3.6). Clearly $n \geq s$, since $c_0 \geq 1$. Hence Λ contains all lower sets of size at most s. We deduce that

$$E_{\Lambda,\infty}(f) \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}},$$

for any lower set of size s. We now use (iii) of Theorem 2.4.10 with $p = 1$ once more, to get

$$E_{\Lambda,\infty}(f) \leq C(d,\gamma,\boldsymbol{\rho})\cdot\exp(-\gamma s^{1/d}).$$

We now combine this with the previous bound to deduce that the quantity ξ in Theorem 3.7.5 satisfies

$$\xi \leq C(d,\gamma,\boldsymbol{\rho}) \cdot \exp(-\gamma s^{1/d}) + E_{\mathsf{disc}}(f) + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}},$$

(here, we also recall that the term $\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) \leq 0$, as in the proof of Theorem 3.3.1). Using the value of s and recalling that $m \geq c_0 2^{d+2} L$, we deduce that

$$\xi \leq C(d,\gamma,\boldsymbol{\rho}) \cdot \begin{cases} \exp\left(-\frac{\gamma}{2}\left(\frac{m}{4c_0 L}\right)^{\frac{1}{d}}\right) & \textit{Chebyshev} \\ \exp\left(-\gamma\left(\frac{m}{4c_0 L}\right)^{\frac{1}{2d}}\right) & \textit{Legendre} \end{cases} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}.$$

However, this bound also clearly holds for all $m \geq 1$, up to a change in the constant $C(d,\gamma,\boldsymbol{\rho})$. After relabelling the universal constant $4c_0$ as c_0, we deduce that $\xi \leq \zeta$, where ζ is as in (3.3.8). This concludes the proof. □

3.9 Conclusions

Sparse polynomial approximation is a useful tool in parametric model problems, including surrogate model construction in UQ. The theory of best s-term approximation supports the use of polynomial based methods, and techniques such as least squares and compressed sensing are known to have desirable sample complexity bounds for obtaining polynomial approximations. In this work, we have closed a key gap between these two areas of research, by showing the existence of mappings that achieve exponential and near-best algebraic rates of the best s-term approximation with respect to the number of samples m.

Keeping this in mind, this Chapter answers Question 1 of §1.6 in the affirmative.

Answer to Question 1
There are mappings for computing approximations to holomorphic finite- or infinite-dimensional, Hilbert-valued functions from limited samples that achieve similar theoretical rates as benchmarks such as the best s-term polynomial approximation.

In addition, we answer Question 9 of §1.6 for the setting in this chapter.

Answer to Question 9
In the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm the errors E_{samp} and E_{disc} enter the error linearly. In the $L^\infty(\mathcal{U};\mathcal{V})$-norm these terms enter the error multiplied by a factor $\sqrt{m/L}$.

3.10 Future work

Note that this chapter is the foundation of Chapter 4, where we introduce efficient algorithms to compute the desired approximations. Therefore, we discuss future work at the end of Chapter 4.

Chapter 4

Efficient algorithms for computing near-best polynomial approximations via compressed sensing from limited samples

This chapter focuses on algorithms (i.e., methods involving finitely-many arithmetic operations) used to compute approximations of Hilbert-valued functions based on a finite set of sample values. We begin in §4.1 with various preliminaries. We recall key notation and present the problem statement in §4.1.2. In §4.2 we describe the main constributions of this chapter. Next, in §4.3, we state our main results on efficient methods approximating smooth Hilbert-valued functions. We provide a discussion on these results in §4.4. In §4.5 we recap the main setup and describe the algorithms. Next in §4.6 we describe the restarting procedure for our efficient algorithm. We continue in §4.7 with the calculations of the computational cost of the algorithms. In §4.8 we provide error bounds for solutions of a minimization problem. In §4.9 we preent the proofs of the main results. Finally, in §4.10 we write our conclusions and address Questions 2–4 of §1.6 for both scalar and Hilbert-valued functions and Question 9, outlining some future work in §4.11.

The content of this chapter is primarily based on [10].

4.1 Preliminaries

Theorems 3.3.1–3.3.3 establishes the existence of methods that takes sample values as input and produces the coefficients of a polynomial approximation that achieves a desired algebraic (in finite and infinite dimensions) or exponential (in infinite dimensions) error bounds. The methods in the previous chapter arise as minimizers of the weighted ℓ^1-minimization problem (2.5.6). However, these methods are not algorithms per se since they involve minimizers of nonlinear optimization problems. The main results in Chapter 3 do not claim that these minimizers can be computed in finitely many arithmetic operations. Thus far in this

thesis, it remains unknown whether rates similar to those proven in the previous chapter in terms of the number of samples m can be achieved through an algorithm computing a polynomial approximation from the sample values. The goal of this chapter is to address this issue.

We now give the formal problem statement, which involves a few technicalities and definitions, particularly the precise definition of an algorithm.

4.1.1 Setup

For clarity and convenience, we will recall some key definitions from the previous chapters. We consider the same setup as in Chapter 3, in particular, §3.1.1. That is, we consider $(\mathcal{V}, \langle\cdot,\cdot\rangle_{\mathcal{V}})$ to be a Hilbert space and $\mathcal{U} = [-1,1]^d$, where $d \in \mathbb{N}$ or $d = \infty$. We consider the associated tensor-product Legendre or Chebyshev polynomials obtained from the uniform or Chebyshev measure ϱ. We assume holomorphy of the continuous target function f (see §2.3) in order to attain the desired rates in §2.4.3. We draw m sample points $\boldsymbol{y}_1, \dots \boldsymbol{y}_m$ i.i.d. from ϱ and consider m noisy sample evaluations of f as in (3.1.1).

The discrete space $\mathcal{V}_K$

Let $N, K \in \mathbb{N}$. As in §3.1.1, we assume that the data (3.1.1) belongs to a finite-dimensional space $\mathcal{V}_K \subseteq \mathcal{V}$, with basis $\{\varphi_k\}_{k=1}^K$ and denote

$$\boldsymbol{G} = (\langle \varphi_j, \varphi_k \rangle_{\mathcal{V}})_{j,k=1}^{K,K} \in \mathbb{C}^{K\times K} \tag{4.1.1}$$

as the Gram matrix of this basis. In what follows, we assume that it is possible to perform matrix-vector multiplications with $\boldsymbol{G}$. In other words, we have access to the function

$$\mathcal{T}_{\boldsymbol{G}} : \mathbb{C}^K \to \mathbb{C}^K, \ \boldsymbol{x} \mapsto \boldsymbol{G}\boldsymbol{x}. \tag{4.1.2}$$

Note that $\boldsymbol{G}$ is self adjoint and positive definite. However, $\boldsymbol{G}$ is only equal to the identity when $\{\varphi_k\}_{k=1}^K$ is orthonormal. Recall from §2.2 and by the discussion in Remark 2.2.1 that $\mathcal{P}_K$ is the orthogonal projector from $\mathcal{V}$ onto $\mathcal{V}_K$, where $\mathcal{P}_K(f)(\boldsymbol{y}) = \mathcal{P}_K(f(\boldsymbol{y}))$ (see (2.2.15)).

4.1.2 Problem statement

We now formally define the input, the output, an algorithm, the computational cost of an algorithm and the problem statement of this chapter.

As in Chapter 3, we consider as an *input* (as in Definition 3.1.1) the collection of sample points $(\boldsymbol{y}_i)_{i=1}^m$ and the array of mK values $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ defined by (3.1.2) and as an *output* (as in Definition 3.1.2) the coefficients $(\hat{c}_{j,k})_{j,k=1}^{N,K} \in \mathbb{C}^{N\times K}$ providing an approximation (3.1.3) to the target function f.

Definition 4.1.1 (Algorithm for polynomial approximation of Hilbert-valued functions)**.** Let $\Lambda \subset \mathcal{F}$ of size $|\Lambda| = N$ be given, along with an indexing $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N$ of the multi-indices in Λ. An *algorithm for polynomial approximation of Hilbert-valued functions from sample values* is a mapping

$$\mathcal{A} : \mathcal{U}^m \times \mathbb{C}^{m \times K} \to \mathbb{C}^{N \times K}, \ \left((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K} \right) \mapsto (\hat{c}_{j,k})_{j,k=1}^{N,K}$$

for which the evaluation of $\mathcal{A}((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$ involves only finitely-many arithmetic operations (including square roots), comparisons and evaluations of the matrix-vector multiplication function $\mathcal{T}_{\boldsymbol{G}}$, where $\mathcal{T}_{\boldsymbol{G}}$ is as (4.1.2) and $\boldsymbol{G}$ is the Gram matrix (4.1.1). If $(d_{i,k})$ is as in (3.1.2) for some function f, then the resulting approximation $\hat{f}$ of f is given by

$$\hat{f} : \boldsymbol{y} \mapsto \sum_{j=1}^{N} \left(\sum_{k=1}^{K} \hat{c}_{j,k} \varphi_k \right) \Psi_{\boldsymbol{\nu}_j}(\boldsymbol{y}), \tag{4.1.3}$$

where $(\hat{c}_{j,k})_{j,k=1}^{N,K} = \mathcal{A}((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$.

Definition 4.1.2 (The computational cost)**.** The *computational cost* of an algorithm $\mathcal{A}$, is the maximum number of arithmetic operations and comparisons used to compute the output from any input.

Note that the number of arithmetic operations to evaluate $\boldsymbol{G}\boldsymbol{x}$ (the Gram matrix $\boldsymbol{G}$ in (4.1.1)) for any $\boldsymbol{x}$ is K^2 in general. For convenience, let

$$\boldsymbol{F} : \mathbb{C}^{K \times K} \to \mathbb{N}, \tag{4.1.4}$$

and write $\boldsymbol{F}(\boldsymbol{G})$ for the maximum number of arithmetic operations and comparisons required to evaluate $\mathcal{T}_{\boldsymbol{G}}(\boldsymbol{x})$ for arbitrary $\boldsymbol{x} \in \mathbb{C}^K$. Note that $\boldsymbol{F}(\boldsymbol{G}) \leq K^2$ in general. However, this may be smaller in certain cases, e.g., when $\boldsymbol{G}$ is structured or sparse or in the case of finite elements, depending on the connectivity of the mesh..

Then, formally stated, the problem we study in this chapter is: *devise algorithms (as in Definition 4.1.1) that take* (3.1.1) *as input (as in Definition 3.1.1) and compute approximations to the coefficients of a polynomial approximation $\hat{f}$ as outputs (as in Definition 3.1.2) to f after a finite number of arithmetic operations and comparisons with guarantees on the computational complexity (as in Definition 4.1.2) while providing guarantees on the error $f - \hat{f}$ in the $L^2_\varrho(\mathcal{U}; \mathcal{V})$- and $L^\infty(\mathcal{U}; \mathcal{V})$-norms.*

4.2 Contributions

Considering the same setup of Chapter 3, our main contribution is six theorems about algorithms (see Tables 4.1–4.2 and Algorithms 1–5), in the sense Definition 4.1.1, for constructing polynomial approximations that achieve the same rates as the theoretical benchmark

provided by the best s-term polynomial approximation in §2.4.3. In other words, polynomial approximations of holomorphic functions can be achieved in a sample efficient manner. Furthermore, they can be computed in subexponential (in the infinite-dimensional case) or algebraic computational cost (in the finite-dimensional case).

To be more specific about the contribution of this chapter we need to introduce an additional concept.

The algorithmic error

The key element of the theory in this chapter (see, e.g., Theorem 4.3.1) is that the same error bound as in Theorems 3.3.1–3.3.3 is attained, up to an additional term. In particular, we have the three sources of errors from §3.2 (see also (i)–(iii) in §2.7), plus the following fifth error (in terms of the count in §2.7):

(v) *The algorithmic error.* This term, which is denoted by E_{alg} depends on the number of iterations t performed by the algorithm that computes the coefficients of the polynomial approximation $\hat{f}$. This is the error committed by the algorithm $\mathcal{A}$ in approximately computing the methods $\mathcal{M}$ in Theorems 3.3.1–3.3.3.

We now have the concepts to describe our main contributions precisely. We assert the existence of algorithms where the algorithmic error E_{alg} decay is $\mathcal{O}(1/t)$, as $t \to \infty$, and efficient algorithms where the algorithmic error E_{alg} decay is, $\mathcal{O}(\mathrm{e}^{-t})$ as $t \to \infty$. Specifically, our error bounds take the form

$$\begin{aligned}\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} &\lesssim E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{alg}},\\ \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} &\lesssim \sqrt{\frac{m}{L}}\left(E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{alg}}\right),\end{aligned}$$

where $\hat{f}$ is an approximation to f as in (4.1.3), $L = L(m,\varepsilon)$ is a (poly)logarithmic factor in m (see (4.3.2)), the terms E_{app}, E_{disc} and E_{samp} are as in §2.7. We also construct efficient versions of these algorithms that requires an additional assumption. In this case the error bounds take the form

$$\begin{aligned}\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} &\lesssim E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{alg}} + \zeta',\\ \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} &\lesssim \sqrt{\frac{m}{L}}\left(E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{alg}} + \zeta'\right).\end{aligned}$$

Here $\zeta' > 0$ plays the role of an upper bound for $E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}$. We assume the upper bound $E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} \leq \zeta'$ as a technicality in the proof. However, in practice, this bound is not needed (for more details see below in item 2. of §4.4).

4.3 Main results

We reiterate at this stage that these results are formulated for Chebyshev and Legendre polynomials. For convenience, we define

$$\alpha = \begin{cases} 1 & \text{Legendre,} \\ \log(3)/\log(4) & \text{Chebyshev.} \end{cases} \tag{4.3.1}$$

Moreover, for convenience we recall (3.3.1). That is, given $m \geq 3$ and $\epsilon \in (0,1)$, we define

$$L = L(m,d,\epsilon) = \begin{cases} \log^2(m) \cdot \min\{\log(m)+d, \log(2d)\cdot\log(m)\} + \log(\epsilon^{-1}) & d < \infty, \\ \log^4(m) + \log(\epsilon^{-1}) & d = \infty. \end{cases} \tag{4.3.2}$$

Algebraic rates of convergence, finite dimensions

Theorem 4.3.1 (Existence of an algorithm; algebraic case, finite dimensions)**.** *Let $d \in \mathbb{N}$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathbb{N}_0^d} \subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis and $\{\varphi_k\}_{k=1}^K$ be a basis for $\mathcal{V}_K$. Then, for every $m \geq 3$, $0 < \epsilon < 1$, $K \geq 1$ and $t \geq 1$, there exists an algorithm*

$$\mathcal{A}_t : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

in the sense of Definition 4.1.1, where $N = \Theta(n,d)$ is as in (2.4.22) with $n = \lceil m/L \rceil$ and $L = L(m,d,\epsilon)$ as in (4.3.2), such that the following property holds. Let $f \in \mathcal{B}(\boldsymbol{\rho})$ for arbitrary $\boldsymbol{\rho} \geq \mathbf{1}$, draw $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m$ randomly and independently according to ϱ and let $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ be as in (3.1.2) for arbitrary noise terms $\boldsymbol{n} = (n_i)_{i=1}^m \in \mathcal{V}$. Let $(\hat{c}_{j,k})_{j,k=1}^{N,K} = \mathcal{A}_t((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$ and define the approximation $\hat{f}$ as in (4.1.3) based on the index set $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$ in (2.4.20). Then the following holds with probability at least $1-\epsilon$. The error satisfies

$$\|f-\hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \left(\zeta + \frac{1}{t}\right), \qquad \|f-\hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \left(\zeta + \frac{1}{t}\right), \tag{4.3.3}$$

where $c_1, c_2 \geq 1$ are universal constants and for any $0 < p \leq 1$,

$$\zeta := C \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}, \tag{4.3.4}$$

is as in (3.3.5) where $c_0 \geq 1$ is a universal constant and $C = C(\boldsymbol{b},\varepsilon,p)$ depends on $\boldsymbol{b}$, ε and p only. The computational cost of the algorithm is bounded by

$$c_3 \cdot \left[m \cdot \Theta(n,d) \cdot d + t \cdot (m \cdot \Theta(n,d) \cdot K + (\Theta(n,d)+m) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)) \cdot (\Theta(n,d))^\alpha\right], \tag{4.3.5}$$

where $n = \lceil m/L \rceil$ *is as in Theorem 3.3.1,* $\Theta(n,d)$ *is as in* (2.4.22), α *is as in* (4.3.1), $\boldsymbol{F}(\boldsymbol{G})$ *is as in* (4.1.4) *and* $c_3 > 0$ *is a universal constant.*

As we mentioned earlier, the same error bound as in Theorem 3.3.1 is attained, up to an additional term $E_{\mathsf{alg}} = 1/t$. Unfortunately, the $1/t$ decay rate of the algorithmic error is slow. Thus, it may be computationally expensive to compute an approximation to within a desired error bound. Fortunately, as explained in the next result, it is possible to improve it to e^{-t} subject to the additional technical assumption mentioned in §4.2.

Theorem 4.3.2 (Existence of an efficient algorithm; algebraic case, finite dimensions)**.** *Consider the setup of Theorem 4.3.1. Then for every* $t \geq 1$ *and* $\zeta' > 0$ *there exists an algorithm*

$$\mathcal{A}_{t,\zeta'} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

in the sense of Definition 4.1.1 such that the same property holds whenever $\zeta' \geq \zeta$*, except with* (4.3.3) *replaced by*

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \left(\zeta + \zeta' + \mathrm{e}^{-t}\right), \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \left(\zeta + \zeta' + \mathrm{e}^{-t}\right), \quad (4.3.6)$$

where $c_1, c_2 \geq 1$ *are universal constants and* ζ *is as in* (4.3.4)*. The computational cost of the algorithm is bounded by*

$$c_3 \cdot \left[m \cdot \Theta(n,d) \cdot d + t \cdot \left(m \cdot \Theta(n,d) \cdot K + (\Theta(n,d) + m) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)\right) \cdot (\Theta(n,d))^\alpha\right],$$

where $n = \lceil m/L \rceil$ *is as in Theorem 3.3.1,* $\Theta(n,d)$ *is as in* (2.4.22), α *is as in* (4.3.1), $\boldsymbol{F}(\boldsymbol{G})$ *is as in* (4.1.4) *and* $c_3 > 0$ *is a universal constant.*

We refer to this as an 'efficient' algorithm, since the parameter t enters linearly in the computational cost but the algorithmic error scales like e^{-t}. The main limitation of this result is that the algorithm parameter ζ' needs to be an upper bound for the true error bound ζ in order for (4.3.6) to hold. As mentioned in §4.2, this is a technical assumption for the proof, and does not appear necessary in practice as shown through numerical experiment in [10, §5].

Algebraic rates of convergence, infinite dimensions

We now consider algebraic rates of convergence in the infinite-dimensional setting.

Theorem 4.3.3 (Existence of an algorithm; algebraic case, infinite dimensions)**.** *Let* $d = \infty$*,* $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}} \subset L^2_\varrho(\mathcal{U})$ *be either the orthonormal Chebyshev or Legendre basis and* $\{\varphi_k\}_{k=1}^K$ *be a basis for* $\mathcal{V}_K$*. Then for every* $m \geq 3$*,* $0 < \epsilon < 1$*,* $K \geq 1$ *and every* $t \geq 1$*, there exists an algorithm*

$$\mathcal{A}_t : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

in the sense of Definition 4.1.1, where $N = \Theta(n,d)$ *is as in* (2.4.22) *with* $n = \lceil m/L \rceil$, *where* $L = L(m,d,\epsilon)$ *is as in* (4.3.2), *with the following property. Let* $\varepsilon > 0$, $0 < p < 1$ *and* $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ *with* $\boldsymbol{b} \geq \boldsymbol{0}$. *Let* $f \in \mathcal{H}(\boldsymbol{b},\varepsilon)$, *draw* $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m$ *randomly and independently according to* ϱ *and let* $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ *be as in* (3.1.2) *for arbitrary noise terms* $\boldsymbol{n} = (n_i)_{i=1}^m \in \mathcal{V}$. *Let* $(\hat{c}_{j,k})_{j,k=1}^{N,K} = \mathcal{A}_t((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$ *and define the approximation* $\hat{f}$ *as in* (4.1.3) *based on the index set* $\Lambda = \Lambda_n^{\mathsf{HCI}}$. *Then the following holds with probability at least* $1-\epsilon$. *The error satisfies*

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \left(\zeta + \frac{1}{t}\right), \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \left(\zeta + \frac{1}{t}\right), \tag{4.3.7}$$

where $c_1, c_2 \geq 1$ *are universal constants and for any* $0 < p \leq 1$,

$$\zeta := C \cdot \left(\frac{m}{c_0 L}\right)^{1/2-1/p} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}, \tag{4.3.8}$$

is as in (3.3.5) *where* $c_0 \geq 1$ *is a universal constant and* $C = C(\boldsymbol{b},\varepsilon,p)$ *depends on* $\boldsymbol{b}$, ε *and* p *only. The computational cost of the algorithm is bounded by*

$$c_3 \cdot \left[m \cdot \Theta(n,\infty) \cdot n + t \cdot (m \cdot \Theta(n,\infty) \cdot K + (\Theta(n,\infty) + m) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)) \cdot (\Theta(n,\infty))^\alpha\right],$$

where $n = \lceil m/L \rceil$ *is as in Theorem 3.3.2,* $\Theta(n,\infty)$ *is as in* (2.4.22), α *is as in* (4.3.1), $\boldsymbol{F}(\boldsymbol{G})$ *is as in* (4.1.4) *and* $c_3 > 0$ *is a universal constant.*

In finite dimensions, the computational cost estimate (4.3.5) is somewhat difficult to interpret, since its behaviour depends on the relative sizes of m and d. Fortunately, in infinite dimensions we can give a more informative assessment. Suppose, for simplicity, that K is fixed (for example, $K = 1$ in the case of a scalar-valued function approximation problem). Then the computational cost is bounded by

$$c \cdot m \cdot \Theta(n,\infty) \cdot n + c_K \cdot t \cdot m \cdot \Theta(n,\infty)^{\alpha+1},$$

where $c > 0$ is a universal constant $c_K > 0$ is a constant depending on K only. Recall from (2.4.22) that $\Theta(n,\infty) = |\Lambda_n^{\mathsf{HCI}}| = |\Lambda_{n,n}^{\mathsf{HC}}|$. Now, when $d = n$ and n is sufficiently large, the minimum in (2.4.23) is attained by the second term $en^{2+\log(n)/\log(2)}$. Substituting this into the above expression and recalling that $n = \lceil m/L \rceil$, where $L = L(m,\infty,\epsilon)$ is as in (4.3.2), we deduce that the computational cost is bounded by

$$c_K \cdot t \cdot m \cdot g(m)^{(\alpha+1)\log(4g(m))/\log(2)}, \qquad g(m) := \left\lceil \frac{m}{\log^4(m) + \log(\epsilon^{-1})} \right\rceil.$$

Since $m \geq 3$ by assumption, we have $\log(m) \geq 1$ and therefore $g(m) \leq m$. Hence, this admits the slightly looser upper bound

$$c_K \cdot t \cdot m^{1+(\alpha+1)\log(4m)/\log(2)}.$$

Therefore the computational cost (for fixed K and t) is *subexponential* in m.

Theorem 4.3.4 (Existence of an efficient algorithm; algebraic case, infinite dimensions)**.** *Consider the setup of Theorem 4.3.3. Then, for every $t \geq 1$ and $\zeta' > 0$ there exists an algorithm*

$$\mathcal{A}_{t,\zeta'} : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

in the sense of Definition 4.1.1 such that the same property holds whenever $\zeta' \geq \zeta$, except with (4.3.7) replaced by

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \left(\zeta + \zeta' + \mathrm{e}^{-t}\right), \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \left(\zeta + \zeta' + \mathrm{e}^{-t}\right), \tag{4.3.9}$$

where $c_1, c_2 \geq 1$ are universal constants and and $\zeta \leq \zeta'$ is as in (4.3.8). The computational cost of the algorithm is bounded by

$$c_3 \cdot \left[m \cdot \Theta(n,\infty) \cdot n + t \cdot (m \cdot \Theta(n,\infty) \cdot K + (\Theta(n,\infty) + m) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)) \cdot (\Theta(n,\infty))^\alpha\right], \tag{4.3.10}$$

where $n = \lceil m/L \rceil$ is as in Theorem 3.3.2, $\Theta(n,\infty)$ is as in (2.4.22), α is as in (4.3.1), $\boldsymbol{F}(\boldsymbol{G})$ is as in (4.1.4) and $c_3 > 0$ is a universal constant.

Similar as Theorem 4.3.2 in the finite dimensional case, Theorem 4.3.4 presents the theoretical results for our 'efficient' algorithm for infinite-dimensional function approximation.

We now recall some similarities between the results presented thus far and those in Chapter 3 (see Theorems 3.3.1 and Theorem 3.3.2). Naturally, we used the results in the previous chapter to derive our error bounds in Theorems 4.3.1–4.3.4. Thus these results are also *nonuniform* and achieve the corresponding algebraic rates for a fixed function f with high probability up to the specified error bound (see §1.4). For the same reasons discussed in Chapter 3, we attain an algebraic rate that scales like m up to a polylogarithmic factor of the order $\mathcal{O}(\log^3(m))$ in terms of m. In contrast, in the infinite-dimensional case, this rate is of the order of $\mathcal{O}(\log^4(m))$. We will see later in Chapter 6 that these algebraic rates in infinite dimensions are near-optimal. As mentioned in Theorem 3.3.1, the method in finite dimensions is more general in that it applies to any function $f \in \mathcal{B}(\boldsymbol{\rho})$ and any $\boldsymbol{\rho} \geq \mathbf{1}$. Conversely, in infinite dimensions, it applies to the class of functions $\mathcal{H}(\boldsymbol{b}, \varepsilon)$.

Exponential rates of convergence, finite dimensions

Finally, we consider exponential rates of convergence in finite dimensions.

Theorem 4.3.5 (Existence of an algorithm; exponential case, finite dimensions)**.** *Let $d \in \mathbb{N}$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathbb{N}_0^d} \subset L^2_\varrho(\mathcal{U})$ be either the orthonormal Chebyshev or Legendre basis and $\{\varphi_k\}_{k=1}^K$ be a basis for $\mathcal{V}_K$. Then for every $m \geq 3$, $0 < \epsilon < 1$, $K \geq 1$, and for every $t \geq 1$, there exists an algorithm*

$$\mathcal{A}_t : \mathcal{U}^m \times \mathbb{C}^{m\times K} \to \mathbb{C}^{N\times K},$$

in the sense of Definition 4.1.1 where $N = \Theta(n,d)$ is as in (2.4.22) with

$$n = \begin{cases} \lceil \sqrt{m/L} \rceil & \text{Legendre,} \\ \lceil m/(2^d L) \rceil & \text{Chebyshev,} \end{cases}$$

and L as in (4.3.2), with the following property. Draw $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ randomly and independently according to ϱ. Then, with probability at least $1-\epsilon$, the following holds. Let $f \in \mathcal{B}(\boldsymbol{\rho})$ for arbitrary $\boldsymbol{\rho} \geq \mathbf{1}$, $(d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m\times K}$ be as in (3.1.2) for arbitrary noise terms $\boldsymbol{n} = (n_i)_{i=1}^m \in \mathcal{V}$. Let $(\hat{c}_{j,k})_{j,k=1}^{N,K} = \mathcal{A}_t((\boldsymbol{y}_i)_{i=1}^m, (d_{i,k})_{i,k=1}^{m,K})$ and define the approximation $\hat{f}$ as in (4.1.3) based on the index set $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$. Then the error satisfies

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \left(\zeta + \frac{1}{t}\right), \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \cdot \left(\zeta + \frac{1}{t}\right), \tag{4.3.11}$$

where $c_1, c_2 \geq 1$ are as in (3.3.7), for any

$$0 < \gamma < (d+1)^{-1} \left(d! \prod_{j=1}^d \log(\rho_j)\right)^{1/d},$$

where

$$\zeta := C \cdot \begin{cases} \exp\left(-\frac{\gamma}{2}\left(\frac{m}{c_0 L}\right)^{\frac{1}{d}}\right) & \text{Chebyshev} \\ \exp\left(-\gamma\left(\frac{m}{c_0 L}\right)^{\frac{1}{2d}}\right) & \text{Legendre} \end{cases} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}, \tag{4.3.12}$$

where $c_0 \geq 1$ is a universal constant and $C = C(d,\gamma,\boldsymbol{\rho})$ depends on d, γ and $\boldsymbol{\rho}$ only. The computational cost of the algorithm is bounded by

$$c_3 \cdot \left[m \cdot \Theta(n,d) \cdot n + t \cdot (m \cdot \Theta(n,d) \cdot K + (\Theta(n,d) + m) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)) \cdot (\Theta(n,d))^\alpha\right],$$

where n is as in (3.3.6), $\Theta(n,d)$ is as in (2.4.22), α is as in (4.3.1), $\boldsymbol{F}(\boldsymbol{G})$ is as in (4.1.4) and $c_3 > 0$ is a universal constant.

Theorem 4.3.6 (Existence of an efficient algorithm; exponential case, finite dimensions)**.** *Consider the setup of Theorem 4.3.5. Suppose that there is a known upper bound $\zeta' \geq \zeta$,*

where ζ is as in (4.3.12). Then, for every $t \geq 1$ and $\zeta' > 0$ there exists an algorithm

$$\mathcal{A}_{t,\zeta'} : \mathcal{U}^m \times \mathbb{C}^{m \times K} \to \mathbb{C}^{N \times K},$$

in the sense of Definition 4.1.1 for which the same property holds whenever $\zeta' \geq \zeta$, except with (4.3.11) replaced by

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \cdot \left(\zeta + \zeta' + \mathrm{e}^{-t}\right), \qquad \|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \cdot \sqrt{\frac{m}{L}} \left(\zeta + \zeta' + \mathrm{e}^{-t}\right), \tag{4.3.13}$$

where $c_1, c_2 \geq 1$ are as in (3.3.7). The computational cost of the algorithm is bounded by

$$c_3 \cdot [m \cdot \Theta(n,d) \cdot n + t \cdot (m \cdot \Theta(n,d) \cdot K + (\Theta(n,d) + m) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)) \cdot (\Theta(n,d))^\alpha], \tag{4.3.14}$$

where n is as in (3.3.6), $\Theta(n,d)$ is as in (2.4.22), α is as in (4.3.1), $\boldsymbol{F}(\boldsymbol{G})$ is as in (4.1.4) and $c_3 > 0$ is a universal constant.

As we did before for Theorem 4.3.3, suppose that K is fixed and, since we consider exponential rates, that d is also fixed. Then, using the third estimate in (2.4.23), we deduce that the computational cost of this algorithm is bounded by

$$c_{K,d} \cdot (m \cdot n^2 \cdot (\log(n))^{d-1} + t \cdot m \cdot (n \cdot (\log(n))^{d-1})^{\alpha+1}).$$

Using the crude bound $n \leq m$, we deduce the bound

$$c_{K,d} \cdot \left(t \cdot m^{\alpha+2} (\log(m))^{(d-1)(\alpha+1)}\right).$$

Thus, for fixed t, the computational cost is polynomial in m as $m \to \infty$.

4.4 Discussion

In addition to the features established in Chapter 3 (see items 1–3 in §3.2 and §3.4), there are several distinguishing features of our analysis that we now highlight:

1. We introduce novel, efficient algorithms designed to compute approximate minimizers within a finite computational time frame. Our algorithms and analysis are based on compressed sensing theory and involve computing approximate minimizers of the weighted ℓ^1-minimization problems defined in (2.5.6). These are, to the best of our knowledge, the first results that show efficient algorithms for polynomial approximation via ℓ^1-minimization with full theoretical guarantees

2. As discussed previously in §4.2, we construct one type of algorithm (see Table 4.1 and Algorithm 2) where the *algorithmic error* E_{alg} is $\mathcal{O}(1/t)$ as $t \to \infty$. This decay

is relatively slow, especially in the regime where E_{app} is exponentially small in m. However, we also present an *efficient* algorithm (Table 4.2 and Algorithm 5) for which this term decays exponentially-fast in t (specifically, $\mathcal{O}(\mathrm{e}^{-t})$ as $t \to \infty$), subject to an additional the theoretical constraint that there is a positive constant ζ' such that $E_{\mathsf{app}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} \leq \zeta'$. This constraint is seemingly an artefact of the proof. The numerical experiments in [10] suggest it is unnecessary in practice.

3. In the infinite-dimensional case (Theorems 4.3.3–4.3.4), the computational cost is *subexponential* in m. Specifically, after t iterations of the algorithm, it is

$$\mathcal{O}\left(t \cdot m^{1+(\alpha+1)\log(4m)/\log(2)}\right), \quad m \to \infty,$$

where $\alpha = 1$ (Chebyshev) or $\alpha = \log(3)/\log(4) \approx 0.79$ (Legendre).

4. In the finite-dimensional, exponential setting (Theorems 4.3.5–4.3.6), the computational cost is *algebraic* in m for fixed d. Namely,

$$\mathcal{O}\left(t \cdot m^{\alpha+2}(\log(m))^{(d-1)(\alpha+1)}\right), \quad m \to \infty.$$

5. While these algorithms are motivated by the desire to have full error bounds, they are also completely practical. For this we refer to the series of numerical experiments in [10] demonstrating their practical efficacy. In fact, these experiments show that our algorithms work even better than what is theoretically suggested.

4.5 The construction of the algorithms in Theorems 4.3.1, 4.3.3 and 4.3.5

Before diving into the details of the proofs, we will recap our main setup, along with important aspects of the algorithms and their construction. Consider a high-dimensional continuous function $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$ with expansion (2.4.2). As in the previous chapter, we follow the setup in §2.5.2. Here, after finitely-many arithmetic operations, our algorithms recover the polynomial coefficients $\boldsymbol{c}_\Lambda \in \mathcal{V}_K^N$ of the truncated expansion (2.5.2) of f from m sample values based on the solution to the minimization problem (2.5.6) for the linear system in (2.5.4).

To develop these algorithms, we use two key ideas. First, we use a powerful, general-purpose first-order optimization method for solving (2.5.6). Second, we use the technique of *restarts* to drastically accelerate its convergence. For the former, we employ the *primal-dual iteration* (also known as the Chambolle–Pock algorithm) [57, 58]. We present error bounds for this method for solving the Hilbert-valued, weighted SR-LASSO (2.5.6), which decay

like $\mathcal{O}(1/t)$, where t is the iteration number. Next, we use a restarting procedure introduced in [78, 79], to obtain faster, exponential decay of the form $\mathcal{O}(\mathrm{e}^{-t})$.

See Remark 2.5.1 for further details about the use of SR-LASSO, as opposed to the classical LASSO or various constrained formulations.

4.5.1 Equivalent minimization problems and method well-definedness

Here we start describing the construction of the algorithms presented in our main results. First, we must show that they are well-defined methods, as outlined in Definition 3.1.3. To do so, we require some additional notation.

Given $1 \leq p \leq \infty$ and $1 \leq q \leq 2$, we define the weighted $\ell^{p,q}_{\boldsymbol{w}}$-norm of a matrix $\boldsymbol{C} = (c_{i,k})_{i,k=1}^{N,K} \in \mathbb{C}^{N \times K}$ as

$$\|\boldsymbol{C}\|_{p,q,\boldsymbol{w}} = \left(\sum_{i=1}^{N} w_i^{2-p} \left(\sum_{k=1}^{K} |c_{i,k}|^q \right)^{p/q} \right)^{1/p}.$$

Note that this is precisely the weighted $\ell^p_{\boldsymbol{w}}$-norm of the vector of $(\|\boldsymbol{c}_i\|_q)_{i=1}^N$, where $\boldsymbol{c}_i = (c_{i,k})_{k=1}^K \in \mathbb{C}^K$ is the ith row of $\boldsymbol{C}$. Further, if $p = q = 2$, then this is just the unweighted $\ell^{2,2}$-norm of a matrix (which is simply its Frobenius norm). In this case, we typically write $\|\cdot\|_{2,2}$.

As mentioned, we first must proved that our algorithm is a well-defined method. Since (2.5.6) yields a vector in $\mathcal{V}_K^N$ and, as methods, the algorithms should yield outputs in $\mathbb{C}^{N \times K}$. Thus, we first need to reformulate

$$\min_{\boldsymbol{z} \in \mathcal{V}_K^N} \mathcal{G}(\boldsymbol{z}), \qquad \mathcal{G}(\boldsymbol{z}) := \lambda \|\boldsymbol{z}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}} \tag{4.5.1}$$

using the basis $\{\varphi_i\}_{i=1}^K$ for $\mathcal{V}_K$. Notice that any vector of coefficients $\boldsymbol{c} = (c_{\boldsymbol{\nu}_i})_{i=1}^N \in \mathcal{V}_K^N$ is equivalent to a matrix of coefficients

$$\boldsymbol{C} = (c_{i,k})_{i,k=1}^{N,K} \in \mathbb{C}^{N \times K},$$

via the relation

$$c_{\boldsymbol{\nu}_i} = \sum_{k=1}^{K} c_{i,k} \varphi_k, \quad i \in [N].$$

Next, observe that if $g = \sum_{k=1}^K d_k \varphi_k \in \mathcal{V}_K$ then

$$\|g\|_{\mathcal{V}} = \|\boldsymbol{d}\|_{\boldsymbol{G}} = \sqrt{\boldsymbol{d}^* \boldsymbol{G} \boldsymbol{d}}, \tag{4.5.2}$$

where $\boldsymbol{d} = (d_k)_{k=1}^K \in \mathbb{C}^K$ and $\boldsymbol{G} \in \mathbb{C}^{K \times K}$ is the Gram matrix for $\{\varphi_k\}_{k=1}^K$, given by (4.1.1). Since $\boldsymbol{G}$ is positive definite, it has a unique positive definite square root matrix $\boldsymbol{G}^{1/2}$. Hence

we may write

$$\|g\|_{\mathcal{V}} = \|\boldsymbol{G}^{1/2}\boldsymbol{d}\|_2.$$

Now let $\boldsymbol{z} \in \mathcal{V}_K^N$ be arbitrary, $\boldsymbol{Z} \in \mathbb{C}^{N\times K}$ be the corresponding matrix and $\boldsymbol{z}_i \in \mathbb{C}^K$ be the ith row of $\boldsymbol{Z}$. Then

$$\|\boldsymbol{z}\|_{1,\boldsymbol{w};\mathcal{V}} = \sum_{i=1}^N w_i \|z_{\boldsymbol{\nu}_i}\|_{\mathcal{V}} = \sum_{i=1}^N w_i \|\boldsymbol{G}^{1/2}\boldsymbol{z}_i\|_2 = \|\boldsymbol{Z}\boldsymbol{G}^{1/2}\|_{2,1,\boldsymbol{w}}.$$

Similarly, let $\boldsymbol{A} = (a_{i,j})_{i,j=1}^{m,N} \in \mathbb{C}^{m\times N}$ and $\boldsymbol{f} = (f_i)_{i=1}^m \in \mathcal{V}_K^m$ be as in (2.5.1) and let $\boldsymbol{B} \in \mathbb{C}^{m\times K}$ be the matrix corresponding to $\boldsymbol{f}$. Then

$$\|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}^2 = \sum_{i=1}^m \left\| \sum_{j=1}^N a_{i,j} z_{\boldsymbol{\nu}_i} - f_i \right\|_{\mathcal{V}}^2 = \|(\boldsymbol{A}\boldsymbol{Z} - \boldsymbol{B})\boldsymbol{G}^{1/2}\|_{2,2}^2.$$

Therefore, we now consider the minimization problem

$$\min_{\boldsymbol{Z}\in\mathbb{C}^{N\times K}} \left\{ \lambda \|\boldsymbol{Z}\|_{2,1,\boldsymbol{w}} + \|(\boldsymbol{A}\boldsymbol{Z} - \boldsymbol{B})\boldsymbol{G}^{1/2}\|_{2,2} \right\}. \tag{4.5.3}$$

This is equivalent to (4.5.1), and so to (2.5.6), in the following sense. A vector $\hat{\boldsymbol{c}} = (\hat{c}_{\boldsymbol{\nu}_i})_{i=1}^N \in \mathcal{V}_K^N$ is a minimizer of (4.5.1) if and only if the matrix $\widehat{\boldsymbol{C}} = (\hat{c}_{i,k})_{i,k=1}^{N,K} \in \mathbb{C}^{N\times K}$ with entries defined by the relation

$$\hat{c}_{\boldsymbol{\nu}_i} = \sum_{k=1}^K \hat{c}_{i,k}\varphi_k, \quad i \in [N],$$

is a minimizer of (4.5.3).

Note that, as in §3.5.1, these are indeed well-defined methods. The minimizer of (4.5.3) with smallest $\ell^{2,2}$-norm is unique because (4.5.3) is a convex problem. Therefore, its set of minimizers is a convex set, and the function $\boldsymbol{Z} \mapsto \|\boldsymbol{Z}\|_{2,2}^2$ is strongly convex.

Recall that the error bounds for these algorithms are based on the theory developed in §3.3 for methods. Therefore, following the setting in the previous chapter, we start by deriving methods for approximately solving the optimization problem (2.5.6), or equivalently (4.5.3).

4.5.2 The primal-dual iteration

A key reason for using the *primal-dual iteration* [57], is that both functions defining the minimization problem in (4.5.4) are not required to be differentiable [19, §7.5]. We first briefly describe the primal-dual iteration in the general case (see [57–59], as well as [19, §7.5]) for more detailed treatments), before specializing to the weighted SR-LASSO problem in the next subsection.

Let $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}})$ and $(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}})$ be (complex) Hilbert spaces, $g : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, $h : \mathcal{Y} \to \mathbb{R} \cup \{\infty\}$ be proper, lower semicontinuous and convex functions and $A \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$ be a bounded linear operator satisfying $\mathrm{dom}(h) \cap A(\mathrm{dom}(g)) \neq \emptyset$. The primal-dual iteration is a general method for solving the convex optimization problem

$$\min_{x \in \mathcal{X}} \{g(x) + h(A(x))\}. \tag{4.5.4}$$

Under this setting the (Fenchel–Rockafeller) dual problem is

$$\min_{\xi \in \mathcal{Y}} \{g^* A^*(\xi)) + h^*(-\xi)\}, \tag{4.5.5}$$

where g^* and h^* are the convex conjugate functions of g and h, respectively. Recall that, for a function $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, its convex conjugate is defined by

$$f^*(z) = \sup_{x \in \mathcal{X}} \left(\mathrm{Re}\,\langle x, z\rangle_{\mathcal{V}} - f(x)\right), \quad z \in \mathcal{X}. \tag{4.5.6}$$

The Lagrangian of (4.5.4) is defined by

$$\mathcal{L}(x, \xi) = g(x) + \mathrm{Re}\,\langle A(x), \xi\rangle_{\mathcal{Y}} - h^*(\xi), \qquad x \in \mathrm{dom}(g),\ \xi \in \mathrm{dom}(h^*), \tag{4.5.7}$$

and $\mathcal{L}(x, \xi) = \infty$ if $x \notin \mathrm{dom}(g)$ or $\mathcal{L}(x, \xi) = -\infty$ if $\xi \notin \mathrm{dom}(h^*)$. This in turn leads to the saddle-point formulation of the problem

$$\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{Y}} \mathcal{L}(x, \xi).$$

The primal-dual iteration seeks a solution $(\hat{x}, \hat{\xi})$ of the saddle-point problem by solving the following fixed-point equation

$$\begin{aligned} \hat{x} &= \mathrm{prox}_{\tau g}(\hat{x} - \tau A^*(\hat{\xi})), \\ \hat{\xi} &= \mathrm{prox}_{\sigma h^*}(\hat{\xi} + \sigma A(\hat{x})), \end{aligned} \tag{4.5.8}$$

where $\tau, \sigma > 0$ are stepsize parameters and prox is the proximal operator, which is defined by

$$\mathrm{prox}_f(z) = \arg\min_{x \in \mathcal{X}} \left\{ f(x) + \frac{1}{2}\|x - z\|_{\mathcal{X}}^2 \right\}, \qquad z \in \mathrm{dom}(f).$$

To be precise, given initial values $(x^{(0)}, \xi^{(0)}) \in \mathcal{X} \times \mathcal{Y}$ the primal-dual iteration defines a sequence $\{(x^{(n)}, \xi^{(n)})\}_{n=1}^{\infty} \subset \mathcal{X} \times \mathcal{Y}$ as follows:

$$\begin{aligned} x^{(n+1)} &= \mathrm{prox}_{\tau g}(x^{(n)} - \tau A^*(\xi^{(n)})), \\ \xi^{(n+1)} &= \mathrm{prox}_{\sigma h^*}(\xi^{(n)} + \sigma A(2x^{(n+1)} - x^{(n)})). \end{aligned} \tag{4.5.9}$$

In the next section now apply this scheme to (2.5.6) and (4.5.3).

4.5.3 The primal-dual iteration for the weighted SR-LASSO problem

It is now convenient to describe an algorithm to approximately solve the Hilbert-valued problem (2.5.6). Then, by using the equivalence between elements of $\mathcal{V}_K^N$ and $\mathbb{C}^{N\times K}$, we obtain an algorithm for approximately solving (4.5.3).

Consider the problem in (2.5.6), in particular consider $\mathcal{X} = (\mathcal{V}_K^N, \langle\cdot,\cdot\rangle_{2;\mathcal{V}})$, $\mathcal{Y} = (\mathcal{V}_K^m, \langle\cdot,\cdot\rangle_{2;\mathcal{V}})$ and $g : \mathcal{X} \to \mathbb{R}\cup\{\infty\}$, $h : \mathcal{Y} \to \mathbb{R}\cup\{\infty\}$ as the proper, lower semicontinuous and convex functions

$$g(\boldsymbol{x}) = \lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}}, \qquad h(\boldsymbol{y}) = \|\boldsymbol{y}-\boldsymbol{f}\|_{2;\mathcal{V}}, \qquad \boldsymbol{x}\in\mathcal{V}_K^N,\ \boldsymbol{y}\in\mathcal{V}_K^m.$$

By using (4.5.6) we first find the proximal maps of g and h^*. For the latter, we see that

$$h^*(\boldsymbol{\xi}) = \sup_{\boldsymbol{v}\in\mathcal{V}_K^m}\left(\mathrm{Re}\,\langle\boldsymbol{v},\boldsymbol{\xi}\rangle_{\mathcal{V}} - \|\boldsymbol{v}-\boldsymbol{f}\|_{2;\mathcal{V}}\right) = \mathrm{Re}\,\langle\boldsymbol{f},\boldsymbol{\xi}\rangle_{\mathcal{V}} + \sup_{\boldsymbol{v}\in\mathcal{V}_K^m}\left(\mathrm{Re}\,\langle\boldsymbol{v},\boldsymbol{\xi}\rangle_{\mathcal{V}} - \|\boldsymbol{v}\|_{2;\mathcal{V}}\right), \quad \forall\boldsymbol{\xi}\in\mathcal{V}_K^m.$$

From [34, Ex. 13.3 & 13.4] it follows that

$$(\|\cdot\|_{\mathcal{V}})^* = \delta_B, \qquad B := \{\boldsymbol{\xi}\in\mathcal{V}_K^m : \|\boldsymbol{\xi}\|_{2;\mathcal{V}}\le 1\},$$

where δ_B is the indicator function of the set B, taking value $\delta_B(\boldsymbol{\xi}) = 0$ when $\boldsymbol{\xi}\in B$ and $+\infty$ otherwise. Hence

$$h^*(\boldsymbol{\xi}) = \mathrm{Re}\,\langle\boldsymbol{f},\boldsymbol{\xi}\rangle_{\mathcal{V}} + \delta_B(\boldsymbol{\xi}). \tag{4.5.10}$$

Using this, we obtain

$$\begin{aligned}\mathrm{prox}_{\sigma h^*}(\boldsymbol{\xi}) &= \arg\min_{\boldsymbol{z}\in\mathcal{V}_K^m}\left\{\sigma\delta_B(\boldsymbol{z}) + \sigma\mathrm{Re}\,\langle\boldsymbol{f},\boldsymbol{z}\rangle_{\mathcal{V}} + \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{\xi}\|_{2;\mathcal{V}}^2\right\}\\ &= \arg\min_{\boldsymbol{z}:\|\boldsymbol{z}\|_{2;\mathcal{V}}\le 1}\left\{\frac{1}{2}\|\boldsymbol{z}-(\boldsymbol{\xi}-\sigma\boldsymbol{f})\|_{2;\mathcal{V}}^2\right\}\\ &= \mathrm{proj}_B(\boldsymbol{\xi}-\sigma\boldsymbol{f}),\end{aligned}$$

where proj_B is the projection onto B, which is given explicitly by

$$\mathrm{proj}_B(\boldsymbol{\xi}) = \min\left\{1, \frac{1}{\|\boldsymbol{\xi}\|_{2;\mathcal{V}}}\right\}\boldsymbol{\xi}.$$

On the other hand, applying the definition of the proximal operator to the function τg with parameter $\tau > 0$, we deduce that

$$\left(\mathrm{prox}_{\tau g}(\boldsymbol{x})\right)_i = \mathrm{prox}_{\tau w_i\lambda\|\cdot\|_{\mathcal{V}}}(x_i),\ i = 1,\dots,N, \qquad \boldsymbol{x} = (x_i)_{i=1}^N\in\mathcal{V}_K^N.$$

Algorithm 1: `primal-dual-wSRLASSO` – the primal-dual iteration for the weighted SR-LASSO problem (2.5.6)

inputs : measurement matrix $\boldsymbol{A} \in \mathbb{C}^{m\times N}$, measurements $\boldsymbol{f} \in \mathcal{V}_K^N$, positive weights $\boldsymbol{w} = (w_i)_{i=1}^N$, parameter $\lambda > 0$, stepsizes $\tau, \sigma > 0$, maximum number of iterations $T \geq 1$, initial values $\boldsymbol{c}^{(0)} \in \mathcal{V}_K^N$, $\boldsymbol{\xi}^{(0)} \in \mathcal{V}_K^m$

output : $\bar{\boldsymbol{c}} = \texttt{primal-dual-wSRLASSO}(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{w}, \lambda, \tau, \sigma, T, \boldsymbol{c}^{(0)}, \boldsymbol{\xi}^{(0)})$, an approximate minimizer of (2.5.6)

initialize: $\bar{\boldsymbol{c}}^{(0)} = \boldsymbol{0} \in \mathcal{V}_K^N$

for $n = 0, 1, \dots T-1$ **do**

$\quad \boldsymbol{p} = (p_i)_{j=1}^N = \boldsymbol{c}^{(n)} - \tau \boldsymbol{A}^* \boldsymbol{\xi}^{(n)}$

$\quad \boldsymbol{c}^{(n+1)} = \left(\max\{\|p_i\|_{\mathcal{V}} - \tau\lambda w_i, 0\} \frac{p_i}{\|p_i\|_{\mathcal{V}}}\right)_{i=1}^N$

$\quad \boldsymbol{q} = \boldsymbol{\xi}^{(n)} + \sigma \boldsymbol{A}(2\boldsymbol{c}^{(n+1)} - \boldsymbol{c}^{(n)}) - \sigma \boldsymbol{f}$

$\quad \boldsymbol{\xi}^{(n+1)} = \min\left\{1, \frac{1}{\|\boldsymbol{q}\|_{2;\mathcal{V}}}\right\} \boldsymbol{q}$

$\quad \bar{\boldsymbol{c}}^{(n+1)} = \frac{n}{n+1} \bar{\boldsymbol{c}}^{(n)} + \frac{1}{n+1} \boldsymbol{c}^{(n+1)}$

end

$\bar{\boldsymbol{c}} = \bar{\boldsymbol{c}}^{(T)}$

Moreover, a simple adaptation of [34, Ex. 14.5] with the $\|\cdot\|_{\mathcal{V}}$-norm gives

$$\operatorname{prox}_{\tau\|\cdot\|_{\mathcal{V}}}(x) = \max\{\|x\|_{\mathcal{V}} - \tau, 0\} \frac{x}{\|x\|_{\mathcal{V}}}, \qquad \forall x \in \mathcal{V}_K \setminus \{0\}.$$

Hence,

$$\operatorname{prox}_{\tau g}(\boldsymbol{x}) = \left(\max\{\|x_i\|_{\mathcal{V}} - \tau\lambda w_i, 0\} \frac{x_i}{\|x_i\|_{\mathcal{V}}}\right)_{i=1}^N, \qquad \boldsymbol{x} = (x_i)_{i=1}^N \in \mathcal{V}_K^N \setminus \{\boldsymbol{0}\}.$$

With this in hand, we are now ready to define the primal-dual iteration for (2.5.6). As we see later, the analysis of convergence for the primal-dual iteration is given in terms of the *ergodic* sequence

$$\bar{\boldsymbol{c}}^{(n)} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{c}^{(i)}, \qquad n = 1, 2, \dots,$$

where $\boldsymbol{c}^{(i)} \in \mathcal{V}_K^N$ is the primal variable obtained at the ith step of the iteration. Hence, we now include the computation of these sequences in the primal-dual iteration for the weighted SR-LASSO problem (2.5.6), and take this as the output. The resulting procedure is described in Algorithm 1.

Having done this, we next adapt Algorithm 1 to obtain an algorithm for (4.5.3). This is given in Algorithm 2.

Algorithm 2: `primal-dual-wSRLASSO-C` – the primal-dual iteration for the weighted SR-LASSO problem (4.5.3)

inputs : measurement matrix $\boldsymbol{A} \in \mathbb{C}^{m\times N}$, measurements $\boldsymbol{B} \in \mathbb{C}^{m\times K}$, positive weights $\boldsymbol{w} = (w_i)_{i=1}^N$, Gram matrix $\boldsymbol{G} \in \mathbb{C}^{K\times K}$, parameter $\lambda > 0$, stepsizes $\tau, \sigma > 0$, maximum number of iterations $T \geq 1$, initial values $\boldsymbol{C}^{(0)} \in \mathbb{C}^{N\times K}$, $\boldsymbol{\Xi}^{(0)} \in \mathbb{C}^{m\times K}$

output : $\overline{\boldsymbol{C}} = \texttt{primal-dual-wSRLASSO-C}(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{w}, \boldsymbol{G}, \lambda, \tau, \sigma, T, \boldsymbol{C}^{(0)}, \boldsymbol{\Xi}^{(0)})$, an approximate minimizer of (4.5.3)

initialize: $\overline{\boldsymbol{C}}^{(0)} = \boldsymbol{0} \in \mathbb{C}^{N\times K}$

for $n = 0, 1, \dots T-1$ **do**
  $\boldsymbol{P} = (p_{i,k})_{j,k=1}^{N,K} = \boldsymbol{C}^{(n)} - \tau \boldsymbol{A}^* \boldsymbol{\Xi}^{(n)}$
  for $i = 1, \dots, N$ **do**
    $\boldsymbol{p}_i = (p_{i,k})_{k=1}^K$
    $(c_{i,k}^{(n+1)})_{k=1}^K = \max\{\|\boldsymbol{G}^{1/2}\boldsymbol{p}_i\|_2 - \tau\lambda w_i, 0\} \frac{\boldsymbol{p}_i}{\|\boldsymbol{G}^{1/2}\boldsymbol{p}_i\|_2}$
  end
  $\boldsymbol{C}^{(n+1)} = (c_{i,k}^{(n+1)})_{i,k=1}^{N,K}$
  $\boldsymbol{Q} = \boldsymbol{\Xi}^{(n)} + \sigma \boldsymbol{A}(2\boldsymbol{C}^{(n+1)} - \boldsymbol{C}^{(n)}) - \sigma \boldsymbol{B}$
  $\boldsymbol{\Xi}^{(n+1)} = \min\left\{1, \frac{1}{\|\boldsymbol{Q}\boldsymbol{G}^{1/2}\|_{2,2}}\right\} \boldsymbol{Q}$
  $\overline{\boldsymbol{C}}^{(n+1)} = \frac{n}{n+1}\overline{\boldsymbol{C}}^{(n)} + \frac{1}{n+1}\boldsymbol{C}^{(n+1)}$
end
$\overline{\boldsymbol{C}} = \overline{\boldsymbol{C}}^{(T)}$

Remark 4.5.1 Note that even though the square-root matrix $\boldsymbol{G}^{1/2}$ is used in Algorithm 2, this matrix does not need to be computed. Indeed,

$$\|\boldsymbol{G}^{1/2}\boldsymbol{d}\|_2 = \sqrt{\boldsymbol{d}^*\boldsymbol{G}\boldsymbol{d}}, \quad \boldsymbol{d} \in \mathbb{C}^K,$$

and for a matrix $\boldsymbol{C} \in \mathbb{C}^{N\times K}$, we have

$$\|\boldsymbol{C}\boldsymbol{G}^{1/2}\|_{2,2} = \sqrt{\sum_{i=1}^N \|\boldsymbol{G}^{1/2}\boldsymbol{c}_i\|_2^2} = \sqrt{\sum_{i=1}^N \boldsymbol{c}_i^*\boldsymbol{G}\boldsymbol{c}_i},$$

where $\boldsymbol{c}_i \in \mathbb{C}^K$ is the ith row of $\boldsymbol{C}$. In particular, computing $\|\boldsymbol{G}^{1/2}\boldsymbol{d}\|_2$ involves $c(\boldsymbol{F}(\boldsymbol{G}) + K)$ arithmetic operations, and computing $\|\boldsymbol{C}\boldsymbol{G}^{1/2}\|_{2,2}$ involves $cm(\boldsymbol{F}(\boldsymbol{G}) + K)$ arithmetic operations, for some universal constant $c > 0$.

4.5.4 The algorithms in Theorems 4.3.1, 4.3.3 and 4.3.5

We are now almost ready to specify the algorithms used in Theorems 4.3.1, 4.3.3 and 4.3.5. Notice that Algorithms 1 and 2 require the measurement matrix $\boldsymbol{A}$ as an input. Hence, we

Algorithm 3: `construct-A` – constructing the measurement matrix (2.5.1)

inputs : sample points $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in \mathcal{U}^d$, finite index set $\Lambda = \{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N\} \subset \mathcal{F}$
output : $\boldsymbol{A} = \texttt{construct-A}((\boldsymbol{y}_i)_{i=1}^m, \Lambda) \in \mathbb{C}^{m \times N}$, the measurement matrix (2.5.1)
initialize: $\bar{\boldsymbol{C}}^{(0)} = \boldsymbol{0} \in \mathbb{C}^{N \times K}$

1 $k = \max\{j : (\boldsymbol{\nu}_i)_j \neq 0,\ i = 1, \dots, N,\ j = 1, \dots, d\}$
2 $n = \max\{(\boldsymbol{\nu}_i)_j : i = 1, \dots, N,\ j = 1, \dots, n\}$
3 **for** $i = 1, \dots, m$ **do**
4 | Set $\boldsymbol{z} = (z_j)_{j=1}^k = ((\boldsymbol{y}_i)_j)_{j=1}^k$
5 | $b_{i,j} = \Psi_j(z_i)$, $i = 1, \dots, k$, $j = 0, \dots, n$,
6 | **for** $j = 1, \dots, N$ **do**
7 | | $a_{i,j} = \prod_{l=1}^n b_{l,(\boldsymbol{\nu}_j)_l}$
8 | **end**
9 **end**
10 $\boldsymbol{A} = \frac{1}{\sqrt{m}}(a_{i,j})_{i,j=1}^{m,N}$

need to describe the computation of this matrix for Chebyshev and Legendre polynomials. This is summarized in Algorithm 3. Notice that line 5 of this algorithm involves evaluating the first k one-dimensional Chebyshev or Legendre polynomials. This can be done efficiently via the three-term recurrence relation, as explained in §4.7 in the proof of Lemma 4.7.2.

Therefore, the specific algorithms used in Theorem 4.3.1, Theorem 4.3.3 and 4.3.5 are given in Table 4.1.

4.6 An efficient restarting procedure; the algorithms used in Theorems 4.3.2, 4.3.4 and 4.3.6

While the primal-dual iteration converges under very general conditions, it typically does so very slowly, with the error in the objective function decreasing like $\mathcal{O}(1/t)$, where t is the iteration number. To obtain exponential convergence (down to some controlled tolerance) we employ a restarting procedure. This is based on recent work of [78, 79].

A restarting procedure

Restarting is a general concept in optimization, where the output of an algorithm after a fixed number of steps is then fed into the algorithm as input, after suitably scaling the parameters of the algorithm [228–230]. In the case of the primal-dual iteration for the weighted SR-LASSO problem, this procedure involves three hyperparameters: a *tolerance* $\zeta' > 0$ and *scale* parameters $0 < r < 1$ and $s > 0$.

The efficient algorithm, step-by-step

- Let m, ϵ, n and t be as given in the particular theorem and set:
  - $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$ (Theorems 3.3.1 and 3.3.3) or $\Lambda = \Lambda_n^{\mathsf{HCI}}$ (Theorem 3.3.2),
  - $\lambda = (4\sqrt{m/L})^{-1}$, where $L = L(m, d, \epsilon)$ is as in (3.3.1),
  - $\tau = \sigma = (\Theta(n, d))^{-\alpha}$, where $\Theta(n, d)$ and α are as in (2.4.22) and (4.3.1), respectively,
  - $T = \lceil 2(\Theta(n, d))^{\alpha} t \rceil$.
- Let $\boldsymbol{D} = (d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m \times K}$ and $\boldsymbol{Y} = (\boldsymbol{y}_i)_{i=1}^{m}$ be an input, as in (3.1.2), and set $\boldsymbol{B} = \frac{1}{\sqrt{m}} \boldsymbol{D}$.
- Compute $\boldsymbol{A} = \texttt{construct-A}(\boldsymbol{Y}, \Lambda)$.
- Let $\boldsymbol{G}$ and $\boldsymbol{w}$ be as in (4.1.1) and (2.4.17), respectively.
- Define the output $\overline{\boldsymbol{C}} = \mathcal{A}(\boldsymbol{D})$, where

$$\mathcal{A}(\boldsymbol{D}) = \texttt{primal-dual-wSRLASSO-C}\,(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{w}, \boldsymbol{G}, \lambda, \tau, \sigma, T, \boldsymbol{0}, \boldsymbol{0})$$

Table 4.1: The algorithms $\mathcal{A} : \mathcal{U}^m \times \mathbb{C}^{m \times K} \to \mathbb{C}^{N \times K}$ used in Theorem 4.3.1, Theorem 4.3.3 and 4.3.5.

After applying one step of the primal-dual iteration (Algorithm 1 or 2) yielding an output $\boldsymbol{c}^{(1)}$, it then scales this vector and the right-hand side vector $\boldsymbol{f}$ by an exponentially-decaying factor a_l (defined in terms of ζ', r and s), before feeding in these values into the primal-dual iteration as input.

We explain the motivations behind the specific form of the restart procedure for the primal-dual iteration later in the proof section. For now, we simply state the procedures in the case of the weighted SR-LASSO problems (2.5.6) and (4.5.3). These are given in Algorithms 4 and 5, respectively. With these in hand, we can also give the algorithms used in Theorems 4.3.2, 4.3.4 and 4.3.6. See Table 4.2.

Note that these algorithms involve a number $c^{\star}$, which is a universal constant. It is possible to provide a precise numerical value of this constant by carefully tracking the constants in several of the proof steps. Since doing so is not especially illuminative, we forgo this additional effort. Instead, we now give a little more detail on this constant:

Remark 4.6.1 From (4.9.2) we see that $c^{\star} = 3296\sqrt{c_0}$, where c_0 is the universal constant that arises in (3.3.3). As shown in the proof of Theorem 3.7.3, the constant c_0 needs to be chosen sufficiently large so that the measurement matrix $\boldsymbol{A}$ satisfies the so-called *wRIP*. In particular, it is related to the universal constant $c > 0$ defined in Lemma 3.7.1. See, in particular, (3.7.2). A numerical value for this constant can indeed be found using results shown in [64]. With this in hand, one can then keep track of the constant c_0 in the proof of Theorem 3.7.3 to find its numerical value. This discussion also highlights why tracking the

Algorithm 4: `primal-dual-rst-wSRLASSO` – the restarted primal-dual iteration for the weighted SR-LASSO problem (2.5.6)

inputs : measurement matrix $\boldsymbol{A} \in \mathbb{C}^{m \times N}$, measurements $\boldsymbol{f} \in \mathcal{V}_K^N$, positive weights $\boldsymbol{w} = (w_i)_{i=1}^N$, parameter $\lambda > 0$, stepsizes $\tau, \sigma > 0$, number of primal-dual iterations $T \geq 1$, number of restarts $R \geq 1$, tolerance $\zeta' > 0$, scale parameter $0 < r < 1$, constant $s > 0$, initial values $\boldsymbol{c}^{(0)} = \boldsymbol{0} \in \mathcal{V}_K^N$ $\boldsymbol{\xi}^{(0)} = \boldsymbol{0} \in \mathcal{V}_K^m$.

output : $\tilde{\boldsymbol{c}} = \texttt{primal-dual-rst-wSRLASSO}(\boldsymbol{A}, \boldsymbol{f}, \boldsymbol{w}, \lambda, \tau, \sigma, T, R, \zeta', r, s)$, an approximate minimizer of (2.5.6)

initialize: $\bar{\boldsymbol{c}}^{(0)} = \boldsymbol{0} \in \mathcal{V}_K^N$, $\varepsilon_0 = \|\boldsymbol{b}\|_{2;\mathcal{V}}$

1 **for** $l = 0, \ldots, R-1$ **do**
2 $\quad \varepsilon_{l+1} = r(\varepsilon_l + \zeta')$
3 $\quad a_l = s\varepsilon_{l+1}$
4 $\quad \tilde{\boldsymbol{c}}^{(l+1)} = a_l \cdot \texttt{primal-dual-wSRLASSO}(\boldsymbol{A}, \boldsymbol{f}/a_l, \boldsymbol{w}, \lambda, \tau, \sigma, T, \tilde{\boldsymbol{c}}^{(l)}/a_l, \boldsymbol{0})$
5 **end**
6 $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{c}}^{(R)}$

value of $c^\star$ is non particularly illuminative. Indeed, it is well-known that universal constants appearing in RIP estimates in compressed sensing are generally very pessimistic [12,19,112].

4.7 The computational cost of the algorithms

This section proves a lemma on the computational cost of Algorithm 2. This will be used later when proving the main theorems:

Lemma 4.7.1 (Computational cost of Algorithm 2)**.** *The computational cost of Algorithm 2 is bounded by*

$$c \cdot (m \cdot N \cdot K + (m+N) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)) \cdot T,$$

where $c > 0$ is a universal constant and $\boldsymbol{F}$ is as in (4.1.4).

Proof. We proceed line-by-line. Line 2 involves a matrix-matrix multiplication and matrix subtraction, for a total of at most

$$c \cdot m \cdot N \cdot K \qquad \text{(line 2)}$$

arithmetic operations for some universal constant c. Now consider lines 3–5. By the previous remark, we may calculate $\|\boldsymbol{G}^{1/2}\boldsymbol{p}_i\|_2 = \sqrt{\boldsymbol{p}_i^* \boldsymbol{G} \boldsymbol{p}_i}$ using one multiplication with the matrix $\boldsymbol{G}$, one inner product of vectors of length K and one square root (recall from Definition 4.1.1 that we count square roots as arithmetic operations). This involves at most $c \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)$ arithmetic operations. Hence the cost of line 5 is at most

$$c \cdot (\boldsymbol{F}(\boldsymbol{G}) + K) \qquad \text{(line 5)},$$

Algorithm 5: `primal-dual-rst-wSRLASSO-C` – the restarted primal-dual iteration for the weighted SR-LASSO problem (4.5.3)

inputs : measurement matrix $\boldsymbol{A} \in \mathbb{C}^{m \times N}$, measurements $\boldsymbol{B} \in \mathbb{C}^{N \times K}$, positive weights $\boldsymbol{w} = (w_i)_{i=1}^N$, Gram matrix $\boldsymbol{G} \in \mathbb{C}^{K \times K}$, parameter $\lambda > 0$, stepsizes $\tau, \sigma > 0$, number of primal-dual iterations $T \geq 1$, number of restarts $R \geq 1$, tolerance $\zeta' > 0$, scale parameter $0 < r < 1$, constant $s > 0$, initial values $\boldsymbol{C}^{(0)} = \boldsymbol{0} \in \mathbb{C}^{N \times K}$, $\boldsymbol{\Xi}^{(0)} = \boldsymbol{0} \in \mathbb{C}^{m \times K}$

output : $\widetilde{\boldsymbol{C}} = \texttt{primal-dual-rst-wSRLASSO-C}(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{w}, \boldsymbol{G}, \lambda, \tau, \sigma, T, R, \zeta', r, s)$, an approximate minimizer of (4.5.3)

initialize: $\widetilde{\boldsymbol{C}}^{(0)} = \boldsymbol{0} \in \mathbb{C}^{N \times K}$, $\varepsilon_0 = \|\boldsymbol{B}\boldsymbol{G}^{1/2}\|_{2;2}$

1 **for** $l = 0, \ldots, R-1$ **do**
2 $\quad \varepsilon_{l+1} = r(\varepsilon_l + \zeta)$
3 $\quad a_l = s\varepsilon_{l+1}$
4 $\quad \widetilde{\boldsymbol{C}}^{(l+1)} = a_l \cdot \texttt{primal-dual-wSRLASSO-C}(\boldsymbol{A}, \boldsymbol{B}/a_l, \boldsymbol{w}, \boldsymbol{G}, \lambda, \tau, \sigma, T, \widetilde{\boldsymbol{C}}^{(l)}/a_l, \boldsymbol{0})$
5 **end**
6 $\widetilde{\boldsymbol{C}} = \widetilde{\boldsymbol{C}}^{(R)}$

for a possibly different universal constant c. Therefore, the total cost of lines 3–5 is

$$c \cdot (\boldsymbol{F}(\boldsymbol{G}) + K) \cdot N \qquad \text{(lines 3–5)}.$$

Line 7 involves no arithmetic operations and line 8 involves at most

$$c \cdot m \cdot N \cdot K \qquad \text{(line 8)}$$

operations. Consider line 9. Due to the previous remark, the computation of $\|\boldsymbol{Q}\boldsymbol{G}^{1/2}\|_{2,2}$ can be performed in at most $c \cdot m \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)$ operations (since $\boldsymbol{Q}$ is of size $m \times K$). Hence line 9 involves at most

$$c \cdot m \cdot (\boldsymbol{F}(\boldsymbol{G}) + K) \qquad \text{(line 9)}$$

operations. Finally, line 10 involves at most

$$c \cdot N \cdot K \qquad \text{(line 10)}$$

operations. After simplifying, we deduce that lines 2–10 involve at most

$$c \cdot (m \cdot N \cdot K + (K + \boldsymbol{F}(\boldsymbol{G})) \cdot (N + m))$$

operations. The result now follows by multiplying this by the number of iterations T. □

Lemma 4.7.2 (Computational cost of Algorithm 3)**.** *The computational cost of Algorithm 3 is bounded by*

$$c \cdot m \cdot (n + N) \cdot k,$$

- Let m, ϵ, n, t and ζ' be as given in the particular theorem and set:
  - $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$ (Theorems 4.3.2 and 4.3.6) or $\Lambda = \Lambda_n^{\mathsf{HCI}}$ (Theorem 4.3.4),
  - $\lambda = (4\sqrt{m/L})^{-1}$, where $L = L(m, d, \epsilon)$ is as in (3.3.1),
  - $\tau = \sigma = (\Theta(n, d))^{-\alpha}$, where $\Theta(n, d)$ and α are as in (2.4.22) and (4.3.1), respectively,
  - $T = \lceil (\Theta(n, d))^{\alpha} c^{\star} \rceil$, where $c^{\star}$ is a universal constant,
  - $R = t$
  - $r = \mathrm{e}^{-1}$
  - $s = \frac{(\Theta(n,d))^{\alpha} T}{2}$
- Let $\boldsymbol{D} = (d_{i,k})_{i,k=1}^{m,K} \in \mathbb{C}^{m \times K}$ and $\boldsymbol{Y} = (\boldsymbol{y}_i)_{i=1}^{m}$ be an input, as in (3.1.2), and set $\boldsymbol{B} = \frac{1}{\sqrt{m}} \boldsymbol{D}$.
- Compute $\boldsymbol{A} = \texttt{construct-A}(\boldsymbol{Y}, \Lambda)$.
- Let $\boldsymbol{G}$, $\boldsymbol{A}$ and $\boldsymbol{w}$ be as in (2.5.1), (4.1.1) and (2.4.17), respectively.
- Define the output $\widetilde{\boldsymbol{C}} = \mathcal{A}(\boldsymbol{D})$, where

$$\mathcal{A}(\boldsymbol{D}) = \texttt{primal-dual-rst-wSRLASSO-C}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{w}, \boldsymbol{G}, \lambda, \tau, \sigma, T, R, \zeta, r, c)$$

Table 4.2: The algorithms $\mathcal{A} : \mathcal{U}^m \times \mathbb{C}^{m \times K} \to \mathbb{C}^{N \times K}$ used in Theorems 4.3.2, 4.3.4 and 4.3.6.

where $c > 0$ is a universal constant and k and n are as in lines 1 and 2 of the algorithm.

Proof. Consider line 5 of the algorithm. Evaluation of the first $k+1$ Chebyshev or Legendre polynomials can be done via the three-term recurrence relation. In the Chebyshev case, this is

$$\Psi_0(z) = 1, \quad \Psi_1(z) = \sqrt{2}z, \qquad \Psi_{j+1}(z) = 2z\Psi_j(z) - c_j \Psi_{j-1}(z), \quad j = 1, \dots, k,$$

where $c_j = 1$ if $j \geq 1$ and $1/\sqrt{2}$ otherwise, and in the Legendre case, it is

$$\begin{aligned} \Psi_0(z) &= 1, \quad \Psi_1(z) = \sqrt{3}z, \\ \Psi_{j+1}(z) &= \frac{\sqrt{j+3/2}}{j+1} \left(\frac{2j+1}{\sqrt{j+1/2}} z\Psi_j(z) - \frac{j}{\sqrt{j-1/2}} \Psi_{j-1}(z) \right), \quad j = 2, \dots, k, \end{aligned}$$

(recall that these polynomials are normalized with respect to their respective probability measures). Hence the computational cost for line 5 is bounded by $c \cdot n \cdot k$. The computational cost for lines 6–8 is precisely $N \cdot (k-1)$. Hence, the computational cost for forming each row of $\boldsymbol{A}$ is bounded by $c \cdot (n \cdot k + N \cdot k)$. The result now follows. □

4.8 Error bounds and the restarting scheme for the primal-dual iteration

We now recall the theory developed in the previous chapter. In particular, the following section builds upon the arguments presented in §3.7. The bounds for polynomial approximations, obtained as inexact minimizers (Theorems 3.7.3–3.7.5), reduce the problem of proving the main results in this chapter (Theorems 4.3.1–4.3.6) to two tasks. The first involves bounding the error in the objective function, i.e., the term

$$\mathcal{G}(\tilde{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)),$$

where $\tilde{\boldsymbol{c}}$ is either an exact minimizer or an approximate minimizer obtained via the primal dual iteration. The second involves the various approximation error terms depending on f and its polynomial coefficients.

4.8.1 Overview

We begin in §4.8.2 by establishing error bounds for inexact minimizers obtained through a finite number of iterations of the primal-dual iteration and address the first task mentioned in §4.8. To be more specific, we provide an error bound for the (unrestarted) primal-dual iteration when applied to Hilbert-valued weighted SR-LASSO problem (3.6.3). This is detailed in Lemma 4.8.2. With these bounds in hand, we proceed to derive a restarting scheme in §4.8.3, which is the crucial part for the efficient algorithm. The error bound for this scheme is presented in Theorem 4.8.4.

In §4.9, we conclude with the final arguments. We utilize the three key theorems from the previous chapter (Theorems 3.7.3–3.7.5) and proceed to estimate each of the error terms mentioned earlier. For the polynomial approximation error, we rely on several results outlined in §2.4.5. To estimate the error in the objective function, we use the results provided in the subsequent section, §4.8.2. After carefully bounding the other two error terms, we finally arrive at the main results.

4.8.2 Error bounds for the primal-dual iteration

We now return to the general setting of the primal-dual iteration, where it is applied to the problem (4.5.4) and takes the form (4.5.9). The following result from [58, Theorem 5.1] establishes an important error bound for the Lagrangian difference.

Theorem 4.8.1. *Let $\tau, \sigma > 0$, initial points $(x^{(0)}, \xi^{(0)}) \in \mathcal{X} \times \mathcal{Y}$ and a bounded linear operator $A \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$, be such that $\|A\|^2_{\mathcal{B}(\mathcal{X},\mathcal{Y})} \leq (\tau\sigma)^{-1}$. Consider the sequence $\{(x^{(n)}, \xi^{(n)})\}_{n=1}^{\infty}$*

generated by the primal-dual iteration (4.5.9). *Then, for any* $(x, \xi) \in \mathcal{X} \times \mathcal{Y}$,

$$\mathcal{L}(\bar{x}^{(n)}, \xi) - \mathcal{L}(x, \bar{\xi}^{(n)}) \leq \frac{\tau^{-1}\|x - x^{(0)}\|_{2;\mathcal{V}}^2 + \sigma^{-1}\|\xi - \xi^{(0)}\|_{2;\mathcal{V}}^2}{n}, \tag{4.8.1}$$

where

$$\bar{x}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} x^{(k)} \quad \textit{and} \quad \bar{\xi}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \xi^{(k)},$$

are the ergodic sequences and $\mathcal{L}$ *is the Lagrangian* (4.5.7).

The following lemma shows a decay rate of $1/n$ on the objective function in the case of the primal-dual iteration when applied to the problem (3.6.3). It is an extension of [19, Lem. 8.6] to the weighted and Hilbert-valued setting.

Lemma 4.8.2. *Let* $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ *and* $\tau, \sigma > 0$ *be such that* $\|\boldsymbol{A}\|_{\mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)}^2 \leq (\tau\sigma)^{-1}$. *Consider the sequence* $\{(\boldsymbol{x}^{(n)}, \boldsymbol{\xi}^{(n)})\}_{n=1}^{\infty}$ *generated by the primal-dual iteration in* (4.5.9) *applied to* (3.6.3) *with* $\boldsymbol{x}^{(0)} \in \mathcal{V}^N$ *and* $\boldsymbol{\xi}^{(0)} = \boldsymbol{0} \in \mathcal{V}^m$. *Then, for any* $\boldsymbol{x} \in \mathcal{V}^N$,

$$\mathcal{G}(\bar{\boldsymbol{x}}^{(n)}) - \mathcal{G}(\boldsymbol{x}) \leq \frac{\tau^{-1}\|\boldsymbol{x} - \boldsymbol{x}_0\|_{2;\mathcal{V}}^2 + \sigma^{-1}}{n}, \qquad \bar{\boldsymbol{x}}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}^{(k)}. \tag{4.8.2}$$

Proof. Using (4.5.7) and (4.5.10), the left-hand side of (4.8.1) is given by

$$\begin{aligned}\mathcal{T}_n(\boldsymbol{x}, \boldsymbol{\xi}) := &\left(\lambda\|\bar{\boldsymbol{x}}^{(n)}\|_{1,\boldsymbol{w};\mathcal{V}} + \operatorname{Re}\langle \boldsymbol{A}\bar{\boldsymbol{x}}^{(n)} - \boldsymbol{f}, \boldsymbol{\xi}\rangle_{2;\mathcal{V}} + \delta_B(\boldsymbol{\xi})\right) \\ &- \left(\lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \operatorname{Re}\langle \boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}, \bar{\boldsymbol{\xi}}^{(n)}\rangle_{2;\mathcal{V}} + \delta_B(\bar{\boldsymbol{\xi}}^{(n)})\right),\end{aligned}$$

where B is the unit ball in $\mathcal{V}^m$. Observe that the term $\boldsymbol{\xi}^{(n)}$ produced by this iteration satisfies $\left\|\boldsymbol{\xi}^{(n)}\right\|_{2;\mathcal{V}} \leq 1$. This follows from the observation shown in §4.5.3 that the proximal mapping

$$\operatorname{prox}_{\sigma h^*}(\boldsymbol{\xi}) = \operatorname{proj}_B(\boldsymbol{\xi} - \sigma\boldsymbol{f})$$

involves the projection onto the unit ball B. Hence the ergodic sequence $\bar{\boldsymbol{\xi}}^{(n)}$ satisfies $\left\|\bar{\boldsymbol{\xi}}^{(n)}\right\|_{2;\mathcal{V}} \leq 1$ as well. Suppose now that $\boldsymbol{A}\boldsymbol{x}^{(n)} - \boldsymbol{f} \neq \boldsymbol{0}$ and set

$$\boldsymbol{\xi} = \frac{\boldsymbol{A}\boldsymbol{x}^{(n)} - \boldsymbol{f}}{\|\boldsymbol{A}\boldsymbol{x}^{(n)} - \boldsymbol{f}\|_{2;\mathcal{V}}}.$$

Then $\delta_B(\boldsymbol{\xi}) = \delta_B(\bar{\boldsymbol{\xi}}^{(n)}) = 1$ and therefore

$$\begin{aligned}\mathcal{T}_n(\boldsymbol{x}, \boldsymbol{\xi}) &= \left(\lambda\|\bar{\boldsymbol{x}}^{(n)}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\bar{\boldsymbol{x}}^{(n)} - \boldsymbol{f}\|_{2;\mathcal{V}}\right) - \left(\lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \operatorname{Re}\langle \boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}, \bar{\boldsymbol{\xi}}^{(n)}\rangle_{2;\mathcal{V}}\right) \\ &\geq \left(\lambda\|\bar{\boldsymbol{x}}^{(n)}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\bar{\boldsymbol{x}}^{(n)} - \boldsymbol{f}\|_{2;\mathcal{V}}\right) - \left(\lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}\|_{2;\mathcal{V}}\right).\end{aligned}$$

Clearly, the same bound also holds in the case $\boldsymbol{A}\boldsymbol{x}^{(n)} - \boldsymbol{f} = \boldsymbol{0}$ where $\boldsymbol{\xi}$ is an arbitrary unit vector. Hence Theorem 4.8.1 and the fact that $\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|_{2;\mathcal{V}} = \|\boldsymbol{\xi}\|_{2;\mathcal{V}} = 1$ gives the result. □

4.8.3 The restarting scheme

For convenience, we now introduce new and slightly modify some existing notation. First, we redefine the objective function $\mathcal{G}$ of the Hilbert-valued weighted SR-LASSO problem (3.6.3) to make the dependence on the term $\boldsymbol{f}$ explicit: namely, we set

$$\mathcal{G}(\boldsymbol{x}, \boldsymbol{f}) = \lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}\|_{2;\mathcal{V}}, \quad \boldsymbol{x} \in \mathcal{V}^N, \ \boldsymbol{f} \in \mathcal{V}^m.$$

We then let

$$\mathcal{E}(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{f}) = \mathcal{G}(\boldsymbol{z}, \boldsymbol{f}) - \mathcal{G}(\boldsymbol{x}, \boldsymbol{f}), \quad \boldsymbol{x}, \boldsymbol{z} \in \mathcal{V}^N, \ \boldsymbol{f} \in \mathcal{V}^m. \tag{4.8.3}$$

Now consider the ergodic sequence $\bar{\boldsymbol{x}}^{(n)}$ produced by n iterations of the primal-dual iteration (4.5.9) applied to (3.6.3) with parameters $\tau, \sigma > 0$, $\boldsymbol{x}_0 \in \mathcal{V}^N$ and $\boldsymbol{\xi}_0 = \boldsymbol{0} \in \mathcal{V}^m$. For reasons that will become clear in a moment, we now make the dependence on the vector $\boldsymbol{f}$ in (3.6.3), the number of iterations $\bar{\boldsymbol{x}}^{(n)}$ and the initial vector $\boldsymbol{x}_0$ explicit, by defining

$$\mathcal{P}(\boldsymbol{x}_0, \boldsymbol{f}, n) = \bar{\boldsymbol{x}}^{(n)}.$$

With this in hand, we conclude this discussion by noting the following two scaling properties:

$$\mathcal{G}(a\boldsymbol{x}, \boldsymbol{f}) = a\mathcal{G}(\boldsymbol{x}, \boldsymbol{f}/a), \quad \mathcal{E}(a\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{f}) = a\mathcal{E}(\boldsymbol{z}, \boldsymbol{x}/a, \boldsymbol{f}/a). \tag{4.8.4}$$

These hold for any $a > 0$ and for any $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{V}^N$ and $\boldsymbol{f} \in \mathcal{V}^m$.

Lemma 4.8.3. *Suppose that $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the weighted rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{w})$ with constants $0 < \rho < 1$ and $\gamma > 0$. Consider the Hilbert-valued weighted SR-LASSO problem (3.6.3) with parameter $\lambda = c/\sqrt{k}$, where $0 < c \leq \frac{(1+\rho)^2}{(3+\rho)\gamma}$. Let $\mathcal{E}$ and $\mathcal{P}$ be as defined above, τ, σ satisfy $\|\boldsymbol{A}\|^2_{\mathcal{B}(\mathcal{V}^N,\mathcal{V}^m)} \leq (\tau\sigma)^{-1}$ and $\boldsymbol{x}, \boldsymbol{x}_0 \in \mathcal{V}^N$, $\boldsymbol{f} \in \mathcal{V}^m$, $a > 0$. Then*

$$\mathcal{E}(a\mathcal{P}(\boldsymbol{x}_0/a, \boldsymbol{f}/a, n), \boldsymbol{x}, \boldsymbol{f}) \leq \frac{C^2}{a\tau n}\left(\mathcal{E}(\boldsymbol{x}_0, \boldsymbol{x}, \boldsymbol{f}) + \xi\right)^2 + \frac{a}{\sigma n},$$

where

$$C = 2\max\left\{C_1'/c, C_2'\right\}, \tag{4.8.5}$$

C_1', C_2' are as in Lemma 3.6.3 and

$$\xi = \xi(\boldsymbol{x}, \boldsymbol{f}) = \frac{\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}\|_{2;\mathcal{V}}. \tag{4.8.6}$$

Proof. The scaling property (4.8.4) and Lemma 4.8.2 give

$$\begin{aligned}\mathcal{E}(a\mathcal{P}(\boldsymbol{x}_0/a, \boldsymbol{f}/a, n), \boldsymbol{x}, \boldsymbol{f}) &= a\mathcal{E}(\mathcal{P}(\boldsymbol{x}_0/a, \boldsymbol{f}/a, n), \boldsymbol{x}/a, \boldsymbol{f}/a) \\ &\leq a\left(\frac{\tau^{-1}\|\boldsymbol{x}/a - \boldsymbol{x}_0/a\|_{2;\mathcal{V}}^2 + \sigma^{-1}}{n}\right) \\ &= \frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{2;\mathcal{V}}^2}{a\tau n} + \frac{a}{\sigma n}.\end{aligned}$$

Now consider the term $\|\boldsymbol{x} - \boldsymbol{x}_0\|_{2;\mathcal{V}}$. Since $\boldsymbol{A}$ has the weighted rNSP and λ satisfies (3.6.4), we may use Lemma 3.6.3 to get

$$\begin{aligned}\|\boldsymbol{x} - \boldsymbol{x}_0\|_{2;\mathcal{V}} &\leq \frac{C_1'}{\sqrt{k}}\left(2\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \frac{\mathcal{G}(\boldsymbol{x}_0, \boldsymbol{f}) - \mathcal{G}(\boldsymbol{x}, \boldsymbol{f})}{\lambda}\right) + \left(\frac{C_1'}{\sqrt{k\lambda}} + C_2'\right)\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}\|_{2;\mathcal{V}} \\ &= \frac{C_1'}{\sqrt{k\lambda}}\mathcal{E}(\boldsymbol{x}_0, \boldsymbol{x}, \boldsymbol{f}) + 2C_1'\frac{\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \left(\frac{C_1'}{\sqrt{k\lambda}} + C_2'\right)\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}\|_{2;\mathcal{V}} \\ &\leq 2\max\left\{C_1'/c, C_2'\right\}\left(\mathcal{E}(\boldsymbol{x}_0, \boldsymbol{x}, \boldsymbol{f}) + \xi\right).\end{aligned}$$

Substituting this into the previous expression now gives the result. □

This lemma gives the rationale behind the restarted scheme. It says the error in the objective function of the scaled output $a\mathcal{P}(\boldsymbol{x}_0/a, \boldsymbol{f}/a, n)$ of the primal-dual iteration with initial value $\boldsymbol{x}_0$ can be bounded in terms of the error in the objective function at the initial value, plus terms depending on the scaling parameter a, the number of iterations n and the compressed sensing error term ξ. By choosing these parameters suitably and iterating this procedure, we obtain the restarting scheme. We summarize this in the following theorem:

Theorem 4.8.4 (Restarting scheme)**.** *Suppose that $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the weighted rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{w})$ with constants $0 < \rho < 1$ and $\gamma > 0$. Consider the Hilbert-valued weighted SR-LASSO problem (3.6.3) with parameter $\lambda = c/\sqrt{k}$, where $0 < c \leq \frac{(1+\rho)^2}{(3+\rho)\gamma}$. Let $\boldsymbol{x} \in \mathcal{V}^N$, $\boldsymbol{f} \in \mathcal{V}^m$, $\zeta' \geq \xi$, where ξ is as in (4.8.6), $0 < r < 1$ and define the sequence*

$$\varepsilon_0 = \|\boldsymbol{f}\|_{2;\mathcal{V}}, \qquad \varepsilon_{k+1} = r(\varepsilon_k + \zeta'), \ \ k = 0, 1, 2, \ldots.$$

Let $\mathcal{E}$ and $\mathcal{P}$ be as defined above, τ, σ satisfy $\|\boldsymbol{A}\|_{\mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)}^2 \leq (\tau\sigma)^{-1}$ and set

$$n = \left\lceil \frac{2C}{r\sqrt{\sigma\tau}} \right\rceil, \qquad a_k = \frac{1}{2}\sigma\varepsilon_{k+1}n, \ \ k = 0, 1, 2, \ldots,$$

where C is as in (4.8.5). Then the iteration $\tilde{\boldsymbol{x}}^{(0)}, \tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(2)}, \ldots$, defined by

$$\tilde{\boldsymbol{x}}^{(0)} = \boldsymbol{0}, \qquad \tilde{\boldsymbol{x}}^{(k+1)} = a_k\mathcal{P}(\tilde{\boldsymbol{x}}^{(k)}/a_k, \boldsymbol{f}/a_k, n), \ \ k = 0, 1, 2, \ldots,$$

satisfies

$$\mathcal{E}(\boldsymbol{x}_k^\star, \boldsymbol{x}, \boldsymbol{f}) \leq \varepsilon_k \leq r^k \|\boldsymbol{f}\|_{2;\mathcal{V}} + \frac{r}{1-r}\zeta', \quad k = 0, 1, 2, \ldots.$$

Proof. We use induction on k. Suppose first that $k = 0$. Then, by definition,

$$\mathcal{E}(\tilde{\boldsymbol{x}}^{(k)}, \boldsymbol{x}, \boldsymbol{f}) = \mathcal{E}(\boldsymbol{0}, \boldsymbol{x}, \boldsymbol{f}) \leq \mathcal{G}(\boldsymbol{0}, \boldsymbol{f}) = \|\boldsymbol{f}\|_{2;\mathcal{V}} = \varepsilon_0.$$

Now suppose that the result holds for k. The previous lemma gives

$$\begin{aligned}\mathcal{E}(\tilde{\boldsymbol{x}}^{(k+1)}, \boldsymbol{x}, \boldsymbol{f}) &= \mathcal{E}(a_k \mathcal{P}(\tilde{\boldsymbol{x}}^{(k)}/a_k, \boldsymbol{f}/a_k, n), \boldsymbol{x}, \boldsymbol{f}) \\ &\leq \frac{C^2}{a_k \tau n}\left(\mathcal{E}(\tilde{\boldsymbol{x}}^{(k)}, \boldsymbol{x}, \boldsymbol{f}) + \zeta\right)^2 + \frac{a_k}{\sigma n} \\ &\leq \frac{C^2}{a_k \tau n}\left(\varepsilon_k + \zeta\right)^2 + \frac{a_k}{\sigma n}.\end{aligned}$$

We now substitute the values of n and a_k to obtain

$$\mathcal{E}(\tilde{\boldsymbol{x}}^{(k+1)}, \boldsymbol{x}, \boldsymbol{f}) = \frac{2C^2(\varepsilon_k + \zeta)}{r\sigma\tau n^2} + \frac{1}{2}r(\varepsilon_k + \zeta) \leq \frac{1}{2}r(\varepsilon_k + \zeta) + \frac{1}{2}r(\varepsilon_k + \zeta) = \varepsilon_{k+1}.$$

This completes the proof. □

This theorem states that the restarted primal-dual iteration $\tilde{\boldsymbol{x}}^{(0)}, \tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(2)}, \cdots$ yields an objective function error $\mathcal{E}(\tilde{\boldsymbol{x}}^{(k)}, \boldsymbol{x}, \boldsymbol{f})$ that converges exponentially fast in the number of restarts k. Further, each (inner) primal-dual iteration involves a number of steps n that depends on the parameters C, r, σ and τ. In other words, n is a constant independent of k. Hence, the restarted scheme converges exponentially fast in the total number of primal-dual iterations as well.

As discussed in the numerical experiments section in [10, §5.1.1], it is typical to use this theorem to optimize the choice of r. Recall that this leads to the explicit choice $r = \mathrm{e}^{-1}$. We use this value in our algorithms – see Table 4.2.

4.9 Proofs of the main results: Theorems 4.3.1–4.3.6

We are now ready to prove the main results of this chapter.

4.9.1 Theorems 4.3.1–4.3.2: algebraic rates of convergence, finite dimensions

Proof of Theorem 4.3.1. The argument is similar to that of Theorem 3.3.1. Recall from §4.5.4 that, in this case the approximation $\hat{f} = \sum_{\boldsymbol{\nu}\in\Lambda} \tilde{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$, where $\hat{\boldsymbol{c}} = \bar{\boldsymbol{c}}^{(T)}$ is the ergodic sequence obtained after T steps of the primal-dual iteration applied to (2.5.6). Hence, the only difference is the estimation of $\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda))$ in Step 2.

We now do this using Lemma 4.8.2. In order to apply this lemma we first need to estimate $\|\boldsymbol{A}\|_{\mathcal{B}(\mathcal{V}_K^N,\mathcal{V}_K^m)}$. Let $\boldsymbol{x} = (x_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda} \in \mathcal{V}_K^N$ and define $p(\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\Lambda} x_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}$. Then

$$\|\boldsymbol{A}\boldsymbol{x}\|_{2;\mathcal{V}} = \sqrt{\frac{1}{m}\sum_{i=1}^m \|p(\boldsymbol{y}_i)\|_{\mathcal{V}}^2} \leq \sup_{\boldsymbol{y}\in\mathcal{U}} \|p(\boldsymbol{y})\|_{\mathcal{V}} \leq \sum_{\boldsymbol{\nu}\in\Lambda} \|x_{\boldsymbol{\nu}}\|_{\mathcal{V}} u_{\boldsymbol{\nu}} \leq \|\boldsymbol{x}\|_{2;\mathcal{V}}\sqrt{|\Lambda|_{\boldsymbol{u}}}.$$

Now the set Λ is lower and of cardinality $|\Lambda| = \Theta(n,d)$. Hence, by (3.8.2) with $s = N$, we have $|\Lambda|_{\boldsymbol{u}} \leq (\Theta(n,d))^{2\alpha}$, where α is as in (4.3.1). Since $\boldsymbol{x}$ was arbitrary, we get

$$\|\boldsymbol{A}\|_{2;\mathcal{V}} \leq (\Theta(n,d))^{\alpha}. \tag{4.9.1}$$

Since the primal-dual iteration in §4.5.4 is used with $\tau = \sigma = (\Theta(n,d))^{-\alpha}$, we have that $\|\boldsymbol{A}\|_{2;\mathcal{V}}^2 \leq (\tau\sigma)^{-1}$. Hence we may apply Lemma 4.8.2. Since the iteration is also initialized with the zero vector and run for a total of $T = \lceil 2(\Theta(n,d))^{\alpha}t\rceil$ iterations (see §4.5.4 once more), this gives

$$\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) \leq (\Theta(n,d))^{\alpha}\frac{\|\mathcal{P}_K(\boldsymbol{c}_\Lambda)\|_{2;\mathcal{V}}^2 + 1}{T}.$$

Observe that

$$\|\mathcal{P}_K(\boldsymbol{c}_\Lambda)\|_{2;\mathcal{V}} \leq \|\boldsymbol{c}_\Lambda\|_{2;\mathcal{V}} \leq \|\boldsymbol{c}\|_{2;\mathcal{V}} = \|f\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq 1.$$

Here, in the last step, we use the fact that $f \in \mathcal{B}(\boldsymbol{\rho})$, and therefore

$$\|f\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq 1.$$

Using this and the value of T, we deduce that

$$\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) \leq \frac{1}{t}.$$

Substituting this into (3.8.4) and combining with the other estimates (3.8.5)–(3.8.7) derived in Step 2 of the proof of Theorem 3.3.1 now gives the desired error bound.

It remains to estimate the computational cost. We do this via Lemmas 4.7.1 and 4.7.2. First observe that the value k in Lemma 4.7.2 is equal to $k = d$ in this case, since the index set $\Lambda = \Lambda_{n,d}^{\mathsf{HC}}$ is a d-dimensional hyperbolic cross index set. Similarly, the value n in Lemma 4.7.2 is bounded by the order n of this hyperbolic cross. As Λ is a lower set, we also have $n \leq N$. Hence, the computational cost for forming the matrix $\boldsymbol{A}$ is bounded by $c \cdot m \cdot N \cdot d$. We now use Lemma 4.7.1 to bound the computational cost of the algorithm. Finally, we recall that $N = \Theta(n,d)$ and $T = \lceil 2(\Theta(n,d))^{\alpha}t\rceil$ in this case. □

Proof of Theorem 4.3.2. As in the previous proof, we only need to estimate the term $\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda))$. Recall from Table 4.2 that in this case $\hat{\boldsymbol{c}} = \tilde{\boldsymbol{c}}^{(R)}$ is the output of the restarted

primal-dual iteration with R restarts. Our goal is to use Theorem 4.8.4 applied to the problem (2.5.6) with weights $\boldsymbol{w} = \boldsymbol{u}$ as in (2.4.17), $\lambda = (4\sqrt{m/L})^{-1}$ and $\boldsymbol{x} = \mathcal{P}_K(\boldsymbol{c}_\Lambda)$.

We first show that the conditions of this theorem hold. Recall from Step 2 of the proof of Theorem 3.7.3 that the matrix $\boldsymbol{A}$ has the weighted rNSP of order $(k, \boldsymbol{u})$ over $\mathcal{V}_K$ with constants $\rho = 2\sqrt{2}/3$ and $\gamma = 2\sqrt{5}/3$. In particular,

$$\frac{(1+\rho)^2}{(3+\rho)\gamma} \geq 0.64.$$

We now use (3.7.4) to see that

$$\lambda = \frac{1}{4\sqrt{c_0}} \frac{1}{\sqrt{k}} \leq \frac{(1+\rho)^2}{(3+\rho)\gamma} \frac{1}{\sqrt{k}},$$

for a sufficiently large choice of c_0.

Next, with this choice of $\boldsymbol{x}$, we see that

$$\xi(\boldsymbol{x}, \boldsymbol{f}) = \frac{\sigma_k(\mathcal{P}_K(\boldsymbol{c}_\Lambda))_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + \|\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}\|_{2;\mathcal{V}}.$$

Using (3.7.5) and (3.7.7), we get

$$\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \sqrt{2}\left(\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f)\right) + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}},$$

with probability at least $1-\epsilon$. Using (3.8.5)–(3.8.7), we deduce that

$$\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \zeta,$$

with probability at least $1-\epsilon$, where ζ is as in (3.3.3). Hence, $\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \zeta'$.

Next, recall from Table 4.2 that $\tau = \sigma = (\Theta(n,d))^{-\alpha}$ in this case. Due to (4.9.1), we see that $\|\boldsymbol{A}\|_{2;\mathcal{V}} \leq (\tau\sigma)^{-1}$ as well.

Now consider the constant C defined in (4.8.5). The values for ρ and γ give that $C_1' \leq C_2' \leq 103$. Since $\lambda = c/\sqrt{k}$ with $c = 1/(4\sqrt{c_0})$, we see that

$$4C \leq 812/c = 3296\sqrt{c_0} := c^\star. \tag{4.9.2}$$

Therefore, recalling that $r = 1/2$ and $\tau = \sigma = (\Theta(n,d))^{-\alpha}$, we see that

$$\left\lceil \frac{2C}{r\sqrt{\sigma\tau}} \right\rceil = \lceil (\Theta(n,d))^\alpha c^\star \rceil = T,$$

where T is as specified in Table 4.2, and

$$\frac{1}{2} r\sigma(\varepsilon_k + \zeta')T = \frac{(\Theta(n,d))^\alpha T}{4} \varepsilon_{k+1} = s\varepsilon_{k+1} = a_k,$$

where s and a_k are as specified in Table 4.2 and Algorithm 4, respectively.

With this in hand, we are now finally in a position to apply Theorem 4.8.4. We deduce that

$$\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) = \mathcal{E}(\tilde{\boldsymbol{c}}^{(R)}, \mathcal{P}_K(\boldsymbol{c}_\Lambda), \boldsymbol{f}) \leq \varepsilon_k = \mathrm{e}^{-R}\|\boldsymbol{f}\|_{2;\mathcal{V}} + \zeta'.$$

To complete the proof of the error bound (4.3.6), we simply note that $\|\boldsymbol{f}\|_{2;\mathcal{V}} \leq \|f\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq 1$, since $f \in \mathcal{B}(\boldsymbol{\rho})$.

It remains to estimate the computational cost. As before, the computational cost for forming the matrix $\boldsymbol{A}$ is bounded by $c \cdot m \cdot N \cdot d$. Next, by construction, we observe that the algorithm consists of $R = t$ primal-dual iterations, each involving $T = \lceil (\Theta(n,d))^\alpha c^\star \rceil$ steps. Therefore, by Lemma 4.7.1 the computational cost for the algorithm is

$$c \cdot (m \cdot N \cdot K + (m+N) \cdot (\boldsymbol{F}(\boldsymbol{G}) + K)) \cdot \lceil (\Theta(n,d))^\alpha c^\star \rceil \cdot t.$$

Since $N = \Theta(n,d)$ and $c^\star$ is a universal constant, the result follows. □

4.9.2 Theorems 4.3.3–4.3.4: algebraic rates of convergence, infinite dimensions

Proof of Theorem 4.3.3. The argument is similar to that of Theorem 4.3.1. Here $\hat{\boldsymbol{c}} = \bar{\boldsymbol{c}}^{(T)}$ is the ergodic sequence obtained after T steps of the primal-dual iteration applied to (2.5.6) as well.

We recall that the set Λ is lower and of cardinality $|\Lambda| = \Theta(n,d)$ with $d = \infty$. Hence, by (3.8.2) with $s = N$, we have $|\Lambda|_{\boldsymbol{u}} \leq (\Theta(n,d))^{2\alpha}$, where α is as in (4.3.1). Using this, we get

$$\|\boldsymbol{A}\|_{2;\mathcal{V}} \leq (\Theta(n,d))^\alpha,$$

as before. Since the primal-dual iteration in Table 4.2 is used with $\tau = \sigma = (\Theta(n,d))^{-\alpha}$, we have that $\|\boldsymbol{A}\|_{2;\mathcal{V}}^2 \leq (\tau\sigma)^{-1}$. Hence, following the same steps we deduce that

$$\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) \leq \frac{1}{t}.$$

Substituting this into (3.8.4) and combining with the other estimates (3.8.5)–(3.8.7) derived in Step 2 of the proof of Theorem 3.3.1 now gives the desired error bound.

The computational cost estimate is similar to the that in the proof of Theorem 4.3.1. In this case, observe that the value k in Lemma 4.7.2 is equal to n. Hence the computational cost of forming $\boldsymbol{A}$ is bounded by $c \cdot m \cdot N \cdot n$ in this case. The computational cost for the algorithm is given by Lemma 4.7.1. To complete the estimate, we substitute the values $N = \Theta(n,d)$ and $T = \lceil 2(\Theta(n,d))^\alpha t \rceil$, as before. □

Proof of Theorem 4.3.4. The proof is similar to that of Theorem 4.3.2 and involves estimating the term $\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda))$. Using the same steps, we deduce that

$$\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \zeta,$$

with probability at least $1 - \epsilon/2$, where ζ is as in (3.3.5). Hence, $\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \zeta'$.

Next, recall from Table 4.2 that $\tau = \sigma = (\Theta(n, d))^{-\alpha}$ with $d = \infty$ in this case. Due to (4.9.1), we see that $\|\boldsymbol{A}\|_{2;\mathcal{V}} \leq (\tau\sigma)^{-1}$ holds. We now apply Theorem 4.8.4 to obtain

$$\mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)) = \mathcal{E}(\tilde{\boldsymbol{c}}^{(R)}, \mathcal{P}_K(\boldsymbol{c}_\Lambda), \boldsymbol{f}) \leq \varepsilon_R = \mathrm{e}^{-R}\|\boldsymbol{f}\|_{2;\mathcal{V}} + \zeta'.$$

To complete the proof of the error bound (4.3.6), we simply note that $\|\boldsymbol{f}\|_{2;\mathcal{V}} \leq \|f\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq 1$, since $f \in \mathcal{B}(\boldsymbol{b}, \varepsilon)$.

The computational cost estimate is as in the previous proof. □

4.9.3 Theorems 4.3.5–4.3.6: exponential rates of convergence, finite dimensions

Proof of Theorem 4.3.5. The argument is the same as the proof of Theorem 4.3.1. The difference relies on the fact that now ζ has the following bound

$$\xi \leq C \cdot \begin{cases} \exp\left(-\frac{\gamma}{2}\left(\frac{m}{4c_0 L}\right)^{\frac{1}{d}}\right) & \textit{Chebyshev} \\ \exp\left(-\gamma\left(\frac{m}{4c_0 L}\right)^{\frac{1}{2d}}\right) & \textit{Legendre} \end{cases} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}} + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \mathcal{G}(\hat{\boldsymbol{c}}) - \mathcal{G}(\mathcal{P}_K(\boldsymbol{c}_\Lambda)),$$

where $C = C(d, \gamma, \boldsymbol{\rho})$. To estimate the final term, we argue exactly as in the proof of Theorem 4.3.1. The computational cost estimate is likewise identical. □

Proof of Theorem 4.3.6. The proof is similar to that of Theorem 4.3.2, except we use Theorem 3.7.5 instead. Recall from Step 2 of the proof of Theorem 3.7.5 that the matrix $\boldsymbol{A}$ has the weighted rNSP of order $(k, \boldsymbol{u})$ over $\mathcal{V}_K$ with constants $\rho = 2\sqrt{2}/3$ and $\gamma = 2\sqrt{5}/3$ with probability $1 - \epsilon$. In particular,

$$\frac{(1+\rho)^2}{(3+\rho)\gamma} \geq 0.64.$$

We now use (3.7.4) to see that

$$\lambda = \frac{1}{4\sqrt{c_0}}\frac{1}{\sqrt{k}} \leq \frac{(1+\rho)^2}{(3+\rho)\gamma}\frac{1}{\sqrt{k}},$$

for a sufficiently large choice of c_0, as before.

Next, with the choice $\boldsymbol{x} = \mathcal{P}_K(\boldsymbol{c}_\Lambda)$ as before, we see that

$$\xi(\boldsymbol{x}, \boldsymbol{f}) = \frac{\sigma_k(\mathcal{P}_K(\boldsymbol{c}_\Lambda))_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + \|\boldsymbol{A}\mathcal{P}_K(\boldsymbol{c}_\Lambda) - \boldsymbol{f}\|_{2;\mathcal{V}}.$$

Using (3.7.10), we get

$$\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + E_{\Lambda,\infty}(f) + \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \frac{\|\boldsymbol{n}\|_{2;\mathcal{V}}}{\sqrt{m}},$$

with probability $1 - \epsilon$. It now follows from the proof of Theorem 3.3.3 that

$$\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \zeta,$$

with probability at least $1 - \epsilon$, where ζ is as in (3.3.8). Hence, $\xi(\boldsymbol{x}, \boldsymbol{f}) \leq \zeta'$.

The rest of the proof follows the same steps as the proof of Theorem 4.3.2. □

4.10 Conclusions

In this chapter, we have closed a key gap between the best s-term polynomial approximation theory, which asserts exponential or algebraic rates of convergence for the approximation of holomorphic functions, and the development of efficient algorithms capable of achieving these rates after a finite number of iterations. Specifically, we have shown that (weighted) ℓ^1-minimization problems can be used to practically compute sparse polynomial approximations to holomorphic functions from limited samples. Our approach also demonstrates robustness to sampling, algorithmic, and physical discretization errors. We consider both scalar- and Hilbert-valued functions in an unknown anisotropy setting, which is particularly relevant in the context of parametric or stochastic DEs. Our results involve several significant advancements of existing techniques, including the introduction of a novel restarted primal-dual iteration for solving weighted ℓ^1-minimization problems in Hilbert spaces. Based on this, we answer Question 2 of §1.6 in the affirmative.

Answer to Question 2

There exist efficient and practical algorithms that achieve the same algebraic (in finite- or infinite-dimensions) and exponential (in finite dimensions) rates as those of the best s-term approximation with respect to the number of samples m.

The computational cost: Algebraic case, infinite dimensions

In finite dimensions, as mentioned in the discussion before Theorem 4.3.4, the computational cost (4.3.10) (for fixed K and t) is *subexponential* in m. Keeping this in mind, we answer Question 3 of §1.6 in the affirmative.

Answer to Question 3

In the infinite-dimensional case, the computational cost is *subexponential* in m. Specifically, it is

$$\mathcal{O}\left(t \cdot m^{1+(\alpha+1)\log(4m)/\log(2)}\right), \quad m \to \infty,$$

where $\alpha = 1$ (Chebyshev) or $\alpha = \log(3)/\log(4) \approx 0.79$ (Legendre).

Whether or not this can be reduced to an algebraic cost is an open problem.

The computational cost: exponential case, finite dimensions

As before, based on the discussion in §4.3, for fixed t, the computational (4.3.14) of the algorithm in Theorem 4.3.6 is polynomial in m as $m \to \infty$. In addition, by using the efficient algorithm of Theorem 4.3.6 (subject to the caveat that an upper bound for the error is known) we answer Question 4 of §1.6 in the affirmative.

Answer to Question 4

In the finite-dimensional, exponential setting, the computational cost is *algebraic* in m for fixed d. Namely,

$$\mathcal{O}\left(t \cdot m^{\alpha+2}(\log(m))^{(d-1)(\alpha+1)}\right), \quad m \to \infty,$$

where $\alpha = 1$ (Chebyshev) or $\alpha = \log(3)/\log(4) \approx 0.79$ (Legendre).

Whether or not the polynomial growth rate described above is sharp is an open problem.

In addition, we answer Question 9 of §1.6 for the setting in this chapter.

Answer to Question 9
In the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm the errors E_{samp}, E_{alg} and E_{disc} enter the error linearly. In the $L^\infty(\mathcal{U};\mathcal{V})$-norm these terms enter the error multiplied by a factor $\sqrt{m/L}$.

4.11 Future work

There are a number of avenues for further research.

- First, this work has focused on Chebyshev and Legendre polynomials on the hypercube $[-1,1]^d$. It is plausible that it can be extended to general ultraspherical or Jacobi poly-

nomials. See [17] for some work in this direction. Other possible extensions involves Hermite or Laguerre polynomials on $\mathbb{R}^d$ or $[0,\infty)^d$. This is an interesting problem for future research.

- It is notable that the algorithms developed in this thesis do not generally compute m-term polynomial approximations. Indeed, (inexact) minimizers of the SR-LASSO problem will generally be nonsparse vectors of length $N = \Theta(n,d)$. It is interesting to investigate whether one can develop algorithms that achieve the same error bounds while computing m-term polynomial approximations. In classical compressed sensing, one can typically compute sparse solutions by using a greedy or iterative procedure (see, e.g., [112]). Unfortunately, it is not clear how to extend these procedures to the weighted case with theoretical guarantees. Nonetheless, certain weighted greedy methods appear to work well in practice for sparse polynomial approximation [5]. On the other hand, a sparse approximation can always be obtained from an (inexact) minimizer of the SR-LASSO problem by thresholding. See [9, Rem. 6.9]

- We have shown that the computational cost is, at worst, subexponential in m in infinite dimensions and algebraic in m (for fixed d) in finite dimensions. This is often not the main computational bottleneck in parametric model problems (generally, computing the samples is the most computationally-intensive step). Whether these are optimal is an interesting open problem. Here, ideas from sublinear-time algorithms [65, 66] may be particularly useful.

- In the case of the exponential rates, it is notable that the best s-term approximation error is exponentially small in $\gamma \cdot s^{1/d}$ (see Theorem 2.4.7). Conversely, the exponents in §3.3 are $\gamma/2(m/(c_0 L))^{1/d}$ (Chebyshev) and $\gamma(m/(c_0 L))^{1/(2d)}$ (Legendre case). The reason for this can be traced to the sample complexity estimate for computing a sparse (and lower) polynomial approximation via compressed sensing with Monte Carlo sampling, i.e., $m \approx c_0 \cdot 2^d \cdot s \cdot L$ (Chebyshev) or $m \approx c_0 \cdot s^2 \cdot L$ (Legendre). To see why this is the case, combine Lemma 3.7.1 with (3.8.2). In the setting of least squares, in which the desired polynomial subspace is known, it is possible to change the sampling measure to obtain sample complexity bounds that are log-linear in s and therefore near optimal. See, e.g., [74, 132]. More recently, several works [32, 101, 102, 160, 182, 253] have also introduced sampling schemes that achieve linear sample complexity in s – i.e., optimal up to a constant. Unfortunately, it is unknown whether linear or log-linear sample complexity is possible in the compressed sensing setting, where the target subspace is unknown. See [15] for further discussion on this issue.

Chapter 5

Deep neural networks for Banach- and Hilbert-valued approximations from limited samples

The purpose of this chapter is to demonstrate that DL is effective at learning holomorphic infinite-dimensional functions that take values in Banach spaces from limited samples. Our main results are practical existence theorems showing that there are DNN architectures and training procedures similar to those used in practice that can efficiently (in terms of sample complexity) approximate the functions described in §2.3.1. Here we consider the Banach-valued case whereas in the previous chapters we consider the Hilbert-valued case only. We begin in §5.1 with various preliminaries. We recall key notation and present the problem statement in §5.1.2 and §5.2. Next, in §5.3, we state our main results about DNN approximation in the unknown and known anisotropy case. We provide a discussion on these results in §5.4. In §5.5 we reformulate the training problem and give the proof strategy. Next in §5.6 we provide two key lemmas to prove the wRIP property. In §5.7, we elaborate on the approximation of orthogonal polynomials by using DNNs. In §5.8 we present the proofs of the main results. Finally, in §5.9 we write our conclusions and address Question 5 of §1.6 for both scalar and Hilbert-valued functions, outlining some future work in §5.10.

In this chapter, we focus on infinite-dimensional, Banach-valued functions. The material in this chapter is primarily based on [8]. For previous work on finite-dimensional, Hilbert-valued functions, see [7].

5.1 Preliminaries

In Chapters 3 and 4 we studied methods and specific methods to construct polynomial approximations that achieved the desired convergence rates of §2.4.5. In these chapters, we focused exclusively on the unknown anisotropy case within a Hilbert-valued function approximation setting. However, driven by their impressive results in a variety of applications

(see §1.8.3), DL is increasingly supplanting classical methods and algorithms, and appears poised to revolutionize scientific computing. Nonetheless, many still question the use of DL in critical applications that require rigorous safety standards. As DL, specifically the process of training DNNs on real-world or synthetic data, is increasingly considered for applications in medicine, science, and engineering, it is important to quantify the efficiency and reliability of DL from both theoretical and practical standpoints. Informally, this leads us to posing the following question: *Are there classes of DNNs and training procedures (i.e., minimizing a loss function) from which one can learn Banach-valued* $(\boldsymbol{b}, \varepsilon)$*-holomorphic functions from limited samples? Are these classes of DNNs stable with respect to various errors that arise in the approximation problem, including those described in §2.7?*

Before stating the problem formally (see §5.1.2), we require some notation and setup.

5.1.1 Setup

Here we consider a similar setup to §3.1.1. We now highlight the main aspects and differences with the previous chapters. We consider continuous functions of the form $f : \mathcal{U} \to \mathcal{V}$, where $\mathcal{U} = [-1, 1]^{\mathbb{N}}$ is as in §2.2 with $d = \infty$ and $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ is a Banach space. Given sample points $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \sim_{\text{i.i.d.}} \varrho$ as in Chapter 1, where ϱ is either the uniform or Chebyshev (arcsine) measure over $\mathcal{U}$, we assume that the measurements take the form

$$d_i = f(\boldsymbol{y}_i) + n_i \in \mathcal{V}, \quad i = 1, \dots, m, \tag{5.1.1}$$

where the n_i represent measurement errors (see (3.1.1)). We consider approximations of $f(\boldsymbol{y})$ in the finite dimensional space $\mathcal{V}_K$. Using the basis $\{\varphi_k\}_{k=1}^K$ we can write this as

$$f(\boldsymbol{y}) \approx \sum_{k=1}^{K} c_k(\boldsymbol{y})\varphi_k. \tag{5.1.2}$$

Recall from §2.2 that we assume the existence of bounded linear operator $\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K$. Note that, in Chapter 3–4 this operator $\mathcal{P}_K$ is the orthogonal projector from $\mathcal{V}$ onto $\mathcal{V}_K$. In contrast, in this chaper we only assume that this bounded linear operator exists (see §2.2), where $\mathcal{P}_K(f)(\boldsymbol{y}) = \mathcal{P}_K(f(\boldsymbol{y}))$ as in (2.2.15) when f is defined everywhere.

Notice that, in contrast to (3.1.3), the coefficients c_k in this approximation are scalar-valued functions of $\boldsymbol{y}$, i.e., $c_k : [-1, 1]^{\mathbb{N}} \to \mathbb{R}$. We construct a DNN (as in Definition 2.6.2) to approximate these coefficient functions.

Observe that, in contrast to Chapters 3–4, we do not assume that the samples d_i to be elements of a finite-dimensional subspace. In typical applications, the samples are computed via some numerical routine, which employs a fine discretization of $\mathcal{V}$ (see §1.4 for further details). In this chapter we do not consider how the evaluations $f(\boldsymbol{y}_i)$ are obtained. It may be done by approximating the DE solution with parameter value $\boldsymbol{y}_i$, where n_i represents the

simulation error. However, it is important to mention that we do not assume any structure to the noise n_i in (5.1.1), other than it be small and bounded in norm.

Now, observe that any DNN as defined in (2.6.1) has domain $\mathbb{R}^n$, whereas the coefficients c_k in (5.1.2) have domain $\mathcal{U} \subset \mathbb{R}^{\mathbb{N}}$. In order to use a DNN to approximate infinite-dimensional functions, we also require a certain restriction operator. Let $\Theta \subset \mathbb{N}$, $|\Theta| = n$. Then we define the *variable restriction operator*

$$\mathcal{T}_\Theta : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^n, \ \boldsymbol{y} = (y_j)_{j=1}^\infty \mapsto (y_j)_{j\in\Theta}. \tag{5.1.3}$$

Given a DNN Φ of the form (2.6.1),with D hidden layers, $N_0 = |\Theta| = n$, $N_{D+2} = K$ and activation function σ we consider the approximation $f_{\Phi,\Theta}(\boldsymbol{y})$ by

$$f(\boldsymbol{y}) \approx f_{\Phi,\Theta}(\boldsymbol{y}) = \sum_{k=1}^K (\Phi \circ \mathcal{T}_\Theta(\boldsymbol{y}))_k \varphi_k. \tag{5.1.4}$$

In our main results, besides describing the DNN architecture $\mathcal{N}$ we also describe a suitable choice of set Θ defining the variable restriction operator.

5.1.2 Problem statement

We now formally define *training* as the process of constructing a DNN Φ of the form (5.1.4) that approximates f from the data $(f(\boldsymbol{y}_i) + n_i)_{i=1}^m \in \mathcal{V}^m$ by minimizing a function $\mathcal{G} : \mathcal{N} \to \mathbb{R}$. That is, for a given family of DNNs $\mathcal{N}$ with a given architecture, we define the DNN training problem with associated objective function $\mathcal{G}$ by

$$\min_{\Phi\in\mathcal{N}} \mathcal{G}(\Phi). \tag{5.1.5}$$

Note that this is equivalent (see §5.5.3–5.5.4) to a minimization problem for the weights and biases (as defined in Definition 2.6.2). In this thesis, we primarily choose $\mathcal{G}$ as

$$\mathcal{G}(\Phi) := \sqrt{\frac{1}{m}\sum_{i=1}^m \|f_{\Phi,\Theta}(\boldsymbol{y}_i) - d_i\|_{\mathcal{V}}^2} + \mathcal{J}(\Phi), \tag{5.1.6}$$

where $\mathcal{J} : \mathcal{N} \to \mathbb{R}$ is a function promoting sparsity or some other desirable feature.

Then, formally stated, the problem we study in this chapter is: *Given a truncation operator $\mathcal{T}_\Theta$ with $|\Theta| = n$ and a class of DNNs $\mathcal{N}$ as in Definition 2.6.2 of the form $\Phi : \mathbb{R}^n \to \mathbb{R}^K$, use the training data $\{(\boldsymbol{y}_i, f(\boldsymbol{y}_i) + n_i)\}_{i=1}^m \subset [-1,1]^{\mathbb{N}} \times \mathcal{V}$ as in (5.1.1) to learn $\hat{\Phi} \in \mathcal{N}$ by solving a certain minimization problem of the form (5.1.6), and construct an approximation to f of the form*

$$f \approx f_{\hat{\Phi},\Theta}(\boldsymbol{y}) = \sum_{k=1}^K (\hat{\Phi} \circ \mathcal{T}_\Theta(\boldsymbol{y}))_k \varphi_k \quad \forall \boldsymbol{y} \in \mathcal{U}, \tag{5.1.7}$$

with guarantees on the error $f - f_{\hat{\Phi},\Theta}$ in the $L^2_\varrho(\mathcal{U};\mathcal{V})$- and $L^\infty(\mathcal{U};\mathcal{V})$-Bochner norms.

5.2 Contributions

In this chapter, we establish results for DL by reinterpreting a polynomial-based approximation based on compressed sensing as a DNN training procedure, particularly by approximately emulating orthogonal polynomials using DNNs. In our previous work [7], we accomplished this for the Hilbert-valued case in finite dimensions. In this chapter, we focus on the infinite-dimensional, Banach-valued case. Specifically, we extend the Hilbert-valued compressed sensing theory developed in Chapter 3 to the Banach-valued case. Furthermore, unlike previous chapters, we address both known and unknown anisotropy cases (see §2.3.2). This involves developing DL strategies that depend on the holomorphy parameter $\boldsymbol{b}$ and lead to smaller DNN architectures, as well as strategies that are independent of $\boldsymbol{b}$.

In summary, our main contribution are four theorems dealing with the known and unknown anisotropy settings, and the case where $\mathcal{V}$ is a Banach space or a Hilbert space (the additional structure in the case of the latter yields rather improved estimates). In each theorem, we assert the existence of a DNN architecture with explicit width and depth bounds and a training procedure based on a (regularized) ℓ^2-loss function from which any resulting DNN approximates f to within an explicit error bound, with high probability. Specifically, our error bounds in the unknown anisotropy case take the form

$$\|f - \hat{f}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim E_{\mathsf{app}} + m^{\theta_1} \left(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right), \tag{5.2.1}$$

and in the $L^\infty(\mathcal{U};\mathcal{V})$-norm

$$\|f - \hat{f}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim E^\infty_{\mathsf{app}} + m^{\theta_2} \left(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right),$$

where $\hat{f}$ is the learned approximation to f. Here $\theta_1 = 0$ and $\theta_2 = 1/2$ if $\mathcal{V}$ is a Hilbert space. If $\mathcal{V}$ is a Banach space $\theta_1 = \theta_2 = 1/2$ in the unknown anisotropy case or $\theta_1 = \theta_2 = 1$ in the known anisotropy case.

There a several distinguishing features of our analysis, that we now highlight:

1. We overcome the curse of dimensionality in the approximation error. The terms E_{app} and E^∞_{app} decay algebraically fast in m/L, where L is a (poly)logarithmic factor in m. Specifically, when f is $(\boldsymbol{b},\varepsilon)$-holomorphic with $\boldsymbol{b} \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, then

$$E_{\mathsf{app}} \lesssim \pi_K \cdot \left(\frac{m}{L}\right)^{-\sigma(p)},$$

where $\sigma(p) > 0$ is given by $\sigma(p) = 1/p - 1/2$ if $\mathcal{V}$ is a Hilbert space or, if $\mathcal{V}$ is a Banach space, $\sigma(p) = 1/p - 2$ (known anisotropy) with $p \leq 1/2$ or $\sigma(p) = \frac{1}{2}(1/p - 1)$ (unknown

anisotropy) with $p \leq 1/2$, and π_K a constant depending on the finite-dimensional subspace $\mathcal{V}_K$.

3. In the Hilbert space case our error bounds are optimal in terms of the number of samples m, up to constants and (poly)logarithmic factors. Specifically, the rate $m^{1/2-1/p}$ is the best achievable for the class of functions considered, regardless of sampling strategy or learning procedure (see Chapter 6). Our results for Banach-valued functions are near-optimal in the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm. For instance, in the unknown anisotropy case they are suboptimal by a factor of $m^{\frac{1}{2p}}$ and in the known anisotropy case they are suboptimal only by a factor of $m^{\frac{3}{2}}$. We conjecture that optimal rates (up to constants and log factors) also hold for Banach-valued functions.

4. We overcome the curse of dimensionality in the DNN architecture $\mathcal{N}$. We consider either the Rectified Linear Unit (ReLU), Rectified Polynomial Unit (RePU) or hyperbolic tangent (tanh) activation function. In the ReLU case the depth of the fully-connected architecture has explicit dependence on the smoothness of f and polylogarithmic-linear scaling in m. For the latter two, the width and depth satisfy

$$\operatorname{depth}(\mathcal{N}) \lesssim \log(m), \qquad \operatorname{width}(\mathcal{N}) \lesssim \begin{cases} m^2 & \text{known anisotropy,} \\ m^{3+\log_2(m)} & \text{unknown anisotropy.} \end{cases} \tag{5.2.2}$$

5. We analyze both the known and unknown anisotropy settings. In the Hilbert space case, the only differences between the two are the width of the DNN architecture and the (poly)-logarithmic term L. Both the depth and the approximation error E_{app} have the same bounds.

6. We analyze Banach-valued functions. As observed, previous work has generally considered either scalar- or Hilbert-valued functions. To the best of our knowledge, these are first theoretical results on learning Banach-valued functions from samples with DNNs. Our results in the Hilbert-valued case are comparable to those in Chapter 4 in infinite dimensions.

The basic idea behind our theorems is to use DNNs to emulate polynomial approximation via least squares (in the known anisotropy case) and compressed sensing (in the unknown anisotropy case). As a by-product, we also show guarantees for polynomial approximation to infinite-dimensional Banach-valued functions from limited samples. To the best of our knowledge, these results are also new.

5.3 Main results

We now present the main results of this chapter. In order to state our results, we require an additional concept. First, let (5.1.5) be a DNN training problem with associated objective function $\mathcal{G}$. Then we say that $\hat{\Phi} \in \mathcal{N}$ is an E_{opt}-*approximate minimizer* of this problem, for some $E_{\mathsf{opt}} \geq 0$, if

$$\mathcal{G}(\hat{\Phi}) \leq E_{\mathsf{opt}} + \min_{\Phi \in \mathcal{N}} \mathcal{G}(\Phi). \tag{5.3.1}$$

We now recall the definition of a minimal monotone majorant (see §2.4.4). Let $\boldsymbol{b} = (b_i)_{i \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be a sequence, its *minimal monotone majorant* is defined by

$$\tilde{\boldsymbol{b}} = (\tilde{b}_i)_{i \in \mathbb{N}}, \quad \text{where } \tilde{b}_i = \sup_{j \geq i} |b_j|, \ \forall i \in \mathbb{N}.$$

Given $0 < p < \infty$, we recall the definition of the *monotone* ℓ^p space $\ell^p_{\mathsf{M}}(\mathbb{N})$ as

$$\ell^p_{\mathsf{M}}(\mathbb{N}) = \{\boldsymbol{b} \in \ell^\infty(\mathbb{N}) : \|\boldsymbol{b}\|_{p,\mathsf{M}} := \|\tilde{\boldsymbol{b}}\|_p < \infty\}. \tag{5.3.2}$$

5.3.1 Learning in the case of unknown anisotropy

Theorem 5.3.1 (Banach-valued learning; unknown anisotropy)**.** *There are universal constants c_0, $c_1, c_2 \geq 1$ such that the following holds. Let $m \geq 3$, $0 < \epsilon < 1$, $0 < p \leq 1/2$, $\varepsilon > 0$, ϱ be either the uniform or Chebyshev probability measure over $\mathcal{U} = [-1,1]^{\mathbb{N}}$, $\mathcal{V}$ be a Banach space, $\mathcal{V}_K \subseteq \mathcal{V}$ be a subspace of dimension K, $\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K$ be a bounded linear operator, π_K be as in (2.2.14),*

$$L = L(m, \epsilon) = \log^4(m) + \log(\epsilon^{-1}) \tag{5.3.3}$$

and

$$\Theta = [n], \quad \textit{where } n = \left\lceil \frac{m}{c_0 L} \right\rceil. \tag{5.3.4}$$

Then there exist

(a) a class $\mathcal{N}^j$ of DNNs $\Phi : \mathbb{R}^n \to \mathbb{R}^K$ with either the ReLU ($j = 1$), RePU ($j = \ell$) or tanh ($j = 0$) activation function with $\ell = 2, 3, \ldots$ and bounds for its depth and width given by

$$\mathrm{width}(\mathcal{N}^1) \leq c_{1,1} \cdot m^{3+\log_2(m)}, \qquad \mathrm{depth}(\mathcal{N}^1) \leq c_{1,2} \cdot \log(m)\Big[\log^2(m) + p^{-1}\log(m) + m\Big],$$

in the ReLU case and

$$\mathrm{width}(\mathcal{N}^j) \leq c_{j,1} \cdot m^{3+\log_2(m)}, \qquad \mathrm{depth}(\mathcal{N}^j) \leq c_{j,2} \cdot \log_2(m),$$

in the tanh ($j = 0$) or RePU ($j = \ell$) cases, where $c_{j,1}$, $c_{j,2}$ are universal constants in the ReLU and tanh cases and $c_{j,1}, c_{j,2}$ depend on $\ell = 2, 3, \ldots$ in the RePU case;

(b) a regularization function $\mathcal{J} : \mathcal{N}^j \to [0, \infty)$ equivalent to a certain norm of the trainable parameters;

(c) a choice of regularization parameter λ involving only m and ϵ;

such that the following holds for every $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\boldsymbol{b} \geq 0$. Let $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$, where $\mathcal{H}(\boldsymbol{b}, \varepsilon)$ is as in (2.3.3), draw $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m \sim_{\text{i.i.d.}} \varrho$ and consider noisy evaluations $d_i = f(\boldsymbol{y}_i) + n_i \in \mathcal{V}$, $i = 1, \ldots, m$, as in (5.1.1). Then, with probability at least $1 - \epsilon$, every E_{opt}-approximate minimizer $\hat{\Phi}$, $E_{\mathsf{opt}} \geq 0$, of the training problem

$$\min_{\Phi \in \mathcal{N}^j} \mathcal{G}(\Phi), \qquad \text{where } \mathcal{G}(\Phi) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \|f_{\Phi,\Theta}(\boldsymbol{y}_i) - d_i\|^2_{\mathcal{V}}} + \lambda \mathcal{J}(\Phi), \tag{5.3.5}$$

satisfies

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \left(E_{\mathsf{app,UB}} + m^{1/2} \cdot (E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}) \right), \tag{5.3.6}$$

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \left(E^\infty_{\mathsf{app,UB}} + m^{1/2} \cdot (E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}) \right), \tag{5.3.7}$$

where $f_{\hat{\Phi},\Theta}$ is as in (5.1.7),

$$E_{\mathsf{app,UB}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{\frac{1}{2}(1-1/p)}, \quad E^\infty_{\mathsf{app,UB}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{\frac{1}{2}(1-1/p)}, \tag{5.3.8}$$

$$E_{\mathsf{samp}} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \|n_i\|^2_{\mathcal{V}}}, \quad E_{\mathsf{disc}} = \|f - \mathcal{P}_K(f)\|_{L^\infty_\varrho(\mathcal{U};\mathcal{V})},$$

and $C = C(\boldsymbol{b}, \varepsilon, p)$ depends on $\boldsymbol{b}$, ε and p only.

Note that Theorem 5.3.1 applies to general Banach spaces. Also note that, due to the proof strategy, in Theorem 5.3.1 (also for Theorem 5.3.3) the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm and the $L^\infty(\mathcal{U};\mathcal{V})$-norm approximation errors follow the same rate. This is because, without Parseval's identity, we are forced to bound most of the significant errors in terms of their $L^\infty(\mathcal{U};\mathcal{V})$-norm. However, in the Hilbert space case we are able to improve the error bound in several ways.

Theorem 5.3.2 (Hilbert-valued learning; unknown anisotropy)**.** *Consider the setup of Theorem 5.3.1, except where $0 < p < 1$ and $\mathcal{V}$ is a Hilbert space. Then the same result holds,*

except with (5.3.6) *and* (5.3.7) *replaced by*

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \left(E_{\mathsf{app,UH}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right), \tag{5.3.9}$$

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \left(E^\infty_{\mathsf{app,UH}} + m^{1/2}(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}})\right), \tag{5.3.10}$$

with $E_{\mathsf{app,UB}}$ *and* $E^\infty_{\mathsf{app,UB}}$ *replaced by*

$$E_{\mathsf{app,UH}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1/2-1/p} \qquad E^\infty_{\mathsf{app,UH}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1-1/p},$$

respectively and potentially different values of the constants c_0, c_1, c_2, $c_{j,1}$, $c_{j,2}$ *and* $C(\boldsymbol{b}, \varepsilon, p)$.

In Theorem 5.3.2, we can obtain near-optimal rates (see Chapter 6) of the form $(m/L)^{1/2-1/p}$, where L is a polylogarithmic factor of the order $\mathcal{O}(\log^4(m))$ for the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm approximation errors. Since these results are also nonuniform (i.e., they achieve the corresponding algebraic rates for a fixed function f) and hold in the unknown anisotropy case, these algebraic rates are comparable to those of Theorem 3.3.2 and Theorem 4.3.3.

Note that Theorem 5.3.1 holds for a fixed $0 < p \leq 1/2$ and Theorem 5.3.2 holds for a fixed $0 < p < 1$. However, this assumption is only needed for the ReLU case, where, as we see, the depth of the DNN architecture behaves likes $\mathcal{O}(1/p)$ for small p. In the RePU and tanh cases, the depth of the architecture is independent of p. This means that the results in fact holds simultaneously for all $0 < p \leq 1/2$ and $0 < p < 1$, respectively, in these cases. Therefore, these activations can fully address the unknown anisotropy case. Indeed, the architectures and training procedures are completely independent of $\boldsymbol{b}$, with the assumption $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ being used only to assert a bound for the approximation error term. ReLU activations lead to schemes that depend on p, but are otherwise independent of $\boldsymbol{b}$ as well.

Upon inspection of the proof (see, e.g., (5.8.20)), we notice that allowing $0 < p \leq p^*$ for some $p^* < 1$ feasible in the case of Theorem 5.3.1 if we enlarge the search space in (5.3.4) to $n = \lceil (m/c_0 L)^{\frac{1}{2(1-p^*)}} \rceil$, resulting in larger width and depth bounds. However, to maintain these bounds as small as possible, we shall abstain from doing so.

5.3.2 Learning in the case of known anisotropy

The previous two theorems address the case of unknown anisotropy, since the DNN architecture and training strategy do not require any knowledge of the anisotropy parameter $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ (and, as noted, in the RePU and tanh cases, the parameter p). We now consider the case of known anisotropy, in which we have knowledge of $\boldsymbol{b}$ to design the architecture and training strategy.

Theorem 5.3.3 (Banach-valued learning; known anisotropy)**.** *There are universal constants* $c_0, c_1, c_2 \geq 1$ *such that the following holds. Let* $m \geq 3$, $0 < \epsilon < 1$, $0 < p \leq 1/2$,

$\varepsilon > 0$, $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with $\boldsymbol{b} \geq 0$, ϱ be either the uniform or Chebyshev probability measure over $\mathcal{U} = [-1,1]^{\mathbb{N}}$, $\mathcal{V}$ be a Banach space, $\mathcal{V}_K \subseteq \mathcal{V}$ be a subspace of dimension K, $\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K$ be a bounded linear operator, π_K be as in (2.2.14) and

$$L = L(m, \epsilon) = \log(m) + \log(\epsilon^{-1}). \tag{5.3.11}$$

Then there exist

(a) a set $\Theta \subset \mathbb{N}$ of size

$$|\Theta| = n := \left\lceil \frac{m}{c_0} \right\rceil, \tag{5.3.12}$$

(b) a class $\mathcal{N}^j$ of DNNs $\Phi : \mathbb{R}^n \to \mathbb{R}^K$ with either ReLU ($j = 1$), RePU ($j = \ell$) or tanh ($j = 0$) activation function and bounds for its depth and with given by

$$\operatorname{width}(\mathcal{N}^1) \leq c_{1,1} \cdot m^2, \qquad \operatorname{depth}(\mathcal{N}^1) \leq c_{1,1} \cdot \log(m) \left(p^{-1} \log(m) + m \right),$$

in the ReLU case and

$$\operatorname{width}(\mathcal{N}^j) \leq c_{j,1} \cdot m^2, \qquad \operatorname{depth}(\mathcal{N}^j) \leq c_{j,2} \cdot \log_2(m),$$

in the tanh ($j = 0$) or RePU ($j = \ell$) cases, where $c_{j,1}$, $c_{j,2}$ are universal constants in the ReLU and tanh cases and $c_{j,1}, c_{j,2}$ depend on $\ell = 2, 3, \ldots$ in the RePU case;

such that the following holds. Let $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$, where $\mathcal{H}(\boldsymbol{b}, \varepsilon)$ is as in (2.3.3), draw $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m \sim_{\text{i.i.d.}} \varrho$ and consider noisy evaluations $d_i = f(\boldsymbol{y}_i) + n_i \in \mathcal{V}$, $i = 1, \ldots, m$, as in (5.1.1). Then, with probability at least $1 - \epsilon$, every E_{opt}-approximate minimizer $\hat{\Phi}$, $E_{\mathsf{opt}} \geq 0$, of the training problem

$$\min_{\Phi \in \mathcal{N}^j} \mathcal{G}(\Phi), \qquad \text{where } \mathcal{G}(\Phi) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \| f_{\Phi,\Theta}(\boldsymbol{y}_i) - d_i \|_{\mathcal{V}}^2}, \tag{5.3.13}$$

satisfies

$$\| f - f_{\hat{\Phi},\Theta} \|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c_1 \Big(E_{\mathsf{app,KB}} + m \left(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}, \right) \Big), \tag{5.3.14}$$

$$\| f - f_{\hat{\Phi},\Theta} \|_{L^\infty(\mathcal{U};\mathcal{V})} \leq c_2 \left(E^\infty_{\mathsf{app,KB}} + m(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}) \right), \tag{5.3.15}$$

where $f_{\hat{\Phi},\Theta}$ is as in (5.1.7),

$$E_{\mathsf{app,KB}} = C \cdot \pi_K \cdot \left(\frac{m}{L} \right)^{2-1/p}, \quad E^\infty_{\mathsf{app,KB}} = C \cdot \pi_K \cdot \left(\frac{m}{L} \right)^{2-1/p}, \tag{5.3.16}$$

$$E_{\mathsf{samp}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \|n_i\|_{\mathcal{V}}^2}, \quad E_{\mathsf{disc}} = \|f - \mathcal{P}_K(f)\|_{L_{\varrho}^{\infty}(\mathcal{U};\mathcal{V})},$$

and $C = C(\boldsymbol{b}, \varepsilon, p)$ depends on $\boldsymbol{b}$, ε and p only. Moreover, if $\boldsymbol{b} \in \ell_{\mathsf{M}}^p(\mathbb{N})$, then the set Θ in (a) may be chosen explicitly as

$$\Theta = [n], \qquad \text{where } n = \left\lceil \frac{m}{c_0 L} \right\rceil. \tag{5.3.17}$$

Theorem 5.3.4 (Hilbert-valued learning; known anisotropy)**.** *Consider the same setup as Theorem 5.3.3, except where $0 < p < 1$ and $\mathcal{V}$ is a Hilbert space. Then the same result holds, except (5.3.14) and (5.3.15) replaced by*

$$\|f - f_{\hat{\Phi},\Theta}\|_{L_{\varrho}^2(\mathcal{U};\mathcal{V})} \leq c_1 \left(E_{\mathsf{app,KH}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right), \tag{5.3.18}$$

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^{\infty}(\mathcal{U};\mathcal{V})} \leq c_2 \left(E_{\mathsf{app,KH}}^{\infty} + m^{1/2}\left(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right)\right), \tag{5.3.19}$$

with $E_{\mathsf{app,KB}}$ and $E_{\mathsf{app,KB}}^{\infty}$ replaced by

$$E_{\mathsf{app,KH}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1/2-1/p} \qquad E_{\mathsf{app,KH}}^{\infty} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1-1/p},$$

and potentially different values of the constants c_0, c_1, c_2, $c_{j,1}$, $c_{j,2}$ and $C(\boldsymbol{b}, \varepsilon, p)$.

In Theorem 5.3.4, we can obtain near-optimal rates (see Chapter 6) of the form $(m/L)^{1/2-1/p}$, where L is now a polylogarithmic factor of the order $\mathcal{O}(\log(m))$ for the $L_{\varrho}^2(\mathcal{U};\mathcal{V})$-norm approximation errors. Theorems 5.3.3 and 5.3.4 are also nonuniform (achieve the corresponding algebraic rates for a fixed function f) as Theorem 3.3.2 and Theorem 4.3.3. However, note that the last two are theorems in the unknown anisotropy setting.

5.4 Discussion

We now comment on five important aspects of the main results.

The various approximation errors

These errors deserve additional discussion. The term E_{app} decays algebraically fast in m/L, where L is a (poly)logarithmic factor in m. In particular, these take the form

$$E_{\mathsf{app,UB}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{\frac{1}{2}(1-1/p)}, \qquad E_{\mathsf{app,KB}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{2-1/p},$$

$$E_{\mathsf{app,UH}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1/2-1/p}, \qquad E_{\mathsf{app,KH}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1/2-1/p},$$

and

$$E^{\infty}_{\mathsf{app},\mathsf{UB}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{\frac{1}{2}(1-1/p)}, \qquad E^{\infty}_{\mathsf{app},\mathsf{KB}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{2-1/p},$$
$$E^{\infty}_{\mathsf{app},\mathsf{UH}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1-1/p}, \qquad E^{\infty}_{\mathsf{app},\mathsf{KH}} = C \cdot \pi_K \cdot \left(\frac{m}{L}\right)^{1-1/p},$$

where L is roughly $\log(m)$ in the known anisotropy case and $\log^4(m)$ in the unknown anisotropy case. This discrepancy arises because of the proof strategy. In the known anisotropy setting we emulate a polynomial least-squares scheme via a DNN. Conversely, in the unknown anisotropy setting we emulate a polynomial (weighted) ℓ^1-minimization scheme, with significantly more intricate analysis via compressed sensing techniques.

The curse of dimensionality

In all cases, we overcome the curse of dimensionality in the sample complexity. Note that the term $E_{\mathsf{app},\mathsf{UH}}$ is the same as what was shown in Chapter 4 for polynomial approximation via compressed sensing of Hilbert-valued functions. Further, both the terms $E_{\mathsf{app},\mathsf{KH}}$ and $E_{\mathsf{app},\mathsf{UH}}$ are optimal up to constants and the logarithmic factors (see Chapter 6). By contrast, the bounds for Banach-valued functions are worse in both cases. Besides not having Parseval's identity, this arises as a consequence of just having the duality pair and not a proper inner product in the proofs, which is a key step in the proofs that lifts the weighted rNSP over $\mathbb{R}$ to the Banach space $\mathcal{V}$ (see Lemma 3.6.7). We expect it may be possible to improve these rates via a different argument.

Architecture

Another key difference between the known and unknown anisotropy cases is the width of the DNN architecture. Here, we overcome the curse of dimensionality in the DNN architecture $\mathcal{N}$. For the RePU and tanh activation functions, it is polynomial in m in the former case; specifically, $\mathcal{O}(m^2)$ for large m. In the latter case, it behaves like $\mathcal{O}(m^{3+\log_2(m)})$, i.e., faster than polynomial, but still subexponential in m. This discrepancy arises from having to include many more coordinate variables in the unknown anisotropy case to guarantee the desired approximation error. Conversely, in the known anisotropy case we may restrict only to those variables that are known to be important. In the ReLU case the depth of the fully-connected architecture has explicit dependence on the smoothness of f and polylogarithmic-linear scaling in m.

Monotonicity of $\boldsymbol{b}$

The unknown anisotropy case also involves the stronger assumption $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$, a condition that was also encountered in Chapters Chapters 3 and 4. This assumption means that the

variables are, on average, ordered in terms of importance. In fact, it is impossible in the unknown anisotropy setting to learn functions with only the assumption $\boldsymbol{b} \in \ell^p(\mathbb{N})$. This is a key result we show in Chapter 6. By contrast, in the known anisotropy case we may in fact assume that $\boldsymbol{b} \in \ell^p(\mathbb{N})$. Yet we do not have any control over the set Θ that defines the truncation operator in this case, except for its size. However, if we suppose that $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ then we may choose Θ explicitly as in (5.3.17).

The handcrafted architecture

Our main results should be interpreted as a primarily theoretical contribution, i.e., showing the existence of DNN architectures and training strategies (similar to those used in practice) for learning such functions from limited datasets that are near-optimal in terms of the amount of training data m. We remark in passing that in our DNNs only the parameters in the final layer are trained, which results in a convex optimization problem (see §5.5.3–5.5.4). The other parameters are *handcrafted* and designed to (approximately) emulate suitable orthonormal polynomials. A consequence of this approach is that the ensuing DNN training strategies are not expected to yield superior performance over the corresponding (least-squares or compressed sensing-based) polynomial approximation procedures.

This provides some justification for the application of DL to parametric PDEs, where superior performance over state-of-art techniques, including polynomial-based methods, has been recently observed [7]. Having said that, our results do also provide credence to various empirical observations about DL for these applications. First, it has been observed that ReLU activations often lead to worse practical performance with similar-sized architectures than smoother activations. In our setting, we require deeper and wider ReLU DNNs to obtain the same rates. Second, it has been observed that width is more important than depth in such applications. This broadly agrees with our width and depth bounds (5.2.2). See also [91] for similar discussion.

However, it is important to stress that there still remains a substantial gap between theory and practice. Practical DL strategies train all (or most) layers via a nonconvex optimization problem and typically employ simpler and smaller architectures (see, e.g., the bounds in (5.2.2)). Yet, they currently lack theoretical guarantees. Further narrowing this gap is an interesting objective for future work. For a practical implementation of DNNs on Banach-valued function approximation see Chapter 7.

Near-optimality

In this chapter, we utilize the results from Chapter 6 to assert the near-optimality of the constructions of our manufactured architecture. However, Chapter 6 only considers the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm. Therefore, we cannot claim that the $L^\infty(\mathcal{U};\mathcal{V})$-norm bounds we obtained are near–optimal. However, in the Hilbert-valued case we obtain the rate $(m/L)^{1-1/p}$ in the

$L^\infty(\mathcal{U};\mathcal{V})$-norm approximation error, which is the same as that obtained, for instance, in Theorems 3.3.2 and Theorem 4.3.3. Proving that these rates are the optimal rates in the $L^\infty(\mathcal{U};\mathcal{V})$-norm is still an open problem.

5.5 Formulating the training problems and proof strategy

The remainder of this chapter is devoted to set up and formulating the training problems and proofs of the main results. As mentioned in §5.2, we first formulate learning problems for Banach-valued functions using orthogonal polynomials, and then use DNNs to emulate these polynomials. The main theorems are then obtained using techniques from compressed sensing theory.

Specifically, we proceed as follows. In this section, we reformulate the problem as a recovery problem for Banach-valued vectors (see §2.5). We then introduce the class of DNNs considered and formulate separate learning problems in the known and unknown anisotropy cases. Then in §5.7 we describe how to emulate polynomials using DNNs, and give bounds for the width and depths of the corresponding architectures. Finally, with the necessary tools in place, in §5.8 we give the proofs of the main results.

5.5.1 Formulation as a vector recovery problem

Observe that the formulation as a vector follows the same steps as that of §2.5. In the following we recall some main aspects. Let $m \in \mathbb{N}$, $\boldsymbol{y}_1,\dots,\boldsymbol{y}_m \in \mathcal{U}$ be the sample points and $f \in L^2_\varrho(\mathcal{U};\mathcal{V})$ be a continuous function. From (2.5.1), let the normalized measurement matrix taking values in $\mathbb{R}$ and the measurement and error vectors by

$$\boldsymbol{A} = \left(\frac{\Psi_{\boldsymbol{\nu}_j}(\boldsymbol{y}_i)}{\sqrt{m}}\right)_{i,j=1}^{m,N} \in \mathbb{R}^{m\times N}, \quad \boldsymbol{f} = \frac{1}{\sqrt{m}}\left(f(\boldsymbol{y}_i)+n_i\right)_{i=1}^m \in \mathcal{V}^m \text{ and } \boldsymbol{e} = \frac{1}{\sqrt{m}}(n_i)_{i=1}^m \in \mathcal{V}^m. \tag{5.5.1}$$

We also define the truncated expansion of f based on the index set Λ and its corresponding vector of coefficients as in (2.5.2). Notice that the matrix $\boldsymbol{A} = (a_{i,j})_{i,j=1}^{m,N}$ immediately extends to a bounded linear operator as in (2.5.3). With this in hand, from §2.5 we recall that

$$\boldsymbol{A}\boldsymbol{c}_\Lambda = \frac{1}{\sqrt{m}}\left(f_\Lambda(\boldsymbol{y}_i)\right)_{i=1}^m = \frac{1}{\sqrt{m}}(f(\boldsymbol{y}_i))_{i=1}^m - \frac{1}{\sqrt{m}}\left(f(\boldsymbol{y}_i)-f_\Lambda(\boldsymbol{y}_i)\right)_{i=1}^m,$$

and therefore

$$\boldsymbol{A}\boldsymbol{c}_\Lambda + \boldsymbol{e} + \widetilde{\boldsymbol{e}} = \boldsymbol{f}, \quad \text{where} \quad \widetilde{\boldsymbol{e}} = \frac{1}{\sqrt{m}}\left(f(\boldsymbol{y}_i)-f_\Lambda(\boldsymbol{y}_i)\right)_{i=1}^m.$$

Hence, vector $\boldsymbol{c}_\Lambda$ of unknown coefficients is a solution of the previous noisy linear system.

5.5.2 The class of DNNs $\mathcal{N}$ and the approximate measurement matrix

We now precisely define the class of DNNs and the approximate measurement matrix that emulates $\boldsymbol{A}$. First, fix $\Theta \subset \mathbb{N}$, $|\Theta| = n$ and let $\Phi_{\boldsymbol{\nu},\delta,\Theta} = \Phi_{\boldsymbol{\nu},\delta} \circ \mathcal{T}_\Theta$ be a DNN approximation to the basis function $\Psi_{\boldsymbol{\nu}}$ for $\boldsymbol{\nu} \in \Lambda$, where $\Phi_{\boldsymbol{\nu},\delta} : \mathbb{R}^n \to \mathbb{R}$ is a DNN of the form (2.6.1) and $\mathcal{T}_\Theta$ is as in (5.1.3). The term $\delta > 0$ is a parameter that controls the accuracy of the approximation $\Phi_{\boldsymbol{\nu},\delta,\Theta} \approx \Psi_{\boldsymbol{\nu}}$. These definitions will be useful later in the proofs. Now let

$$\Phi_{\Lambda,\delta} : \mathbb{R}^n \to \mathbb{R}^N, \bar{\boldsymbol{y}} \mapsto (\Phi_{\boldsymbol{\nu},\delta}(\bar{\boldsymbol{y}}))_{\boldsymbol{\nu}\in\Lambda}, \qquad \Phi_{\Lambda,\delta,\Theta} = \Phi_{\Lambda,\delta} \circ \mathcal{T}_\Theta, \quad \forall \bar{\boldsymbol{y}} = (\bar{y}_j)_{j\in\Theta} \in \mathbb{R}^n,$$

and define the class of DNNs $\mathcal{N}$ by

$$\mathcal{N} = \left\{ \Phi : \mathbb{R}^n \to \mathbb{R}^K : \Phi_\Theta(\bar{\boldsymbol{y}}) = \boldsymbol{Z}^\top \Phi_{\Lambda,\delta}(\bar{\boldsymbol{y}}),\ \boldsymbol{Z} \in \mathbb{R}^{N\times K}, \bar{\boldsymbol{y}} = (\bar{y}_j)_{j\in\Theta} \in \mathbb{R}^n \right\}, \tag{5.5.2}$$

where $\boldsymbol{Z} \in \mathbb{R}^{N\times K}$ is the matrix of trainable parameters. We now also define the approximate measurement matrix $\boldsymbol{A}' \approx \boldsymbol{A}$ by

$$\boldsymbol{A}' = \left(\frac{\Phi_{\boldsymbol{\nu}_j,\delta,\Theta}(\boldsymbol{y}_i)}{\sqrt{m}} \right)_{i,j=1}^{m,N} \in \mathbb{R}^{m\times N}. \tag{5.5.3}$$

5.5.3 Unknown anisotropy recovery

As discussed in §2.4.6, choosing an appropriate index set is a vital step towards obtaining the desired approximation rates in §5.3. Recall the definition of Λ_n^{HCI} in (2.4.21), which is isomorphic to the n-dimensional *hyperbolic cross* index set of order $n-1$. Notice from (2.4.23) that

$$N := |\Lambda_n^{\mathsf{HCI}}| \leq e n^{2+\log(n)/\log(2)}, \quad \forall n \in \mathbb{N}. \tag{5.5.4}$$

Let $\boldsymbol{w} = \boldsymbol{u} \geq \boldsymbol{1}$ be the so-called *intrinsic weights*, given by (2.4.17). We now construct the DNN training problem considered in Theorem 5.3.1. As in §2.5.2, we consider the Banach-valued, weighted SR-LASSO optimization problem

$$\min_{\boldsymbol{z}\in\mathcal{V}_K^N} \mathcal{G}(\boldsymbol{z}), \qquad \mathcal{G}(\boldsymbol{z}) := \lambda \|\boldsymbol{z}\|_{1,\boldsymbol{u};\mathcal{V}} + \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}. \tag{5.5.5}$$

As in (2.5.6) $\lambda > 0$ is a tuning parameter and $\boldsymbol{A}$ and $\boldsymbol{f}$ are as in (5.5.1). To obtain a DNN training problem, we replace $\boldsymbol{A}$ with its approximation $\boldsymbol{A}'$, defined by (5.5.3), giving the optimization problem

$$\min_{\boldsymbol{z}\in\mathcal{V}_K^N} \mathcal{G}'(\boldsymbol{z}), \qquad \mathcal{G}'(\boldsymbol{z}) := \lambda \|\boldsymbol{z}\|_{1,\boldsymbol{u};\mathcal{V}} + \|\boldsymbol{A}'\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}. \tag{5.5.6}$$

To show that (5.5.6) is equivalent to a DNN training problem we argue as follows. Let $\{\varphi_i\}_{i=1}^K$ be the basis of $\mathcal{V}_K$ and $\boldsymbol{z} = (z_{\boldsymbol{\nu}_j})_{j=1}^N$ be an arbitrary element of $\mathcal{V}_K^N$. Now, recall

that $\mathcal{N}$ is the class of DNNs defined in (5.5.2). Then, we can associate $\boldsymbol{z}$ with a DNN $\Phi \in \mathcal{N}$ via the relation

$$\Phi = \boldsymbol{Z}^\top \Phi_{\Lambda,\delta}, \quad \text{where } \boldsymbol{Z} = (Z_{j,k})_{j,k=1}^{N,K} \in \mathbb{R}^{N\times K} \text{ is such that } z_{\boldsymbol{\nu}_j} = \sum_{k=1}^{K} Z_{j,k}\varphi_k, \ \forall j \in [N]. \tag{5.5.7}$$

Now observe that

$$\begin{aligned} f_{\Phi,\Theta}(\boldsymbol{y}) &= \sum_{k=1}^{K} ((\Phi \circ \mathcal{T}_\Theta)(\boldsymbol{y}))_k \varphi_k = \sum_{k=1}^{K} (\boldsymbol{Z}^\top \Phi_{\Lambda,\delta,\Theta}(\boldsymbol{y}))_k \varphi_k \\ &= \sum_{k=1}^{K}\sum_{j=1}^{N} Z_{j,k} \Phi_{\boldsymbol{\nu}_j,\delta,\Theta}(\boldsymbol{y})\varphi_k = \sum_{\boldsymbol{\nu}\in\Lambda} z_{\boldsymbol{\nu}} \Phi_{\boldsymbol{\nu},\delta,\Theta}(\boldsymbol{y}). \end{aligned}$$

Hence, if $d_i = f(\boldsymbol{y}_i) + n_i \in \mathcal{V}$ are the noisy evaluations of f, then

$$\|\boldsymbol{A}'\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \left\| \sum_{\boldsymbol{\nu}\in\Lambda} z_{\boldsymbol{\nu}} \Phi_{\boldsymbol{\nu},\delta,\Theta}(\boldsymbol{y}_i) - d_i \right\|_{\mathcal{V}}^2} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \|f_{\Phi,\Theta}(\boldsymbol{y}_i) - d_i\|_{\mathcal{V}}^2}.$$

Now let $\mathcal{J} : \mathcal{N} \to [0,\infty)$ be the regularization functional defined by

$$\|\boldsymbol{z}\|_{1,\boldsymbol{u};\mathcal{V}} = \sum_{j=1}^{N} u_{\boldsymbol{\nu}_j} \|z_{\boldsymbol{\nu}_j}\|_{\mathcal{V}} = \sum_{j=1}^{N} u_{\boldsymbol{\nu}_j} \left\| \sum_{k=1}^{K} Z_{j,k}\varphi_k \right\|_{\mathcal{V}} := \mathcal{J}(\Phi),$$

where $\Phi \in \mathcal{N}$ is as in (5.5.7). Clearly $\mathcal{J}$ is a norm over the trainable parameters, as claimed. Using this and the previous relation, we deduce that (5.5.6) is equivalent to the DNN training problem

$$\min_{\Phi\in\mathcal{N}} \sqrt{\frac{1}{m}\sum_{i=1}^{m} \|f_{\Phi,\Theta}(\boldsymbol{y}_i) - d_i\|_{\mathcal{V}}^2} + \lambda \mathcal{J}(\Phi).$$

By this, we mean that every minimizer $\hat{\Phi} \in \mathcal{N}$ of this problem corresponds to a minimizer $\hat{\boldsymbol{z}}$ of (5.5.6) via the relation (5.5.7), and vice versa.

5.5.4 Known anisotropy recovery

Recall the discussion in §2.5.3. Analogously to §2.5, we define the normalized measurement matrix and the approximate normalized measurement matrix by

$$\boldsymbol{A} = \left(\frac{\Psi_{\boldsymbol{\nu}_j}(\boldsymbol{y}_i)}{\sqrt{m}} \right)_{i,j=1}^{m,s} \in \mathbb{R}^{m\times s}, \quad \text{and} \quad \boldsymbol{A}' := \left(\frac{\Phi_{\boldsymbol{\nu}_j,\Theta}(\boldsymbol{y}_i)}{\sqrt{m}} \right)_{i,j=1}^{m,s} \in \mathbb{R}^{m\times s}, \tag{5.5.8}$$

where $\{\boldsymbol{\nu}_1,\dots,\boldsymbol{\nu}_s\}$ is an ordering of S. Likewise, we truncate the expansion of f and its vector coefficients based on (2.5.2) for the index set S. Defining the class of DNNs $\mathcal{N}$ as in

(5.5.2), except with Λ and N replaced by S and s, respectively, we now see that the training problem (5.3.13) can be expressed as the Banach-valued minimization problem

$$\min_{\boldsymbol{z}\in\mathcal{V}_K^s} \mathcal{G}'(\boldsymbol{z}), \qquad \mathcal{G}'(\boldsymbol{z}) := \|\boldsymbol{A}'\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}, \tag{5.5.9}$$

where $\boldsymbol{f} = \frac{1}{\sqrt{m}}(d_i)_{i=1}^m \in \mathcal{V}^m$. To be precise, any $\hat{\boldsymbol{z}} = (\hat{z}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in S}$ that is a minimizer of (5.5.9) defines a minimizer $\hat{\Phi}$ of (5.3.13) via the relation (5.5.7), except with Λ and N replaced by S and s, respectively. As before, we also consider (5.5.9) as an approximation to a minimization problem with matrix $\boldsymbol{A}$ for the polynomial coefficients $\boldsymbol{c}_S$:

$$\min_{\boldsymbol{z}\in\mathcal{V}_K^s} \mathcal{G}(\boldsymbol{z}), \qquad \mathcal{G}(\boldsymbol{z}) := \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{f}\|_{2;\mathcal{V}}.$$

To end this section, it is worth mentioning here that, for ease of notation we denote $\mathcal{G}$ and $\mathcal{G}'$ as the objective function for both the known and unknown anisotropy case. Mathematically, the known anisotropy setting is just a particular case of the unknown anisotropy setting with $\lambda = 0$ and Λ and N replaced by S and s, respectively.

5.6 Matrices satisfying the weighted rNSP over Banach spaces

We now assert conditions on m under which the measurement matrix (2.5.1) satisfies the wRIP. For this, we use the following result, which is Lemma 3.7.1 applied to the infinite-dimensional case only.

Lemma 5.6.1 (wRIP for Chebyshev and Legendre polynomials)**.** *Let ϱ be the tensor-product uniform or Chebyshev measure on $\mathcal{U} = [-1,1]^{\mathbb{N}}$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$ be the corresponding tensor-product orthonormal Legendre or Chebyshev polynomial basis of $L^2_\varrho(\mathcal{U})$, $\Lambda = \Lambda_n^{\mathsf{HCI}}$ be as in (2.4.21) for some $n \geq 1$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ be drawn independently and identically from the measure ϱ. Let c_0 be a universal constant, $0 < \delta, \epsilon < 1$ and $k \geq 1$, suppose that*

$$m \geq c_0 \cdot \delta^{-2} \cdot k \cdot \left(\log^2(k/\delta) \cdot \log^2(en) + \log(2/\epsilon)\right), \tag{5.6.1}$$

then, with probability at least $1-\epsilon$, the matrix $\boldsymbol{A}$ defined in (2.5.1) satisfies the wRIP over $\mathbb{R}$ of order $(k, \boldsymbol{u})$ with constant $\delta_{k,\boldsymbol{u}}$, where $\boldsymbol{u}$ are the intrinsic weights (2.4.17).

Recall from Remark 3.7.2 that the previous lemma only considers the case $k \geq 1$. The case $k < 1$ is trivially satisfied.

The previous result will be used in the case of unknown anisotropy. In the case of known anisotropy, we require a different argument in order to obtain the better scaling with respect to m in Theorem 5.3.3. As we see later, this case can be analyzed by asserting conditions

under which the relevant measurement matrix $\boldsymbol{A}$ defined in (5.5.8) has the rNSP of 'full' order $k = |S|_{\boldsymbol{u}}$, where S is the index set used to construct $\boldsymbol{A}$.

Lemma 5.6.2 (Weighted rNSP for Chebyshev and Legendre polynomials in the 'full' case)**.** *Let ϱ be the tensor-product uniform or Chebyshev measure on $\mathcal{U} = [-1,1]^{\mathbb{N}}$, $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$ be the corresponding tensor-product orthonormal Legendre or Chebyshev polynomial basis of $L^2_\varrho(\mathcal{U})$, $S \subset \mathcal{F}$ and $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ be drawn independently and identically from the measure ϱ. Let $0 < \delta, \epsilon < 1$, $k = |S|_{\boldsymbol{u}}$, where $\boldsymbol{u}$ are the intrinsic weights (2.4.17), and suppose that*

$$m \geq ((1-\delta)\log(1-\delta) + \delta)^{-1} \cdot k \cdot \log(k/\epsilon). \tag{5.6.2}$$

Then, with probability at least $1-\epsilon$, the matrix $\boldsymbol{A} \in \mathbb{R}^{m\times s}$, $s = |S|$, defined in (5.5.8) satisfies the weighted rNSP over $\mathbb{R}$ of order $(k, \boldsymbol{u})$ with constants $\rho = 0$ and $\gamma \leq (1-\delta)^{-1/2}$.

Proof. Since $k = |S|_{\boldsymbol{u}}$ is the weighted cardinality of the 'full' index set S, the wrNSP is equivalent to the condition

$$\|\boldsymbol{x}\|_2 \leq \gamma \|\boldsymbol{A}\boldsymbol{x}\|_2, \quad \forall \boldsymbol{x} \in \mathbb{R}^s.$$

Define the space $\mathcal{P} = \operatorname{span}\{\Psi_{\boldsymbol{\nu}} : \boldsymbol{\nu} \in S\} \subset L^2_\varrho(\mathcal{U})$. Then by Parseval's identity, this inequality is equivalent to

$$\|p\|_{L^2_\varrho(\mathcal{U})} \leq \gamma \|p\|_{\mathsf{disc}}, \quad \forall p \in \mathcal{P},$$

where $\|p\|_{\mathsf{disc}} = \sqrt{\frac{1}{m}\sum_{i=1}^m |p(\boldsymbol{y}_i)|^2}$. Thus, by [12, §5.2] and [12, Thm. 5.7], we have $\gamma \leq (1-\delta)^{-1/2}$, provided

$$m \geq ((1-\delta)\log(1-\delta) + \delta)^{-1} \cdot \kappa \cdot \log(s/\epsilon), \tag{5.6.3}$$

where $s = |S|$ and $\kappa = \kappa(\mathcal{P})$ is defined by (see [12, Eqn. (5.15)])

$$\kappa = \sup_{\boldsymbol{y}\in\mathcal{U}} \sum_{\boldsymbol{\nu}\in S} |\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})|^2.$$

Observe that $\kappa \leq \sum_{\boldsymbol{\nu}\in S} u_{\boldsymbol{\nu}}^2 = |S|_{\boldsymbol{u}} = k$ and $s = |S| \leq |S|_{\boldsymbol{u}} = k$, since $\boldsymbol{u} \geq \boldsymbol{1}$. Hence (5.6.3) is implied by (5.6.2). This gives the result. □

5.7 Deep neural network approximation

In this section we detail the second key component of our proofs, which is the approximation of the orthonormal polynomials $\Psi_{\boldsymbol{\nu}}$ by DNNs.

5.7.1 Approximate multiplication via DNNs

Our results are based on three different different DNN architectures (tanh, ReLU and RePU) that emulate the product of n numbers. The first two follow from [91] and [215], respectively. The third is based on [179] and [215].

Lemma 5.7.1 (Approximate multiplication of n numbers by ReLU and tanh DNNs)**.** *Let $0 < \delta < 1$, $n \in \mathbb{N}$ and consider constants $\{M_i\}_{i=1}^n \subset \mathbb{R}_+^n$. Then there exists a ReLU ($j = 1$) or a tanh ($j = 0$) DNN $\Phi_\delta^j : \prod_{i=1}^n [-M_i, M_i] \to \mathbb{R}$ satisfying*

$$\sup_{|x_i| \leq M_i} \left| \prod_{i=1}^n x_i - \Phi_\delta^j(\boldsymbol{x}) \right| \leq \delta, \qquad \text{where } \boldsymbol{x} = (x_i)_{i=1}^n, \tag{5.7.1}$$

for $j \in \{0, 1\}$. The width and depth are bounded by

$$\begin{aligned} &\mathrm{width}(\Phi_\delta^1) \leq c_{1,1} \cdot n, \\ &\mathrm{depth}(\Phi_\delta^1) \leq c_{1,2} \Big(1 + \log(n) \Big[\log(n\delta^{-1}) + \log(M) \Big] \Big), \end{aligned}$$

in the ReLU case, where $M = \prod_i^n M_i$ and

$$\mathrm{width}(\Phi_\delta^0) \leq c_{1,1} \cdot n, \quad \mathrm{depth}(\Phi_\delta^0) \leq c_{1,2} \cdot \log_2(n),$$

in the tanh ($j = 0$) case and $c_{j,1}$, $c_{j,2}$ are universal constants for $j \in \{0, 1\}$.

Proof. First, in the tanh case, let $N \geq \max_{i \in [n]} \{M_i\}$ be such that $\boldsymbol{x} \in [-N, N]^n$. Then the result in the tanh case is a direct application of [91, Lem. 3.8]. We now focus on the ReLU case. Let $M_i > 0$ such that $|x_i| \leq M_i$ for all $i \in [n]$. Now, write the multiplication of these terms as

$$\prod_{i=1}^n x_i = \left(\prod_{i=1}^n x_i / M_i \right) \cdot \prod_{i=1}^n M_i = \left(\prod_{i=1}^n x_i / M_i \right) \cdot M.$$

Then, using [215, Prop. 2.6], for any $\widetilde{\delta} \leq \delta$ there exists a ReLU DNN $\Phi_{\widetilde{\delta},1}$ such that

$$\left| M \left(\prod_{i=1}^n \frac{x_i}{M_i} \right) - M \cdot \Phi_{\widetilde{\delta},1} \left(\frac{x_1}{M_1}, \ldots, \frac{x_n}{M_n} \right) \right| \leq M\widetilde{\delta}.$$

We now set $\widetilde{\delta} = \delta / M$. Since the composition of affine maps is an affine map, we can define a ReLU DNN of the same architecture as

$$\widetilde{\Phi}_{\delta,M}(x_1, \ldots, x_n) = M \cdot \Phi_{\delta/M,1} \left(\frac{x_1}{M_1}, \ldots, \frac{x_n}{M_n} \right). \tag{5.7.2}$$

This implies that

$$\left| \prod_{i=1}^n x_i - \widetilde{\Phi}_{\delta,M}(x_1, \ldots, x_n) \right| \leq \delta.$$

Taking the supremum over $|x_i| \leq M_i$ we obtain (5.7.1). We now bound the width and depth. From [215, Prop. 2.6] notice that there exists a constants $c_2 > 0$ such that

$$\operatorname{depth}(\widetilde{\Phi}_{\delta,M}) \leq \operatorname{depth}(\Phi_{\delta/M,1}) \leq c_2 \left(1 + \log(n) \log(nM/\delta)\right).$$

Next, from the construction of the DNN for the product of n numbers as a binary tree in [235, §3.2], observe that the product of two numbers involves a maximum of 12 nodes per layer. Thus, for the product of n numbers, the width is bounded by

$$\operatorname{width}(\widetilde{\Phi}_{\delta,M}) \leq 12 \left\lceil \frac{n}{2} \right\rceil.$$

This completes the proof. □

The following lemma asserts the existence of a RePU DNN to calculate the multiplication of two numbers. Its proof can be found in [179, Lem. 2.1] (see also [215, Appx. A] and [179, Thm. 2.5]).

Lemma 5.7.2 (Exact multiplication of two numbers by a RePU DNN)**.** *For $\ell = 2, 3, \ldots$, there exists a RePU DNN $\bar{\Phi}^\ell : \mathbb{R}^2 \to \mathbb{R}$ such that*

$$\bar{\Phi}^\ell(x_1, x_2) = x_1 x_2, \quad \forall x_1, x_2 \in \mathbb{R}.$$

The width and depth of this DNN are

$$\operatorname{width}(\bar{\Phi}^\ell) = c_\ell, \quad \operatorname{depth}(\bar{\Phi}^\ell) = 1,$$

where c_ℓ is a constant depending on ℓ.

We now use similar arguments to those in [215, §2.3], and notice that the previous lemma implies the existence of RePU DNN for multiplying n different numbers. The next lemma and its proof are based on [215, Prop. 2.6], which, in turn, employs techniques from [235, Prop. 3.3]. Basically, the idea here is to construct a DNN Φ^ℓ as a binary tree of $\bar{\Phi}^\ell$-networks using Lemma 5.7.2. Unlike in the ReLU and tanh cases (see Lemma 5.7.1), the resulting multiplication is exact and we do not require the assumption $|x_i| \leq M_i$, $i \in [n]$.

Lemma 5.7.3 (Exact multiplication of n numbers by RePU)**.** *For $\ell = 2, 3, \ldots$, there exists a RePU DNN $\Phi^\ell : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\Phi^\ell(\boldsymbol{x}) = \prod_{i=1}^{n} x_i, \quad \forall \boldsymbol{x} = (x_i)_{i=1}^n \in \mathbb{R}^n. \tag{5.7.3}$$

The width and depth are bounded by

$$\operatorname{width}(\Phi^\ell) \leq c_{\ell,1} \cdot n, \quad \operatorname{depth}(\Phi^\ell) \leq c_2 \log_2(n),$$

where $c_{\ell,1}$ and c_2 are positive constants and only $c_{\ell,1}$ depends on ℓ.

Proof. Let $\widetilde{n} := \min\{2^k : k \in \mathbb{N}, 2^k \geq n\}$. For every $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we define the vector $\widetilde{\boldsymbol{x}} = (x_1, \ldots, x_n, x_{n+1}, \ldots, x_{\widetilde{n}}) \in \mathbb{R}^{\widetilde{n}}$, where $x_{n+1} = \ldots = x_{\widetilde{n}} = 1$. Observe that the map $\boldsymbol{x} \mapsto \widetilde{\boldsymbol{x}}$ is affine and, hence, can be implemented by a suitable choice of weights and biases in the first layer. Arguing as in [215, Prop. 2.6], let $l \in \mathbb{N}$, consider vectors in $\mathbb{R}^{2l}$ and define the mapping

$$R_l(z_1, \ldots, z_{2l}) := \left(\bar{\Phi}^\ell(z_1, z_2), \ldots, \bar{\Phi}^\ell(z_{2l-1}, z_{2l})\right) \in \mathbb{R}^l,$$

where $\bar{\Phi}^\ell$ is as in Lemma 5.7.2. Now, for $k \in \mathbb{N}$ we consider the composition

$$\mathcal{R}^k := R_1 \circ R_2 \circ R_{2^2} \circ \ldots \circ R_{2^{k-1}}. \tag{5.7.4}$$

Keeping this in mind we now define $\Phi^\ell : \mathbb{R}^n \to \mathbb{R}$ by

$$\Phi^\ell(x_1, \ldots, x_n) := \mathcal{R}^{\log_2(\widetilde{n})}(x_1, \ldots, x_{\widetilde{n}}).$$

Observe that in this case there are $\log_2(\widetilde{n})$ terms in (5.7.4), then $\mathcal{R}^{\log_2(\widetilde{n})}$ and Φ^ℓ are well defined. We immediately deduce (5.7.3). We now estimate the depth and width of Φ^ℓ. First, note that $\widetilde{n} \leq 2n$. Then, following [215, Prop. 2.6] and using Lemma 5.7.2, there exists a positive constant c_2 such that

$$\operatorname{depth}(\Phi^\ell) \leq c_2 \log_2(n).$$

Moreover, from Lemma 5.7.2 and [179, Lem. 2.5], the construction of the DNN implies the existence of a constant $c_\ell > 0$ depending on ℓ such that

$$\operatorname{width}(\Phi^\ell) \leq c_\ell \cdot n.$$

Thus the bound for the width holds.

□

5.7.2 Emulation of orthogonal polynomials via DNNs

Having shown that DNNs can emulate the multiplication of n numbers, we are now able to emulate orthonormal polynomials. We do this by appealing to the fundamental theorem of algebra, which allows us to represent any such polynomial as a product of its roots. This approach is based on [87], which considered only ReLU DNNs and Legendre polynomials. This approach differs from other constructions (see, e.g., [215, Prop. 2.13]) which first emulate univariate orthogonal polynomials and then use the previously derived multiplication networks.

Theorem 5.7.4. *Let $\Lambda \subset \mathcal{F}$ be a finite multi-index set, $m(\Lambda) = \max_{\boldsymbol{\nu}\in\Lambda} \|\boldsymbol{\nu}\|_1$ and $\Theta \subset \mathbb{N}$, $|\Theta| = n$, satisfy*

$$\bigcup_{\boldsymbol{\nu}\in\Lambda} \operatorname{supp}(\boldsymbol{\nu}) \subseteq \Theta.$$

Let $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}} \subset L^2_\varrho(\mathcal{U})$ be the orthonormal Legendre or Chebyshev polynomial basis of $L^2_\varrho(\mathcal{U})$. Then for every $0 < \delta < 1$ there exists a ReLU ($j = 1$), RePU ($j = \ell$) or tanh ($j = 0$) DNN $\Phi^j_{\Lambda,\delta} : \mathbb{R}^n \to \mathbb{R}^{|\Lambda|}$, such that, if $\Phi^j_{\Lambda,\delta}(\boldsymbol{z}) = (\Phi^j_{\boldsymbol{\nu},\delta}(\boldsymbol{z}))_{\boldsymbol{\nu}\in\Lambda}$, $\boldsymbol{z} = (z_j)_{j\in\Theta} \in \mathbb{R}^n$ and $\mathcal{T}_\Theta$ is as in (5.1.3), then

$$\|\Psi_{\boldsymbol{\nu}} - \Phi^j_{\boldsymbol{\nu},\delta} \circ \mathcal{T}_\Theta\|_{L^\infty(\mathcal{U})} \leq \delta, \quad \forall \boldsymbol{\nu} \in \Lambda, \quad j \in \{0, 1, \ell\}.$$

In the case of the ReLU ($j = 1$) activation function, the width and depth of this DNN satisfy

$$\begin{aligned}
&\operatorname{width}(\Phi^1_{\boldsymbol{\nu}}) \leq c_{1,1} \cdot |\Lambda| \cdot m(\Lambda), \\
&\operatorname{depth}(\Phi^1_{\boldsymbol{\nu}}) \leq c_{1,2} \cdot \Big(1 + \log(m(\Lambda)) \cdot \Big[\log(m(\Lambda)\delta^{-1}) + m(\Lambda) + n\Big]\Big).
\end{aligned}$$

In the case of the RePU ($j = \ell$) or tanh ($j = 0$) activation function, the width and depth of this DNN satisfy

$$\operatorname{width}(\Phi^j_{\Lambda,\delta}) \leq c_{j,1} \cdot |\Lambda| \cdot m(\Lambda), \quad \operatorname{depth}(\Phi^j_{\Lambda,\delta}) \leq c_{j,2} \cdot \log_2(m(\Lambda)).$$

Here $c_{j,1}$, $c_{j,2}$ are universal constants in the ReLU, RePU and tanh cases, with only $c_{\ell,1}$ depending on $\ell = 2, 3, \ldots$.

Proof. We divide the proof into two cases.

Case 1: Legendre polynomials. The univariate Legendre polynomials $\{\Psi_\nu\}_{\nu\in\mathbb{N}_0}$ are given by (see, e.g., [12, §2.2.2])

$$P_\nu(y) = \frac{1}{2^\nu \nu!} \frac{\mathrm{d}^\nu}{\mathrm{d}y^\nu}(y^2 - 1)^\nu \quad \text{and} \quad \Psi_\nu(y) = \sqrt{2\nu + 1} P_\nu(y), \quad \forall \nu \in \mathbb{N}_0. \tag{5.7.5}$$

Hence, their multivariate counterparts can be written as

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{i\in\operatorname{supp}(\boldsymbol{\nu})} \sqrt{2\nu_i + 1} P_{\nu_i}(y_i), \quad \forall \boldsymbol{y} \in \mathcal{U}, \ \boldsymbol{\nu} \in \mathcal{F}, \tag{5.7.6}$$

where $\operatorname{supp}(\boldsymbol{\nu})$ is as in (2.4.7). Using the fundamental theorem of algebra we may write

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{i\in\operatorname{supp}(\boldsymbol{\nu})} \prod_{j=1}^{\nu_i} \Big(\sqrt{2\nu_i + 1} d_{\nu_i}\Big)^{1/\nu_i} (y_i - r_j^{(\nu_i)}), \quad \forall \boldsymbol{y} \in \mathcal{U}, \quad \boldsymbol{\nu} \in \mathcal{F}. \tag{5.7.7}$$

Here, $\{r_j^{(\nu_i)}\}_{j=1}^{\nu_i}$ are the ν_i roots of the polynomial P_{ν_i} and d_{ν_i} is a scaling factor. Using (5.7.5), we see that the leading coefficient of P_ν is $d_\nu = 2^{-\nu}\frac{(2\nu)!}{(\nu!)^2}$. Then, by Stirling's formula

for factorials $\sqrt{2\pi}n^{n+1/2}\mathrm{e}^{-n} \leq n! \leq \mathrm{e}n^{n+1/2}\mathrm{e}^{-n}$, this coefficient satisfies

$$d_0 = 1 \quad \text{and} \quad d_\nu \leq \frac{\mathrm{e}2^\nu}{\pi\sqrt{2\nu}}, \quad 1 \leq \nu.$$

Next, for $\boldsymbol{\nu} \in \Lambda$, we define the affine map $\mathcal{A}_{\boldsymbol{\nu}} : \mathbb{R}^n \to \mathbb{R}^{\|\boldsymbol{\nu}\|_1}$, $\mathcal{A}_{\boldsymbol{\nu}}(\boldsymbol{y}) = (a_{i,j}(\boldsymbol{y}))_{i\in\mathrm{supp}(\boldsymbol{\nu}),j\in[\nu_i]}$, by

$$a_{i,j}(\boldsymbol{y}) = \left(\sqrt{2\nu_i+1}d_{\nu_i}\right)^{1/\nu_i}(y_i - r_j^{(\nu_i)}), \quad i \in \mathrm{supp}(\boldsymbol{\nu}),\ j \in [\nu_i], \qquad \forall \boldsymbol{y} \in \mathcal{U}. \tag{5.7.8}$$

With this in mind, given $\boldsymbol{\nu} \in \Lambda$, this allows us to define the product of the $\|\boldsymbol{\nu}\|_1$ terms (5.7.7) that comprise $\Psi_{\boldsymbol{\nu}}$. It is useful, however, to make the number of factors in this multiplication constant for all $\boldsymbol{\nu} \in \Lambda$. To this end, we now redefine the affine map $\mathcal{A}_{\boldsymbol{\nu}} : \mathbb{R}^n \to \mathbb{R}^{m(\Lambda)}$ by padding the output vector with $m(\Lambda) - \|\boldsymbol{\nu}\|_1$ terms equal to one.

Our aim now is to apply Lemma 5.7.1 to show that there exists a DNN that approximates the multiplication of the factors in $\mathcal{A}_{\boldsymbol{\nu}}(\boldsymbol{y})$. To do so, we need to identify bounds M_k for each of the terms in the output vector $\mathcal{A}_{\boldsymbol{\nu}}(\boldsymbol{y})$, with $k \in [m(\Lambda)]$. First, notice that the roots of the Legendre polynomials are in $[-1,1]$ for (5.7.7). Then each factor in (5.7.8) is bounded by

$$\widetilde{M}_i := 2(\sqrt{2\nu_i+1})^{1/\nu_i}\left(\frac{\mathrm{e}2^{\nu_i}}{\pi\sqrt{2\nu_i}}\right)^{1/\nu_i}, \quad i \in \mathrm{supp}(\boldsymbol{\nu}),\ j \in [\nu_i].$$

Clearly, the other terms in $\mathcal{A}_{\boldsymbol{\nu}}(\boldsymbol{y})$ are bounded by $M_k = 1$. Therefore, since $|\mathrm{supp}(\boldsymbol{\nu})| \leq |\Theta| = n$ we get

$$\begin{aligned}
\prod_{k=1}^{m(\Lambda)} M_k &= \left(\prod_{i\in\mathrm{supp}(\boldsymbol{\nu})}\prod_{j=1}^{\nu_i}\widetilde{M}_i\right)\cdot\left(\prod_{k=1}^{m(\Lambda)-\|\boldsymbol{\nu}\|_1} 1\right) \\
&= \prod_{i\in\mathrm{supp}(\boldsymbol{\nu})}\prod_{j=1}^{\nu_i} 2(\sqrt{2\nu_i+1})^{1/\nu_i}\left(\frac{\mathrm{e}2^{\nu_i}}{\pi\sqrt{2\nu_i}}\right)^{1/\nu_i} \\
&\leq 2^{2\|\boldsymbol{\nu}\|_1}\left(\frac{\mathrm{e}}{\pi}\right)^n \prod_{i\in\mathrm{supp}(\boldsymbol{\nu})}\sqrt{\frac{2\nu_i+1}{2\nu_i}} \\
&\leq 2^{2m(\Lambda)}\left(\frac{\mathrm{e}}{\pi}\right)^n\left(\frac{3}{2}\right)^{n/2} =: M.
\end{aligned}$$

This defines the multiplication of $m(\Lambda)$ factors. Using Lemma 5.7.1 (or Lemma 5.7.3 in the case of RePU, in which case the previous calculation is unnecessary), for any $0 < \delta < 1$ there exists a ReLU ($j = 1$), a RePU ($j = \ell$) or a tanh ($j = 0$) DNN $\Phi^j_{\delta,M,\boldsymbol{\nu}}$ that approximates the multiplication of the $m(\Lambda)$ factors defining $\mathcal{A}_{\boldsymbol{\nu}}$. Thus, we define the DNN $\Phi^j_{\boldsymbol{\nu}} : \mathbb{R}^n \to \mathbb{R}$ by

$$\Phi^j_{\boldsymbol{\nu}} = \Phi^j_{\delta,M,\boldsymbol{\nu}} \circ \mathcal{A}_{\boldsymbol{\nu}}, \quad j \in \{0,1,\ell\}.$$

By construction, we have

$$\|\Psi_{\boldsymbol{\nu}} - \Phi^j_{\boldsymbol{\nu}} \circ \mathcal{T}_\Theta\|_{L^\infty_{\tilde{\varrho}}(\mathcal{U})} \leq \delta, \quad j \in \{0, 1, \ell\}, \quad \forall \boldsymbol{\nu} \in \Lambda.$$

Applying the bound for M, and some basic algebra we obtain the respective bounds for the width and depth of each $\Phi^j_{\boldsymbol{\nu}}$. Specifically,

$$\begin{aligned}
&\mathrm{width}(\Phi^1_{\boldsymbol{\nu}}) \leq c_{1,1} \cdot m(\Lambda), \\
&\mathrm{depth}(\Phi^1_{\boldsymbol{\nu}}) \leq c_{1,2} \cdot \Big(1 + \log(m(\Lambda)) \cdot \Big[\log(m(\Lambda)\delta^{-1}) + m(\Lambda) + n\Big]\Big),
\end{aligned}$$

in the ReLU case, and

$$\mathrm{width}(\Phi^j_{\boldsymbol{\nu}}) \leq c_{j,1} \cdot m(\Lambda), \qquad \mathrm{depth}(\Phi^j_{\boldsymbol{\nu}}) \leq c_{j,2} \cdot \log_2(m(\Lambda))$$

otherwise. Here $c_{j,1}$, $c_{j,2}$ are universal constants in the ReLU, RePU, tanh cases, with only $c_{\ell,1}$ depending on ℓ. Observe that we have found DNNs $\Phi^j_{\boldsymbol{\nu}}$ of the same depth that approximate each polynomial $\Psi_{\boldsymbol{\nu}}$ for $\boldsymbol{\nu} \in \Lambda$. We consider now the DNN formed by vertically stacking these DNNs, i.e., $\Phi^j_{\Lambda,\delta} = (\Phi^j_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$. It follows immediately that the depth and width of this DNN satisfy

$$\begin{aligned}
&\mathrm{width}(\Phi^1_{\boldsymbol{\nu}}) \leq c_{1,1} \cdot |\Lambda| \cdot m(\Lambda), \\
&\mathrm{depth}(\Phi^1_{\boldsymbol{\nu}}) \leq c_{1,2} \cdot \Big(1 + \log(m(\Lambda)) \cdot \Big[\log(m(\Lambda)\delta^{-1}) + m(\Lambda) + n\Big]\Big),
\end{aligned}$$

in the ReLU case, and

$$\mathrm{width}(\Phi^j_{\boldsymbol{\nu}}) \leq c_{j,1} \cdot |\Lambda| \cdot m(\Lambda), \qquad \mathrm{depth}(\Phi^j_{\boldsymbol{\nu}}) \leq c_{j,2} \cdot \log_2(m(\Lambda)).$$

in the RePU and tanh cases. This completes the proof for the Legendre polynomials.

Case 2: Chebyshev polynomials. The orthonormal Chebyshev polynomials are defined by

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = 2^{\|\boldsymbol{\nu}\|_0/2} \prod_{i\in\mathrm{supp}(\boldsymbol{\nu})} \cos(\nu_i \arccos(y_i)), \quad \forall \boldsymbol{y} \in \mathcal{U},\ \boldsymbol{\nu} \in \mathcal{F}. \tag{5.7.9}$$

We can write each factor as a product over the roots of the polynomials $\cos(\nu_i \arccos(y_i))$, to give

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{i\in\mathrm{supp}(\boldsymbol{\nu})} \prod_{j=1}^{\nu_i} \left(2^{1/2} d_{\nu_i}\right)^{1/\nu_i} (y_i - r_j^{(\nu_i)}), \quad \forall \boldsymbol{y} \in \mathcal{U},\ \boldsymbol{\nu} \in \mathcal{F}. \tag{5.7.10}$$

Define the affine mapping $\mathcal{A}_{\boldsymbol{\nu}} : \mathbb{R}^n \to \mathbb{R}^{m(\Lambda)}$ with entries

$$a_{i,j}(\boldsymbol{y}) = \left(2^{1/2} d_{\nu_i}\right)^{1/\nu_i} (y_i - r_j^{(\nu_i)}), \quad i \in \operatorname{supp}(\boldsymbol{\nu}), j \in [\nu_i],$$

where $d_\nu = 2^{\nu-1}$, and the remaining $m(\Lambda) - \|\boldsymbol{\nu}\|_1$ entries being equal to one. As in the previous case, the roots $r_j^{(\nu_i)} \in [-1, 1]$. Hence we define M as

$$M = \prod_{i=1}^{m(\Lambda)} M_i = \prod_{i \in \operatorname{supp}(\boldsymbol{\nu})} \prod_{j=1}^{\nu_i} \left(2^{1/2} d_{\nu_i}\right)^{1/\nu_i} 2 = 2^{2\|\boldsymbol{\nu}\|_1 - \|\boldsymbol{\nu}\|_0/2} \leq 2^{2m(\Lambda)}.$$

Then, using same arguments as those for the Legendre case, by Lemma 5.7.3 and Lemma 5.7.1 for any $0 < \delta < 1$ there exists a ReLU ($j = 1$), a RePU ($j = \ell$) or a tanh ($j = 0$) DNN $\Phi^j_{\delta,\mathcal{M}}$ that approximates the multiplication of $m(\Lambda)$ factors defined in (5.7.10) with this specific choice of M. Thus, we define the DNNs $\Phi^j_{\boldsymbol{\nu}} : \mathbb{R}^n \to \mathbb{R}$ by

$$\Phi^j_{\boldsymbol{\nu}} = \Phi^j_{\delta,M} \circ \mathcal{A}_{\boldsymbol{\nu}}, \quad j \in \{0, 1, \ell\},$$

for which the following bound holds

$$\left\| \Psi_{\boldsymbol{\nu}} - \Phi^j_{\boldsymbol{\nu}} \circ \mathcal{T}_\Theta \right\|_{L^\infty_\varrho(\mathcal{U})} \leq \delta, \quad j \in \{0, 1, \ell\}, \quad \forall \boldsymbol{\nu} \in \Lambda.$$

We now obtain the result using the bound for M and the same arguments as in the Legendre case. □

5.8 Proofs of main results: Theorems 5.3.1–5.3.4

We are now ready to prove Theorems 5.3.1-5.3.4. The general strategy comes from the proof of the main results in §3.8 (see also [7, §B.4]) in the Hilbert-valued setting. We first show that the polynomial matrix $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ has the wrNSP. Then, we show that its approximation via DNNs $\boldsymbol{A}' \in \mathbb{R}^{m \times N}$ has the wrNSP by using a perturbation result from [7, Lem. 12] (see also [19, Lem. 8.5]). This implies that the operator $\boldsymbol{A}' \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the wrNSP by Lemma 5.6.1 (also Lem.3.7.1). Next, after splitting the error into various terms, we use Lemma 3.6.3 in combination with the results in §2.4.5 to derive the desired error bounds. Finally, we use the results of §5.7 to estimate the depths and widths of the DNN architectures.

Recall that we use the notation $a \lesssim b$ to mean that there exists a constant $c > 0$ independent of a and b such that $a \leq cb$.

5.8.1 Theorem 5.3.1: unknown anisotropy, Banach-valued case

We commence with the unknown anisotropy case.

Proof of Theorem 5.3.1. The proof is divided in several steps.

Step 1: Problem setup. Let Θ, $|\Theta| = n$ be as in (5.3.4), $\Lambda = \Lambda_n^{\mathsf{HCI}}$ be as in (2.4.21), $N = |\Lambda|$ and $\delta > 0$ be a constant whose value will be chosen later in Step 4. Let $\Phi_{\Lambda,\delta}$ be as in Theorem 5.7.4 and consider the class of DNNs (5.5.2). Then, as shown in §5.5.1–5.5.3, we can reformulate the DNN training problem (5.3.5) as the Banach-valued, weighted SR-LASSO problem (5.5.6).

Step 2: Establishing the weighted rNSP. Let $\boldsymbol{A}' \in \mathbb{R}^{m\times N}$ be given by (5.5.3). We now prove that the induced operator $\boldsymbol{A}' \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{u})$ with constants $\gamma' > 0$ and $0 < \rho' < 1$ to be specified. We do this first by establishing the wrNSP for $\boldsymbol{A}$, and then by using a perturbation result [7, Lem. 12] to establish it for $\boldsymbol{A}'$.

First, define the weighted sparsity parameter

$$k := \sqrt{\frac{m}{c_0 L}}, \tag{5.8.1}$$

where $L = L(m,\epsilon)$ is as in (5.3.3) and $c_0 \geq 1$ is a universal constant. Observe that $m \geq m/L \geq m/(c_0 L) = k^2$, since $m \geq 3$ by assumption and therefore $L(m,\epsilon) \geq 1$ for all $0 < \epsilon < 1$. Our aim now is to apply Lemma 3.7.1 to show that $\boldsymbol{A}$ has the wRIP over $\mathbb{R}$ of order $(2k, \boldsymbol{u})$. Let $\bar{c}_0$ be the constant considered therein (related, in turn, to the constants in [51, Thm. 2.14]). Note that we now use the notation $\bar{c}_0$ to avoid confusion with the constants c_0 in Theorem 5.3.1. Set

$$\bar{\delta} = \frac{1}{(4\sqrt{2}\sqrt{k}+1)}.$$

Consider (5.6.1). Since $k \leq \sqrt{m}$ and $m \geq 3$, we have $\log^2(k/\bar{\delta}) \lesssim \log^2(m)$ and since $n = \lceil m/(c_0 L)\rceil \leq 2m$ we have $\log(en) \lesssim \log(m)$. Hence, using the fact that $\log(m) \gtrsim 1$ once more, we deduce that

$$\log^2(k/\bar{\delta}) \cdot \log^2(en) + \log(4/\epsilon) \lesssim \log^4(m) + \log(\epsilon^{-1}) = L(m,\epsilon).$$

We now assume that

$$k \geq 1. \tag{5.8.2}$$

In particular, this implies that $m/(c_0 L) \geq 1$. We discuss the case $k < 1$ at the end of the proof. Using this, we get

$$\bar{c}_0 \cdot \bar{\delta}^{-2} \cdot 2k \cdot \left(\log^2(k/\bar{\delta}) \cdot \log^2(en) + \log(4/\epsilon)\right) \leq c_0 \cdot k^2 \cdot L(m,\epsilon) = m,$$

for a suitably-large choice of the universal constant c_0. It follows that condition (5.6.1), with k and ϵ replaced by $2k$ and $\epsilon/2$, respectively, holds. Therefore, with probability at

least $1-\epsilon/2$, the matrix $\boldsymbol{A}$ has the wRIP over $\mathbb{R}$ of order $(2k, \boldsymbol{u})$ with constant

$$\delta_{2k,\boldsymbol{u}} = \bar{\delta} = \frac{1}{4\sqrt{2}\sqrt{k}+1}. \tag{5.8.3}$$

We now seek to apply Lemma 3.6.7. Notice that, with this value of $\delta_{2k,\boldsymbol{u}}$, we have

$$2\sqrt{2}\frac{\delta_{2k,\boldsymbol{u}}}{1-\delta_{2k,\boldsymbol{u}}} = \frac{1}{2\sqrt{k}}, \qquad \frac{\sqrt{1+\delta_{2k,\boldsymbol{u}}}}{1-\delta_{2k,\boldsymbol{u}}} \leq \frac{3}{2}.$$

Here, in the second step, we used the fact that $k \geq 1$, by assumption. Thus, with probability at least $1-\epsilon/2$, $\boldsymbol{A} \in \mathbb{R}^{m\times N}$ has the weighted rNSP over $\mathbb{R}$ of order $(k, \boldsymbol{u})$ with constants

$$\rho = \frac{1}{2\sqrt{k}}, \quad \gamma = \frac{3}{2}. \tag{5.8.4}$$

Next, we turn our attention to the matrix $\boldsymbol{A}'$. It is a short argument (see Step 3 of the proof of [7, Thm. 5]) based on the definition of $\Phi_{\boldsymbol{\nu},\delta,\Theta}$ to show that

$$\|\boldsymbol{A}-\boldsymbol{A}'\|_2 \leq \sqrt{N}\delta. \tag{5.8.5}$$

Now suppose that δ satisfies

$$\sqrt{N}\delta \leq \tilde{\delta} := \frac{2}{3(3+4k)}. \tag{5.8.6}$$

Later in Step 4 we will ensure that this condition is fulfilled. Then, using a straightforward extension of [19, Lem. 8.5] from the unweighted to the weighted case we deduce that, with probability at least $1-\epsilon/2$, $\boldsymbol{A}'$ has the weighted rNSP over $\mathbb{R}$ of order $(k, \boldsymbol{u})$ with constants

$$\frac{\rho+\gamma\tilde{\delta}\sqrt{k}}{1-\gamma\tilde{\delta}} = \frac{3}{4\sqrt{k}} := \tilde{\rho}, \qquad \frac{\gamma}{1-\gamma\tilde{\delta}} \leq \frac{7}{4} := \widetilde{\gamma}.$$

Here, in the second step we used the fact that $k \geq 1$ once more. Finally, we now apply Lemma 3.6.7. First note that $s^*(k) \leq k$ since $\boldsymbol{u} \geq 1$. Hence, with probability at least $1-\epsilon/2$, the corresponding operator $\boldsymbol{A}' \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the weighted rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{u})$ with constants

$$\sqrt{k}\widetilde{\rho} = \frac{3}{4} := \rho', \qquad \sqrt{k}\widetilde{\gamma} \leq 2\sqrt{k} := \gamma'. \tag{5.8.7}$$

Step 3: Estimating the error. First, we recall that $\mathcal{P}_K : \mathcal{V} \to \mathcal{V}_K$ is a bounded linear operator and $\pi_K = \max\{\|\mathcal{P}_K\|_{\mathcal{V}\to\mathcal{V}_K}, 1\}$. Let $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$ and consider its expansion (2.4.2). As in (2.5.2), let $f_\Lambda = \sum_{\boldsymbol{\nu}\in\Lambda} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}$ be the truncated expansion of f. For convenience, we now recall some notation from §3.7.3

$$E_{\Lambda,2}(f) = \|f-f_\Lambda\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})}, \quad E_{\Lambda,\infty}(f) = \|f-f_\Lambda\|_{L^\infty(\mathcal{U};\mathcal{V})}, \quad E_{\mathsf{disc}} = \|f-\mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})}. \tag{5.8.8}$$

We now derive an error bound for $f - f_{\hat{\Phi},\Theta}$, where $\hat{\Phi}$ is an approximate minimizer of (5.3.5) and $f_{\hat{\Phi},\Theta}$ is as in (5.1.7). Write $\hat{\Phi} = \widehat{\boldsymbol{C}}^\top \Phi_{\Lambda,\delta}$ for $\widehat{\boldsymbol{C}} \in \mathbb{R}^{N\times K}$ and let $\hat{\boldsymbol{c}} = (\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ be the corresponding approximate minimizer of (5.5.6) defined via the relation (5.5.7). Set

$$f_{\hat{\Psi}} = \sum_{\boldsymbol{\nu}\in\Lambda} \hat{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}.$$

Then

$$\begin{aligned}
&\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \\
&\quad \le \|f - \mathcal{P}_K(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f) - \mathcal{P}_K(f_\Lambda)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f_\Lambda) - f_{\hat{\Psi}}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|f_{\hat{\Psi}} - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \\
&\quad =: A_1 + A_2 + A_3 + A_4.
\end{aligned}$$

In addition, for the $L^\infty(\mathcal{U};\mathcal{V})$-norm error, we have

$$\begin{aligned}
&\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \\
&\quad \le \|f - \mathcal{P}_K(f)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f) - \mathcal{P}_K(f_\Lambda)\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|\mathcal{P}_K(f_\Lambda) - f_{\hat{\Psi}}\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|f_{\hat{\Psi}} - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \\
&\quad =: B_1 + B_2 + B_3 + B_4.
\end{aligned}$$

We now bound the terms A_1, A_2, A_3, A_4, B_1, B_2, B_3 and B_4 in several substeps.

Step 3(i): Bounding A_1 and B_1. First, we have $A_1 \le B_1 = E_{\mathsf{disc}}$

Step 3(ii): Bounding A_2 and B_2. Using the linearity of $\mathcal{P}_K$ and the fact that it is a bounded operator, we have

$$A_2 \le \pi_K \|f - f_\Lambda\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \pi_K E_{\Lambda,2}(f). \tag{5.8.9}$$

and

$$B_2 \le \pi_K \|f - f_\Lambda\|_{L^\infty(\mathcal{U};\mathcal{V})} = \pi_K E_{\Lambda,\infty}(f). \tag{5.8.10}$$

Step 3(iii): Bounding A_3 and B_3. Let $\boldsymbol{u}$ be the intrinsic weights in (2.4.17). Recall that $\mathcal{V}$ is a Banach space. Then, we bound A_3 by B_3 in order to obtain a bound in terms of the coefficients in the $\ell^1_{\boldsymbol{u}}(\Lambda;\mathcal{V})$-norm. This step is not necessary in the Hilbert-valued case when Parseval's identity is available, and we can bound A_3 by the coefficients $\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}$ in the $\ell^2(\Lambda;\mathcal{V})$-norm. This issue causes an extra m-dependent factors in the final error bound. Therefore, we have

$$A_3 = \|\mathcal{P}_K(f_\Lambda) - f_{\hat{\Psi}}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \le B_3 = \|\mathcal{P}_K(f_\Lambda) - f_{\hat{\Psi}}\|_{L^\infty(\mathcal{U};\mathcal{V})} \le \|\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}}$$

where $\boldsymbol{c}_{\Lambda,K} = (\mathcal{P}_K(c_{\boldsymbol{\nu}}))_{\boldsymbol{\nu}\in\Lambda}$.

We now apply Lemma 3.6.3 to the problem (5.5.6). In Step 2 we showed that, with probability at least $1-\epsilon/2$, $\boldsymbol{A}' \in \mathcal{B}(\mathcal{V}_K^N, \mathcal{V}_K^m)$ has the weighted rNSP of order $(k, \boldsymbol{u})$ with constants ρ' and γ' given by (5.8.7) and k given by (5.8.1). Hence, this lemma gives

$$\|\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \leq C_1 \left(2\sigma_k(\boldsymbol{c}_{\Lambda,K})_{1,\boldsymbol{u};\mathcal{V}} + \frac{\mathcal{G}'(\hat{\boldsymbol{c}}) - \mathcal{G}'(\boldsymbol{c}_{\Lambda,K})}{\lambda}\right) + \left(\frac{C_1}{\lambda} + C_2\sqrt{k}\right)\|\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K} - \boldsymbol{f}\|_{2;\mathcal{V}},$$

with probability at least $1-\epsilon/2$, where $C_1 = (1+\rho')/(1-\rho')$, $C_2 = 2\gamma'/(1-\rho')$ and $\mathcal{G}'$ is as in (5.5.6). Notice that this holds provided $\lambda \leq C_1'/(C_2'\sqrt{k})$, where $C_1' = (1+\rho')^2/(1-\rho')$, $C_2' = (3+\rho')\gamma'/(1-\rho')$. Using the values for γ' and ρ' we get

$$\frac{1}{6\sqrt{k}} = \left(\frac{1}{\gamma'}\right)\frac{1}{3} < \frac{1}{\gamma'}\left[1 - \frac{2}{(3+\rho')}\right] = \frac{1}{\gamma'}\left[\frac{(1+\rho')}{(3+\rho')}\right] \leq \frac{1}{\gamma'}\left[\frac{(1+\rho')^2}{(3+\rho')}\right],$$

which implies that

$$\lambda := \frac{1}{6\sqrt{m/L}} = \frac{1}{6\sqrt{c_0 k}} \leq \frac{1}{\sqrt{k}}\cdot\frac{1}{6\sqrt{k}} \leq \frac{1}{\sqrt{k}}\cdot\frac{1}{\gamma'}\left[\frac{(1+\rho')^2}{(3+\rho')}\right] = \frac{C_1'}{C_2'\sqrt{k}}, \tag{5.8.11}$$

as required. Here we used the fact that $c_0 \geq 1$. Now, since $\hat{\boldsymbol{c}}$ is an approximate minimizer and $\boldsymbol{c}_{\Lambda,K} \in \mathcal{V}_K^N$ is feasible for (5.5.6), we have $\mathcal{G}'(\hat{\boldsymbol{c}}) - \mathcal{G}'(\boldsymbol{c}_{\Lambda,K}) \leq E_{\mathsf{opt}}$, where E_{opt} is as in (5.3.1). Also, due to the values of ρ' and γ' given by (5.8.7), we notice that $C_1, C_1' \lesssim 1$ and $C_2, C_2' \lesssim \sqrt{k}$. Substituting this into the previous bound and noticing that $1/\lambda \lesssim k$, we obtain

$$\|\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \lesssim \sigma_k(\boldsymbol{c}_{\Lambda,K})_{1,\boldsymbol{u};\mathcal{V}} + kE_{\mathsf{opt}} + k\|\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K} - \boldsymbol{f}\|_{2;\mathcal{V}},$$

with probability at least $1-\epsilon/2$. Consider the first term. Since $\mathcal{P}_K$ satisfies (2.2.14), (2.4.10) implies that

$$\sigma_k(\boldsymbol{c}_{\Lambda,K})_{1,\boldsymbol{u};\mathcal{V}} = \inf\left\{\sum_{\boldsymbol{\nu}\in\Lambda\setminus S} u_{\boldsymbol{\nu}}\|\mathcal{P}_K(c_{\boldsymbol{\nu}})\|_{\mathcal{V}} : S \subseteq \Lambda,\ |S|_{\boldsymbol{u}} \leq k\right\} \leq \pi_K\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}.$$

Hence

$$\|\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \lesssim \pi_K\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}} + k\|\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K} - \boldsymbol{f}\|_{2;\mathcal{V}} + kE_{\mathsf{opt}}, \tag{5.8.12}$$

with probability at least $1-\epsilon/2$. We now estimate the second term. Let $i = 1, \ldots, m$ and write

$$\begin{aligned}\sqrt{m}\left(\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K} - \boldsymbol{f}\right)_i &= \sum_{\boldsymbol{\nu}\in\Lambda} \mathcal{P}_K(c_{\boldsymbol{\nu}})\Phi_{\boldsymbol{\nu},\delta,n}(\boldsymbol{y}_i) - f(\boldsymbol{y}_i) - n_i \\ &= \sum_{\boldsymbol{\nu}\in\Lambda} \mathcal{P}_K(c_{\boldsymbol{\nu}})\left(\Phi_{\boldsymbol{\nu},\delta,n}(\boldsymbol{y}_i) - \Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i)\right) + \sum_{\boldsymbol{\nu}\in\Lambda} \mathcal{P}_K(c_{\boldsymbol{\nu}})\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i) - f(\boldsymbol{y}_i) - n_i.\end{aligned}$$

Then, using Theorem 5.7.4, the triangle inequality and the fact that $\mathcal{P}_K$ is a bounded linear operator, we get

$$\begin{aligned}\|\sqrt{m}(\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K}-\boldsymbol{f})_i\|_{\mathcal{V}} &\leq \sum_{\boldsymbol{\nu}\in\Lambda}\|\mathcal{P}_K(c_{\boldsymbol{\nu}})\|_{\mathcal{V}}\delta+\left\|\sum_{\boldsymbol{\nu}\in\Lambda}\mathcal{P}_K(c_{\boldsymbol{\nu}})\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i)-f(\boldsymbol{y}_i)\right\|_{\mathcal{V}}+\|n_i\|_{\mathcal{V}}\\ &\leq \delta\sum_{\boldsymbol{\nu}\in\Lambda}\|\mathcal{P}_K(c_{\boldsymbol{\nu}})\|_{\mathcal{V}}+\left\|\sum_{\boldsymbol{\nu}\notin\Lambda}\mathcal{P}_K(c_{\boldsymbol{\nu}})\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i)\right\|_{\mathcal{V}}+\|f(\boldsymbol{y}_i)-\mathcal{P}_K(f)(\boldsymbol{y}_i)\|_{\mathcal{V}}+\|n_i\|_{\mathcal{V}}\\ &\leq \pi_K\sqrt{N}\delta\|\boldsymbol{c}_\Lambda\|_{2;\mathcal{V}}+\pi_K\left\|\sum_{\boldsymbol{\nu}\notin\Lambda}c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i)\right\|_{\mathcal{V}}+\|f-\mathcal{P}_K(f)\|_{L^\infty_\varrho(\mathcal{U};\mathcal{V})}+\|n_i\|_{\mathcal{V}}\\ &= \pi_K\left(\sqrt{N}\delta\|\boldsymbol{c}_\Lambda\|_{2;\mathcal{V}}+\|f(\boldsymbol{y}_i)-f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}}\right)+E_{\mathsf{disc}}+\|n_i\|_{\mathcal{V}},\end{aligned}$$

where E_{disc} is as in (5.3.8). Notice that, by (2.4.2), the Cauchy-Schwarz inequality and the orthonormality of $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$, we have

$$\|\boldsymbol{c}_{\boldsymbol{\nu}}\|_{\mathcal{V}}=\left\|\int_{\mathcal{U}}f(\boldsymbol{y})\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})\,\mathrm{d}\varrho(\boldsymbol{y})\right\|_{\mathcal{V}}\leq\int_{\mathcal{U}}\|f(\boldsymbol{y})\|_{\mathcal{V}}|\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})|\,\mathrm{d}\varrho(\boldsymbol{y})\leq\|f\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}\cdot 1\leq 1,\quad\forall\boldsymbol{\nu}\in\Lambda.$$

In the last inequality we used the fact that $\|f\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}\leq\|f\|_{L^\infty(\mathcal{U};\mathcal{V})}\leq 1$. Hence

$$\|\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K}-\boldsymbol{f}\|_{2;\mathcal{V}}\leq\pi_K N\delta+\pi_K\sqrt{\frac{1}{m}\sum_{i=1}^m\|f(\boldsymbol{y}_i)-f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}}^2}+E_{\mathsf{disc}}+E_{\mathsf{samp}},$$

where E_{samp} is as in (5.3.8). Now, since $m=c_0\cdot L\cdot k^2\geq 2\cdot k^2\cdot\log(4/\epsilon)$ for a sufficiently large choice of the universal constant c_0, the arguments in [12, Lem. 7.11] imply that

$$\sqrt{\frac{1}{m}\sum_{i=1}^m\|f(\boldsymbol{y}_i)-f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}}^2}\leq\sqrt{2}\left(\frac{E_{\Lambda,\infty}(f)}{k}+E_{\Lambda,2}(f)\right),\tag{5.8.13}$$

with probability at least $1-\epsilon/2$, where $E_{\Lambda,2}(f)$ and $E_{\Lambda,\infty}(f)$ are as in (5.8.8). We deduce that

$$\|\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K}-\boldsymbol{f}\|_{2;\mathcal{V}}\lesssim\pi_K\left(N\delta+\frac{E_{\Lambda,\infty}(f)}{k}+E_{\Lambda,2}(f)\right)+E_{\mathsf{disc}}+E_{\mathsf{samp}},$$

with probability at least $1-\epsilon/2$. Substituting this into (5.8.12) and applying the union bound now yields

$$A_3\leq B_3\lesssim\pi_K\sqrt{k}\left(\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}}+\sqrt{k}N\delta+\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}}+\sqrt{k}E_{\Lambda,2}(f)\right)+k\left(E_{\mathsf{opt}}+E_{\mathsf{samp}}+E_{\mathsf{disc}}\right),\tag{5.8.14}$$

with probability at least $1-\epsilon$.

Step 3(iv): Bounding A_4 and B_4. Recalling that $f_{\hat{\Phi},\Theta} = \sum_{\boldsymbol{\nu}\in\Lambda} \hat{c}_{\boldsymbol{\nu}} \Phi_{\boldsymbol{\nu},\delta,\Theta}$ and that $\boldsymbol{u} \geq \mathbf{1}$, we first write

$$\|f_{\hat{\Psi}} - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f_{\hat{\Psi}} - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq \sum_{\boldsymbol{\nu}\in\Lambda} \|\Psi_{\boldsymbol{\nu}} - \Phi_{\boldsymbol{\nu},\delta,\Theta}\|_{L^\infty(\mathcal{U})} u_{\boldsymbol{\nu}} \|\hat{c}_{\boldsymbol{\nu}}\|_{\mathcal{V}} \leq \delta \|\hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}}.$$

Recall that $\hat{\boldsymbol{c}}$ is an approximate minimizer of (5.5.6). Hence

$$\lambda \|\hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \leq \lambda \|\mathbf{0}\|_{1,\boldsymbol{u};\mathcal{V}} + \|\boldsymbol{A}'\mathbf{0} - \boldsymbol{f}\|_{2;\mathcal{V}} + E_{\mathsf{opt}} = \|\boldsymbol{f}\|_{2;\mathcal{V}} + E_{\mathsf{opt}},$$

where $\mathbf{0} \in \mathcal{V}_K^N$ is the zero vector. Using the definitions of $\boldsymbol{f}$ and λ in (2.5.1) and (5.8.11), respectively, we see that

$$\|\hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \lesssim k \left(\|\boldsymbol{e}\|_{2;\mathcal{V}} + \|f\|_{L^\infty_\varrho(\mathcal{U};\mathcal{V})} + E_{\mathsf{opt}} \right) \leq k \left(E_{\mathsf{samp}} + 1 + E_{\mathsf{opt}} \right).$$

Note that $\delta k \lesssim 1$ due to (5.8.6). Since $k \geq 1$, we get

$$A_4 \leq B_4 \lesssim k\delta + \sqrt{k} \left(E_{\mathsf{samp}} + E_{\mathsf{opt}} \right). \tag{5.8.15}$$

Step 3(v): Final bound. Combining the estimates in Step 3(i), (5.8.9),(5.8.10), (5.8.14) and (5.8.15) and using the facts that $\pi_K \geq 1$ and $k \geq 1$ we deduce that

$$\begin{aligned} \|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim \pi_K \sqrt{k} \left(\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + \sqrt{k}\delta N + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + \sqrt{k} E_{\Lambda,2}(f) \right) \\ + k \left(E_{\mathsf{samp}} + E_{\mathsf{opt}} + E_{\mathsf{disc}} \right), \end{aligned} \tag{5.8.16}$$

which results in the same bound for the L^∞-norm error

$$\begin{aligned} \|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \pi_K \sqrt{k} \left(\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + \sqrt{k}\delta N + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + \sqrt{k} E_{\Lambda,2}(f) \right) \\ + k \left(E_{\mathsf{samp}} + E_{\mathsf{opt}} + E_{\mathsf{disc}} \right). \end{aligned} \tag{5.8.17}$$

The fact that the $L^2_\varrho(\mathcal{U};\mathcal{V})$ and the $L^\infty(\mathcal{U};\mathcal{V})$-norm error are bounded by the same terms in this case it is not surprising as we bound most of the important terms by their corresponding $L^\infty(\mathcal{U};\mathcal{V})$ norm. This concludes Step 3.

Step 4: Establishing the algebraic rates. Here, we bound the first four terms in (5.8.17) (and consequently in (5.8.16)). Recall the definition of k in (5.8.1). Then part (ii) of Theorem 2.4.13 with $q = 1 > p$ gives

$$\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} \leq C(\boldsymbol{b},\varepsilon,p) \cdot k^{1/2-1/p} = C(\boldsymbol{b},\varepsilon,p) \cdot \left(\frac{m}{c_0 L} \right)^{\frac{1}{2}(1/2-1/p)},$$

where $C(\boldsymbol{b}, \varepsilon, p) > 0$ depends on $\boldsymbol{b}$, ϵ and p only. Next, define the following term:

$$E_\Lambda(f) = \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + \sqrt{k} E_{\Lambda,2}(f).$$

As noted, the set $\Lambda = \Lambda_n^{\mathsf{HCI}}$ contains the union of all anchored sets of size at most n (see [12, Prop. 2.18]). We now use Corollary 2.4.15 with $s = n$ and $q = 1$. This implies that there exists an anchored set $S \subset \mathcal{F}$ of size $|S| \leq n$ such that

$$E_{\Lambda,\infty}(f) = \|f - f_\Lambda\|_{L_\varrho^\infty(\mathcal{U};\mathcal{V})} \leq \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot n^{1-1/p}. \quad (5.8.18)$$

Similarly, using Corollary 2.4.15 with $q = 1$ we get

$$E_{\Lambda,2}(f) = \|f - f_\Lambda\|_{L_\varrho^2(\mathcal{U};\mathcal{V})} \leq \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot n^{1-1/p}. \quad (5.8.19)$$

Therefore

$$E_\Lambda(f) \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \left(k^{-1/2} \cdot n^{1-1/p} + \sqrt{k} \cdot n^{1-1/p}\right) \leq C(\boldsymbol{b}, \varepsilon, p) \cdot k^{1/2} \cdot n^{1-1/p}. \quad (5.8.20)$$

Since $p \leq 1/2$, the exponent $1 - 1/p$ is negative. Using the definitions of n and k in (5.3.4) and (5.8.1), respectively, we see that $n \geq k^2$. Hence

$$E_\Lambda(f) \lesssim C(\boldsymbol{b}, \varepsilon, p) \cdot k^{5/2-2/p} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot k^{1/2-1/p} = C(\boldsymbol{b}, \varepsilon, p) \cdot \left(\frac{m}{c_0 L}\right)^{\frac{1}{2}(1/2-1/p)}.$$

Here, in the penultimate step we use the fact that $k \geq 1$ by assumption and $p \leq 1/2$. Returning to (5.8.17) , we deduce that

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq \pi_K \cdot C(\boldsymbol{b}, \varepsilon, p) \cdot \left(\frac{m}{c_0 L}\right)^{\frac{1}{2}(1-1/p)} + k\left(E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}\right), \quad (5.8.21)$$

provided δ satisfies $k\delta N \leq k^{1-1/p}$. Hence it suffices to to choose

$$\delta \leq N^{-1} k^{-\frac{1}{p}}.$$

Therefore, recalling (5.8.6), we now set

$$\delta = \min\left\{\frac{2}{3(3+4k)\sqrt{N}}, \frac{k^{-\frac{1}{p}}}{N}\right\}. \quad (5.8.22)$$

In this way, using the definition of $E_{\mathsf{app,UB}}$ in (5.3.8), (5.8.21) and the bound $k \lesssim \sqrt{m}$ we get

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim E_{\mathsf{app,UB}} + m^{1/2} \cdot (E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}),$$
$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim E^\infty_{\mathsf{app,UB}} + m^{1/2} \cdot (E_{\mathsf{disc}} + E_{\mathsf{samp}} + E_{\mathsf{opt}}),$$

as required.

Step 5: Bounding the width and depth of the DNN architecture. We have now established the main error bounds (5.3.6) and 5.3.7. In this penultimate step, we derive the bounds for the width and depth of the class of DNNs $\mathcal{N}$. To do this, we follow similar arguments to those in Step 6 of the proof of [7, Thm. 5]. Using (5.8.22) and the facts that $k \geq 1$ and $p < 1$, we first see that

$$\delta \gtrsim N^{-1} k^{-\frac{1}{p}} \quad \Rightarrow \quad \log(\delta^{-1}) \lesssim \log(N) + \frac{1}{p}\log(k). \tag{5.8.23}$$

From the definition of Λ_n^{HCI} in (2.4.21) notice that $m(\Lambda) = \max_{\boldsymbol{\nu}\in\Lambda} \|\boldsymbol{\nu}\|_1 \leq n$. We now apply Theorem 5.7.4 with the ReLU activation function and the choice $\Theta = [n]$ as in (5.3.4). Notice that this choice is valid, since every $\boldsymbol{\nu} \in \Lambda = \Lambda_n^{\mathsf{HCI}}$ satisfies $\mathrm{supp}(\boldsymbol{\nu}) \subseteq [n]$. We deduce that the width and depth of the DNN $\mathcal{N} = \mathcal{N}^1$ satisfies

$$\begin{aligned}
&\mathrm{width}(\mathcal{N}^1) \lesssim Nn, \\
&\mathrm{depth}(\mathcal{N}^1) \lesssim \Big(1 + \log(n)\Big[\log(n) + \log(N) + p^{-1}\log(k) + n\Big]\Big),
\end{aligned}$$

Noticing that, $3 \leq m$, $k^2 \leq n \leq m$ and $N \lesssim n^{2+\log_2(n)}$, $\log(N) \lesssim \log^2(n)$ (these follow from (2.4.23)) now gives the result in the ReLU case. On the other hand, for the RePU or hyperbolic tangent activation function, the width and depth of this DNN satisfy

$$\mathrm{width}(\mathcal{N}^j) \leq c_{j,1} \cdot n^{3+\log_2(n)}, \quad \mathrm{depth}(\mathcal{N}^j) \leq c_{j,2} \cdot \log_2(n),$$

where $c_{j,1}$, $c_{j,2}$ are universal constants for the hyperbolic tangent activation function ($j = 0$) and $c_{j,1}, c_{j,2}$ depend on ℓ for the RePU activation function ($j = \ell$). The bounds in these cases now follow from the fact that $n \leq m$.

Step 6: The case $k < 1$. So far, we have assumed that $k \geq 1$. We now address the case $k < 1$. In this case, since $p < 1$ we have

$$k = \sqrt{\frac{m}{c_0 L}} < 1 \Rightarrow 1 < \left(\frac{m}{c_0 L}\right)^{\frac{1}{2}(1-1/p)}. \tag{5.8.24}$$

Next, we skip Step 2 of the above argument, and go directly to Step 3. Note that

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq B_1 + B_2 + B_3 + B_4.$$

We now bound the various terms.

Step 6(i): Bounding B_1. Once more we have $B_1 \leq E_{\mathsf{disc}}$.

Step 6(ii): Bounding B_2. Using (5.8.10) and (5.8.18) with $n = 1$ we get

$$B_2 \leq \pi_K E_{\Lambda,\infty}(f) \leq \pi_K C(\boldsymbol{b}, \varepsilon, p).$$

Step 6(iii): Bounding B_3. Once more, using (5.8.18) with $n = 1$ and triangle inequality we obtain

$$\begin{aligned}\|\mathcal{P}_K(f_\Lambda) - f_{\hat{\Psi}}\|_{L^\infty(\mathcal{U};\mathcal{V})} &\leq \|\mathcal{P}_K(f_\Lambda) - \mathcal{P}_K(f) + \mathcal{P}_K(f) - f_{\hat{\Psi}}\|_{L^\infty(\mathcal{U};\mathcal{V})} \\ &\leq \pi_K \cdot C(\boldsymbol{b}, \varepsilon, p) + \pi_K \|f\|_{L^\infty(\mathcal{U};\mathcal{V})} + \|\hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}}.\end{aligned}$$

Since $\hat{\boldsymbol{c}}$ is an approximate minimizer, following the same analysis as Step 3(iv) gives

$$\|\hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \lesssim k(E_{\mathsf{samp}} + 1 + E_{\mathsf{opt}}).$$

Therefore $B_3 \lesssim k(E_{\mathsf{samp}} + 1 + E_{\mathsf{opt}}) + \pi_K$.

Step 6(iv): Bounding B_4. This step is almost identical and gives $B_4 \lesssim k\delta(E_{\mathsf{samp}} + 1 + E_{\mathsf{opt}})$.

Step 6(v): Final bound. Combining the previous estimates and using the fact that $\pi_K \geq 1$, $\delta k < k < 1$ (the first inequality follows from (5.8.22)), we deduce that

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \pi_K(C(\boldsymbol{b}, \varepsilon, p) + 1) + k(E_{\mathsf{samp}} + E_{\mathsf{opt}}) + E_{\mathsf{disc}}.$$

The condition (5.8.24) and the fact that $m \geq 3$ give that $m^{1/2} > 1 > k$. Using (5.8.24) once more, we deduce that

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \pi_K \cdot \left(\frac{m}{c_0 L}\right)^{\frac{1}{2}(1-1/p)} + m^{1/2}(E_{\mathsf{samp}} + E_{\mathsf{opt}} + E_{\mathsf{disc}}).$$

The error bounds (5.3.6)–(5.3.7) in the case $k < 1$ now follows with $C = C(\boldsymbol{b}, \varepsilon, p) + 1$, depending on $\boldsymbol{b}$, ε and p only.

Step 6(v): Bounding the width and depth of the DNN architecture. It remains to bound the width and depth of the DNN architecture in the case $k < 1$. Since $m/(c_0 L) < 1$, by (5.3.4)

we see that $n = 1$ in this case. Recall also that δ is defined by (5.8.22). Hence $\delta \gtrsim 1/N$ in this case, since $k < 1$ and $N \geq 1$. We deduce that $\log(\delta^{-1}) \lesssim \log(N)$. We now argue as in Step 5, using this bound and the fact that $n = 1 < m$. Thus, the bounds still hold in this case.

□

5.8.2 Theorem 5.3.2: unknown anisotropy, Hilbert-valued case

Proof of Theoorem 5.3.2. The proof is similar to that of Theorem 5.3.1, except that it uses Parseval's identity and Lemma 3.6.6 instead of Lemma 3.6.7.

Step 1: Problem setup. This step is identical.

Step 2: Establishing the weighted rNSP. In this case, we define

$$k := \frac{m}{c_0 L}, \tag{5.8.25}$$

where $L = L(m, \epsilon)$ is as in (5.3.3) and $c_0 \geq 1$ is a universal constant. Observe that $m \geq m/L \geq m/(c_0 L) = k$. Let $\bar{c}_0$ be the constant considered in Lemma 3.7.1. Set $\bar{\delta} = \left(4\sqrt{2}+1\right)^{-1}$. Observe that in this case we do not need the assumption $k > 1$ as in the previous case. Indeed, since $k \leq m$, $m \geq 3$ and $n \leq 2m$ we deduce that

$$\bar{c}_0 \cdot \bar{\delta}^{-2} \cdot 2k \cdot \left(\log^2(k/\bar{\delta}) \cdot \log^2(en) + \log(4/\epsilon)\right) \leq c_0 \cdot k \cdot L(m, \epsilon) = m,$$

for a suitably-large choice of the universal constant c_0. By similar arguments to those of the previous theorem, considering Lemma 3.7.1 and Remark 3.7.2, we deduce that, with probability at least $1-\epsilon/2$, the matrix $\boldsymbol{A}$ has the wRIP over $\mathbb{R}$ of order $(2k, \boldsymbol{u})$ with constant

$$\delta_{2k,\boldsymbol{u}} = \bar{\delta} = \frac{1}{4\sqrt{2}+1}. \tag{5.8.26}$$

Hence, by Lemma 3.6.5 it has the weighted rNSP over $\mathbb{R}$ of order $(k, \boldsymbol{u})$ with constants

$$\rho = \frac{1}{2}, \quad \gamma = \frac{3}{2}. \tag{5.8.27}$$

Note that (5.8.5) holds in this case, since this property pertains to the matrices $\boldsymbol{A}$ and $\boldsymbol{A}'$ and not the associated linear operators. We now assume that δ satisfies

$$\sqrt{N}\delta \leq \tilde{\delta} := \frac{2}{3(3+4\sqrt{k})}. \tag{5.8.28}$$

By the same extension of [19, Lem. 8.5] and from the values of ρ and γ in (5.8.27) the matrix $\boldsymbol{A}'$ has the weighted rNSP of order $(k, \boldsymbol{u})$ over $\mathbb{R}$ with constants

$$\frac{\rho + \gamma\tilde{\delta}\sqrt{k}}{1-\gamma\tilde{\delta}} = \frac{3}{4} := \tilde{\rho}, \qquad \frac{\gamma}{1-\gamma\tilde{\delta}} \leq \frac{9}{4} := \widetilde{\gamma}, \tag{5.8.29}$$

with probability at least $1-\epsilon/2$. Then, applying Lemma 3.6.6 the corresponding operator $\boldsymbol{A}' \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ satiesfies the weighted rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{u})$ with constants $\rho' < 1$ and $\gamma' > 0$, where $\rho' = \widetilde{\rho}$ and $\gamma' = \widetilde{\gamma}$, with probability $1-\epsilon/2$.

Step 3: Estimating the error. The setup, Step 3(i) and Step 3(ii) are identical. For Step 3(iii), we can use Parseval's identity to bound A_3 in terms of the $\ell^2(\Lambda; \mathcal{V})$-norm of the coefficients $\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}$. That is,

$$A_3 = \|\mathcal{P}_K(f_\Lambda) - f_{\hat{\Psi}}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \|\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}\|_{2;\mathcal{V}},$$

where $\boldsymbol{c}_{\Lambda,K} = (\mathcal{P}_K(c_{\boldsymbol{\nu}}))_{\boldsymbol{\nu}\in\Lambda}$. We now apply Lemma 3.6.3 to the problem (5.5.6). This lemma gives

$$\|\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}\|_{2;\mathcal{V}} \leq \frac{C_1'}{\sqrt{k}}\left(2\sigma_k(\boldsymbol{c}_{\Lambda,K})_{1,\boldsymbol{u};\mathcal{V}} + \frac{\mathcal{G}'(\hat{\boldsymbol{c}}) - \mathcal{G}'(\boldsymbol{c}_{\Lambda,K})}{\lambda}\right) + \left(\frac{C_1'}{\sqrt{k}\lambda} + C_2'\right)\|\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K} - \boldsymbol{f}\|_{2;\mathcal{V}},$$

with probability at least $1-\epsilon/2$, where $C_1' = (1+\rho')^2/(1-\rho')$, $C_2' = (3+\rho')\gamma'/(1-\rho')$ and $\mathcal{G}'$ is as in (5.5.6). The values of ρ' and γ' in (5.8.29) give

$$\lambda := \frac{1}{6\sqrt{m/L}} = \frac{1}{6\sqrt{c_0 k}} \leq \frac{1}{6\sqrt{k}} < \frac{(1+\rho')^2}{(3+\rho')\gamma'}\frac{1}{\sqrt{k}}.$$

Now, since $m = c_0 \cdot L \cdot k \geq 2 \cdot k \cdot \log(4/\epsilon)$ for a sufficiently large choice of the universal constant c_0, once more the arguments in [12, Lem. 7.11] imply that

$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}\|f(\boldsymbol{y}_i) - f_\Lambda(\boldsymbol{y}_i)\|_{\mathcal{V}}^2} \leq \sqrt{2}\left(\frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f)\right), \tag{5.8.30}$$

with probability at least $1-\epsilon/2$. Notice that (5.8.13) is slightly different to (5.8.30). Following similar arguments to Step 3(iii), we deduce that A_3 satisfies

$$A_3 \lesssim \pi_K\left(\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + N\delta + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f)\right) + E_{\mathsf{opt}} + E_{\mathsf{samp}} + E_{\mathsf{disc}}. \tag{5.8.31}$$

For B_3, notice once again that applying Lemma 3.6.3 to the problem (5.5.6) gives

$$B_3 \leq \|\boldsymbol{c}_{\Lambda,K} - \hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} \leq C_1\left(2\sigma_k(\boldsymbol{c}_{\Lambda,K})_{1,\boldsymbol{u};\mathcal{V}} + \frac{\mathcal{G}'(\hat{\boldsymbol{c}}) - \mathcal{G}'(\boldsymbol{c}_{\Lambda,K})}{\lambda}\right) + \left(\frac{C_1}{\lambda} + C_2\sqrt{k}\right)\|\boldsymbol{A}'\boldsymbol{c}_{\Lambda,K} - \boldsymbol{f}\|_{2;\mathcal{V}},$$

with probability at least $1-\epsilon/2$, where $C_1 = (1+\rho')/(1-\rho')$, $C_2 = 2\gamma'/(1-\rho')$, and where ρ' and γ' are as in (5.8.29), and $\mathcal{G}'$ is as in (5.5.6). Therefore, we deduce that

$$B_3 \lesssim \pi_K \sqrt{k} \left(\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + N\delta + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f) \right) + \sqrt{k}(E_{\mathsf{opt}} + E_{\mathsf{samp}} + E_{\mathsf{disc}}) \quad (5.8.32)$$

with probability at least $1-\epsilon$. Step 3(iv) is essentially the same except that we use the bounds $1/\lambda \lesssim \sqrt{k}$, $\sqrt{k}\delta \lesssim 1$ and $\delta\sqrt{k} \leq \delta\sqrt{N} \leq \delta N$ in this case to get the bound

$$A_4 \leq B_4 \lesssim \sqrt{k}\delta(E_{\mathsf{samp}} + 1 + E_{\mathsf{opt}}) \lesssim N\delta + E_{\mathsf{samp}} + E_{\mathsf{opt}}.$$

Hence, combining the estimates and using the fact that $k \leq n \leq N$ once more, we deduce that

$$\begin{aligned} \|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim \pi_K \left(\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + N\delta + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f) \right) \\ + E_{\mathsf{samp}} + E_{\mathsf{opt}} + E_{\mathsf{disc}}. \end{aligned} \quad (5.8.33)$$

and

$$\begin{aligned} \|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \pi_K \sqrt{k} \left(\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} + N\delta + \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f) \right) \\ + \sqrt{k}(E_{\mathsf{samp}} + E_{\mathsf{opt}} + E_{\mathsf{disc}}). \end{aligned} \quad (5.8.34)$$

This concludes Step 3.

Step 4: Establishing the algebraic rates. By the same arguments, except using (5.8.25), we get

$$\frac{\sigma_k(\boldsymbol{c}_\Lambda)_{1,\boldsymbol{u};\mathcal{V}}}{\sqrt{k}} \leq C(\boldsymbol{b},\varepsilon,p) \cdot k^{1/2-1/p} = C(\boldsymbol{b},\varepsilon,p) \cdot \left(\frac{m}{c_0 L} \right)^{(1/2-1/p)}.$$

Similarly as in Step 4, there exists an anchored set $S \subset \mathcal{F}$ of size $|S| \leq n$ such that

$$E_{\Lambda,\infty}(f) = \|f - f_\Lambda\|_{L^\infty_\varrho(\mathcal{U};\mathcal{V})} \leq \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b},\varepsilon,p) \cdot n^{1-1/p}.$$

Now, using Corollary 2.4.15 with $q = 2$ we get

$$E_{\Lambda,2}(f) = \|f - f_\Lambda\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{2;\mathcal{V}} \leq C(\boldsymbol{b},\varepsilon,p) \cdot n^{1/2-1/p}.$$

Hence

$$E_\Lambda(f) = \frac{E_{\Lambda,\infty}(f)}{\sqrt{k}} + E_{\Lambda,2}(f) \lesssim C(\boldsymbol{b},\varepsilon,p) \cdot \left(k^{-1/2} n^{1-1/p} + n^{1/2-1/p} \right) \lesssim C(\boldsymbol{b},\varepsilon,p) \cdot \left(\frac{m}{c_0 L} \right)^{1/2-1/p}.$$

Here, in the final step, we used the definitions of k and n and the fact that $p < 1$. Therefore, using (5.8.33) and (5.8.34) we conclude that

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \lesssim C(\boldsymbol{b},\varepsilon,p) \cdot \pi_K \cdot \left(\frac{m}{c_0 L}\right)^{(1/2-1/p)} + E_{\mathsf{samp}} + E_{\mathsf{opt}} + E_{\mathsf{disc}},$$
$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim C(\boldsymbol{b},\varepsilon,p) \cdot \pi_K \cdot \left(\frac{m}{c_0 L}\right)^{(1-1/p)} + m^{1/2}(E_{\mathsf{samp}} + E_{\mathsf{opt}} + E_{\mathsf{disc}}),$$

for potentially different values of $C(\boldsymbol{b},\varepsilon,p)$ provided $\delta \leq N^{-1}k^{1/2-1/p}$. Hence, in view of (5.8.28), we now set $\delta = \min\left\{\frac{2}{3(3+4\sqrt{k})\sqrt{N}}, \frac{k^{1/2-1/p}}{N}\right\}$.

Step 5: Bounding the width and depth of the DNN architecture. This step is essentially the same. Since bound (5.8.23) remains valid, the only possible difference is in the various universal constants.

Note that in this proof we do not need the assumption $k < 1$. Hence, Step 6 is not necessary. □

5.8.3 Theorem 5.3.3: known anisotropy, Banach-valued case

Proof of Theorem 5.3.3. We proceed in similar steps to those of the previous two theorems.

Step 1: Problem setup. Let $S \subset \mathcal{F}$ be a finite index set and write $s = |S|$. We will choose a suitable S in Step 4 below. Now let $\Theta \subset \mathbb{N}$ be any set for which

$$\bigcup_{\boldsymbol{\nu}\in S} \operatorname{supp}(\boldsymbol{\nu}) \subseteq \Theta. \tag{5.8.35}$$

Notice that the left-hand side is a finite set, since S is finite and any multi-index $\boldsymbol{\nu} \in \mathcal{F}$ has only finitely many nonzero terms. Hence Θ can be chosen as a finite set. We make a precise choice of Θ in Step 4 once we have defined S. Next, let $\Phi_{S,\delta}$ be as in Theorem 5.7.4, where $\delta > 0$ will also be chosen in Step 4, and consider the class of DNNs (5.5.2) with S in place of Λ and s in place of N.

Step 2: Establishing the weighted rNSP. The main difference in this case is the use of Lemma 5.6.2 to assert the weighted rNSP instead of Lemma 5.6.1. Let $\boldsymbol{A}, \boldsymbol{A}' \in \mathbb{R}^{m\times s}$ be given by (5.5.8) and set

$$\bar{k} := \frac{m}{11L} \leq \frac{m}{2}, \tag{5.8.36}$$

where $L = L(m,\epsilon) \geq 1$ is as in (5.3.11). We now make the following assumption:

$$\bar{k} \geq k := |S|_{\boldsymbol{u}}. \tag{5.8.37}$$

Later, when we construct the set S in Step 4 we will verify that this holds. We now apply Lemma 5.6.2 with $\delta = 2/5$. Notice that

$$m = 11 \cdot \bar{k} \cdot L(m, \epsilon) \geq ((1-\delta)\log(1-\delta) + \delta)^{-1} \cdot k \cdot \log(2k/\epsilon).$$

Hence, with probability at least $1 - \epsilon/2$, the matrix $\boldsymbol{A}$ has the weighted rNSP over $\mathbb{R}$ of order $(k, \boldsymbol{u})$ with constants $\rho = 0$ and $\gamma = \sqrt{5/3}$. Or equivalently (recall the proof of Lemma 5.6.2), the bound

$$\|\boldsymbol{x}\|_2 \leq \sqrt{5/3}\|\boldsymbol{A}\boldsymbol{x}\|_2, \quad \forall \boldsymbol{x} \in \mathbb{R}^s, \tag{5.8.38}$$

holds with probability at least $1 - \epsilon/2$. Now, much as before, we have

$$\|\boldsymbol{A} - \boldsymbol{A}'\|_2 \leq \sqrt{s}\delta. \tag{5.8.39}$$

Suppose that

$$\sqrt{s}\delta \leq \frac{\sqrt{3}}{2\sqrt{5}}. \tag{5.8.40}$$

Then, if (5.8.38) holds, we have

$$\|\boldsymbol{x}\|_2 \leq \sqrt{5/3}\|\boldsymbol{A}'\boldsymbol{x}\|_2 + \|\boldsymbol{x}\|_2/2, \quad \forall \boldsymbol{x} \in \mathbb{R}^s,$$

which implies that

$$\|\boldsymbol{x}\|_2 \leq 2\sqrt{5/3}\|\boldsymbol{A}'\boldsymbol{x}\|_2, \quad \forall \boldsymbol{x} \in \mathbb{R}^s.$$

We deduce that, with probability at least $1 - \epsilon/2$, $\boldsymbol{A}'$ has the weighted rNSP over $\mathbb{R}$ of order $(k, \boldsymbol{u})$ with constants $\rho = 0$ and $\gamma = 2\sqrt{5/3}$. Finally, applying Lemma 3.6.7 (and recalling that $\boldsymbol{u} \geq 1$), we deduce that the corresponding operator $\boldsymbol{A}' \in \mathcal{B}(\mathcal{V}^s, \mathcal{V}^m)$ satisfies the weighted rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{u})$ with constants $\rho' = 0$ and $\gamma' = 2\sqrt{k}\sqrt{5/3}$, with the same probability.

Step 3: Estimating the error. Let $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$. This step is again similar to Step 3 of the proof of Theorem 5.3.1. We first write

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \leq A_1 + A_2 + A_3 + A_4,$$
$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq B_1 + B_2 + B_3 + B_4,$$

with A_1, A_2, A_3, A_4, B_1, B_2, B_3 and B_4, defined in the same way, except with Λ replaced by S throughout.

Step 3(i): Bounding A_1 and B_1. This step is identical and gives $A_1 \leq B_2 \leq E_{\mathsf{disc}}$.

Step 3(ii): Bounding A_2 and B_2. This step is identical and gives $A_2 \leq \pi_K E_{S,2}(f)$ and $B_2 \leq \pi_K E_{S,\infty}(f)$.

Step 3(iii): Bounding A_3 and B_3. Let $\boldsymbol{u}$ be the intrinsic weights in (2.4.17), using the definition of k in (5.8.37) and the Cauchy-Schwarz inequality we get

$$A_3 \leq B_3 \leq \|\boldsymbol{c}_{S,K} - \hat{\boldsymbol{c}}\|_{1,\boldsymbol{u};\mathcal{V}} = \sum_{\boldsymbol{\nu}\in S} \boldsymbol{u}_{\boldsymbol{\nu}} \|\boldsymbol{c}_{\boldsymbol{\nu},K} - \hat{\boldsymbol{c}}_{\boldsymbol{\nu}}\|_{\mathcal{V}} \leq \left(\sum_{\boldsymbol{\nu}\in S} \boldsymbol{u}_{\boldsymbol{\nu}}^2\right)^{1/2} \left(\sum_{\boldsymbol{\nu}\in S} \|\boldsymbol{c}_{\boldsymbol{\nu},K} - \hat{\boldsymbol{c}}_{\boldsymbol{\nu}}\|_{\mathcal{V}}^2\right)^{1/2}$$

and therefore

$$B_3 \leq \sqrt{k} \|\boldsymbol{c}_{S,K} - \hat{\boldsymbol{c}}\|_{2;\mathcal{V}}.$$

We now use the weighted rNSP for $\boldsymbol{A}'$ to deduce that

$$B_3 \lesssim k \|\boldsymbol{A}'(\boldsymbol{c}_{S,K} - \hat{\boldsymbol{c}})\|_{2;\mathcal{V}} \leq k \left(\|\boldsymbol{A}'\boldsymbol{c}_{S,K} - \boldsymbol{f}\|_{2;\mathcal{V}} + \|\boldsymbol{A}'\hat{\boldsymbol{c}} - \boldsymbol{f}\|_{2;\mathcal{V}} \right).$$

Now $\hat{\boldsymbol{c}}$ is an approximate minimizer of (5.5.9) and $\boldsymbol{c}_{S,K} \in \mathcal{V}_K^s$ is feasible. Therefore

$$B_3 \lesssim k \left(2\|\boldsymbol{A}'\boldsymbol{c}_{S,K} - \boldsymbol{f}\|_{2;\mathcal{V}} + E_{\mathsf{opt}} \right).$$

Via the same arguments as before, we now bound

$$\|\boldsymbol{A}'\boldsymbol{c}_{S,K} - \boldsymbol{f}\|_{2;\mathcal{V}} \leq \pi_K s\delta + \pi_K \sqrt{\frac{1}{m}\sum_{i=1}^m \|f(\boldsymbol{y}_i) - f_S(\boldsymbol{y}_i)\|_{\mathcal{V}}^2} + E_{\mathsf{disc}} + E_{\mathsf{samp}}.$$

Now, it follows from (5.8.36) and (5.8.37) that $m = 11\bar{k}L \geq 11k(\log(m) + \log(1/\epsilon)) \geq 2k\log(4/\epsilon)$. Hence, a minor modification of [12, Lemma 7.11] gives

$$\sqrt{\frac{1}{m}\sum_i \|f(\boldsymbol{y}_i) - f_S(\boldsymbol{y}_i)\|_{\mathcal{V}}^2} \leq \sqrt{2}\left(\frac{E_{S,\infty}(f)}{\sqrt{k}} + E_{S,2}(f) \right)$$

with probability at least $1 - \epsilon/2$. Combining this with the previous bound, we deduce that

$$A_3 \leq B_3 \lesssim \pi_K \sqrt{k} \left(\sqrt{k} s\delta + E_{S,\infty}(f) + \sqrt{k} E_{S,2}(f) \right) + k \left(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} \right), \quad (5.8.41)$$

with probability at least $1 - \epsilon$.

Step 3(iv). Bounding A_4 and B_4. As in the corresponding step in the previous proofs, we first write

$$A_4 \leq B_4 \leq \|f_{\hat{\Psi}} - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq \sum_{\boldsymbol{\nu}\in S} \|\Psi_{\boldsymbol{\nu}} - \Phi_{\boldsymbol{\nu},\delta,\Theta}\|_{L^\infty(\mathcal{U})} \|\hat{c}_{\boldsymbol{\nu}}\|_{\mathcal{V}} \leq \delta\sqrt{s} \|\hat{\boldsymbol{c}}\|_{2;\mathcal{V}}.$$

Since $\boldsymbol{A}'$ has the weighted rNSP and $\hat{\boldsymbol{c}}$ is an approximate minimizer, we get

$$\begin{aligned}\|\hat{\boldsymbol{c}}\|_{2;\mathcal{V}} &\lesssim \sqrt{k}\|\boldsymbol{A}'\hat{\boldsymbol{c}}\|_{2;\mathcal{V}} \\ &\leq \sqrt{k}(\|\boldsymbol{A}'\hat{\boldsymbol{c}}-\boldsymbol{f}\|_{2;\mathcal{V}}+\|\boldsymbol{f}\|_{2;\mathcal{V}}) \\ &\leq \sqrt{k}(\|\boldsymbol{A}'\boldsymbol{0}-\boldsymbol{f}\|_{2;\mathcal{V}}+\|\boldsymbol{f}\|_{2;\mathcal{V}}+E_{\mathsf{opt}}) \\ &\leq \sqrt{k}(E_{\mathsf{samp}}+1+E_{\mathsf{opt}}).\end{aligned}$$

Note that $\delta\sqrt{s} \lesssim 1$ due to (5.8.40). Hence we obtain

$$B_4 \lesssim \sqrt{s}\sqrt{k}\delta + \sqrt{k}(E_{\mathsf{samp}}+E_{\mathsf{opt}}).$$

Step 3(v). Final bound. Combining the estimates for A_1, A_2, A_3, A_4, B_1, B_2, B_3 and B_4 from the previous substeps and noticing that $\pi_K \geq 1$ by definition and $k = |S|_{\boldsymbol{u}} \geq 1$ (since $\boldsymbol{u} \geq \boldsymbol{1}$), we obtain

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \lesssim \pi_K\left(ks\delta + \sqrt{k}E_{S,\infty}(f) + kE_{S,2}(f)\right) + k\left(E_{\mathsf{opt}}+E_{\mathsf{disc}}+E_{\mathsf{samp}}\right), \tag{5.8.42}$$

and the same bound for the L^∞-norm error

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \pi_K\left(ks\delta + \sqrt{k}E_{S,\infty}(f) + kE_{S,2}(f)\right) + k\left(E_{\mathsf{opt}}+E_{\mathsf{disc}}+E_{\mathsf{samp}}\right). \tag{5.8.43}$$

Step 4: Establishing the algebraic rates. Suppose that $\boldsymbol{b} \in \ell^p(\mathbb{N})$ (we address the case $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ in Step 6 below). We now make a suitable choice of S so as to obtain the desired algebraic rates of convergence.

We first apply part (ii) of Theorem 2.4.13 with $\bar{k}$ in place of k. Let $q = 1$. Then this guarantees the existence of a set S_1 with $|S_1|_{\boldsymbol{u}} \leq \bar{k}$ such that

$$\|\boldsymbol{c}-\boldsymbol{c}_{S_1}\|_{2;\mathcal{V}} \leq \|\boldsymbol{c}-\boldsymbol{c}_{S_1}\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b},\varepsilon,p)\cdot\bar{k}^{1-1/p}. \tag{5.8.44}$$

We now define

$$S = S_1 \cap \Lambda, \quad \text{where } \Lambda = \Lambda^{\mathsf{HC}}_{\lceil\bar{k}\rceil,\infty} = \left\{\boldsymbol{\nu}=(\nu_k)_{k=1}^\infty \in \mathcal{F} : \prod_{k:\nu_k\neq 0}(\nu_k+1) \leq \lceil\bar{k}\rceil\right\}. \tag{5.8.45}$$

Observe that $|S|_{\boldsymbol{u}} \leq |S_1|_{\boldsymbol{u}} \leq \bar{k}$. Therefore (5.8.37) holds for this choice of S. Note also that S is independent of $f \in \mathcal{H}(\boldsymbol{b},\varepsilon)$ and depends only on $\boldsymbol{b}$, ε (see Remark 2.4.16).

Having defined S, we now bound

$$E_{S,2}(f) \leq E_{S,\infty}(f) = \|f-f_S\|_{L^\infty_{\varrho}(\mathcal{U};\mathcal{V})} \leq \|\boldsymbol{c}-\boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b},\varepsilon,p)\cdot\bar{k}^{1-1/p} + \|\boldsymbol{c}-\boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}}.$$

Now, the set Λ is precisely the union of all lower sets (see Definition 2.4.9) of size at most $\lceil \bar{k} \rceil$ (see, e.g., [12, Prop. 2.5]). Hence, by part (i) of Theorem 2.4.13,

$$\|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1-1/p}.$$

Since $k \leq \bar{k}$, we deduce that

$$\sqrt{k} E_{S,\infty}(f) + k E_{S,2}(f) \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{2-1/p}.$$

Substituting this bound into (5.8.42) and (5.8.43) gives

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim C(\boldsymbol{b}, \varepsilon, p) \cdot \pi_K \left(ks\delta + \bar{k}^{2-1/p} \right) + k \left(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} \right)$$

and

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim C(\boldsymbol{b}, \varepsilon, p) \cdot \pi_K \left(ks\delta + \bar{k}^{2-1/p} \right) + k \left(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} \right).$$

We now set

$$\delta = \min \left\{ \frac{\sqrt{3}}{2\sqrt{5}\sqrt{\bar{k}}}, \bar{k}^{-\frac{1}{p}} \right\}. \tag{5.8.46}$$

Notice that (5.8.40) holds for this choice of δ, since $s = |S| \leq |S|_{\boldsymbol{u}} = k \leq \bar{k}$. Substituting this into the previous expression and using the definition (5.8.36) of $\bar{k}$ and (5.8.37) now gives

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim E_{\mathsf{app,KB}} + m \left(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} \right), \tag{5.8.47}$$

and

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim E^\infty_{\mathsf{app,KB}} + m \left(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}} \right), \tag{5.8.48}$$

as required.

Step 5: Bounding the width and depth of the DNN architecture. We first consider Θ. Recall that Θ must satisfy (5.8.35). By construction, any $\boldsymbol{\nu} \in S$ is also an element of Λ, and therefore $2^{\|\boldsymbol{\nu}\|_0} \leq \lceil \bar{k} \rceil$. Hence $|\mathrm{supp}(\boldsymbol{\nu})| = \|\boldsymbol{\nu}\|_0 \leq \log_2(\lceil \bar{k} \rceil)$. Since $|S| = s$, we deduce that

$$\left| \bigcup_{\boldsymbol{\nu} \in S} \mathrm{supp}(\boldsymbol{\nu}) \right| \leq s \log_2(\lceil \bar{k} \rceil).$$

Observe that $s = |S| \leq |S|_{\boldsymbol{u}} \leq \bar{k}$, since $\boldsymbol{u} \geq 1$. Using the definition (5.8.36) of $\bar{k}$, the fact that $m \geq 3$ and the definition (5.3.11) of L, we see that

$$s \log_2(\lceil \bar{k} \rceil) \leq \frac{m}{11L} \frac{\log(m)}{\log(2)} \leq \frac{m}{11 \log(2)} \leq n,$$

where n is as in (5.3.12). Thus, we now choose Θ as any set of size n that satisfies (5.8.35).

We now estimate the width and depth of the DNN architecture. First, observe that

$$\log(\delta^{-1}) \lesssim p^{-1}\log(\bar{k}) \leq p^{-1}\log(m).$$

In addition, due to the choice (5.8.45), we have

$$m(S) = \max_{\boldsymbol{\nu}\in S} \|\boldsymbol{\nu}\|_1 \leq \max_{\boldsymbol{\nu}\in\Lambda} \|\boldsymbol{\nu}\|_1 \leq \lceil \bar{k} \rceil \leq m. \tag{5.8.49}$$

Hence, applying Theorem 5.7.4 with the set S in place of Λ and Θ as chosen above, we deduce that the width and depth in the case of the ReLU activation function satisfy

$$\mathrm{width}(\mathcal{N}^1) \lesssim m^2, \quad \mathrm{depth}(\mathcal{N}^1) \lesssim \left(1 + \log(m)\left(p^{-1}\log(m) + m\right)\right).$$

Here, we also used the facts that $s = |S| \leq k \leq \bar{k} \leq m$ and $n \lesssim m$. Now, for either the RePU or tanh activation function, we have

$$\mathrm{width}(\mathcal{N}^j) \leq c_{j,1}\cdot m^2, \quad \mathrm{depth}(\mathcal{N}^j) \leq c_{j,2}\cdot \log_2(m),$$

where $c_{j,1}$, $c_{j,2}$ are universal constants for the tanh activation function ($j = 0$) and $c_{j,1}, c_{j,2}$ depend on ℓ for the RePU activation function ($j = \ell$). This gives the desired bounds.

Step 6: Modifying the proof in the case $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$. In this case, we replace the definition of S in (5.8.45) with

$$S = S_1 \cap \Lambda, \quad \text{where } \Lambda = \Lambda^{\mathsf{HCI}}_{\lceil \bar{k} \rceil},$$

and $\Lambda^{\mathsf{HCI}}_{\lceil \bar{k} \rceil}$ is as in (2.4.21). Recall from the discussion in §2.5.2 that this set contains all anchored sets of size at most $\lceil \bar{k} \rceil$. Thus, we may argue as in Step 4, but using Corollary 2.4.15 instead of Theorem 2.4.13 to bound the error $\boldsymbol{c} - \boldsymbol{c}_\Lambda$. Doing so, and using exactly the same value for δ yields an identical bounds (5.8.47) and (5.8.48).

We now modify Step 5 accordingly. By definition of $\Lambda^{\mathsf{HCI}}_{\lceil \bar{k} \rceil}$, any multi-index $\boldsymbol{\nu} \in S$ must satisfy $\mathrm{supp}(\boldsymbol{\nu}) \subseteq \{1, \dots, \lceil \bar{k} \rceil\}$. It follows from (5.8.36) that $\lceil \bar{k} \rceil \leq n$, where n is as in (5.3.17). Hence we may take Θ as in (5.3.17). Finally, we note that (5.8.49) also holds for this choice of S. Thus the bounds for the widths and depths of the various DNN classes hold in this case as well. □

5.8.4 Theorem 5.3.4: known anisotropy, Hilbert-valued case

Proof of Theoorem 5.3.4. The proof involves several modifications to that of the previous theorem. Step 1 is identical. In Step 2, instead of Lemma 3.6.7 we use Lemma 3.6.6 and Lemma 3.6.5 to deduce that the operator $\boldsymbol{A}' \in \mathcal{B}(\mathcal{V}^s, \mathcal{V}^m)$ has the weighted rNSP over $\mathcal{V}$ of

order $(k, \boldsymbol{u})$ with k-independent constants $\rho' = 0$ and $\gamma' = 2\sqrt{5/3}$, with probability at least $1 - \epsilon/2$.

Step 3: Estimating the error. Let $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$. Consider the same setup as before, with

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq A_1 + A_2 + A_3 + A_4,$$

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \leq B_1 + B_2 + B_3 + B_4,$$

and A_1, A_2, A_3, A_4 and B_1, B_2, B_3 and B_4 defined in the same way. Step 3(i) and Step 3(ii) are identical.

Step 3(iii): Bounding A_3 and B_3. As in the corresponding step in the proof of Theorem 5.3.2, using Parseval's identity we first write

$$A_3 = \|\boldsymbol{c}_{S,K} - \hat{\boldsymbol{c}}\|_{2;\mathcal{V}}.$$

Following the same arguments as before, noticing that $\gamma' \lesssim 1$, we deduce that

$$A_3 \lesssim \pi_K \left(\sqrt{s}\delta + \frac{E_{S,\infty}(f)}{\sqrt{k}} + E_{S,2}(f) \right) + E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}.$$

with probability at least $1 - \epsilon$. By using the same arguments as in the previous proof, noticing that $\gamma' \lesssim 1$, we only see one additional factor $\sqrt{k}$ bounding B_3, so we obtain

$$B_3 \lesssim \pi_K \sqrt{k} \left(\sqrt{s}\delta + \frac{E_{S,\infty}(f)}{\sqrt{k}} + E_{S,2}(f) \right) + \sqrt{k}(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}),$$

with probability at least $1 - \epsilon$.

Step 3(iv). Bounding A_4 and B_4. Using the same arguments as in the corresponding step in the previous proofs, noticing once more that $\gamma' \lesssim 1$, we obtain

$$A_4 \leq B_4 \lesssim \sqrt{s}\delta + E_{\mathsf{samp}} + E_{\mathsf{opt}}. \tag{5.8.50}$$

Step 3(v). Final bound. Combining the estimates for A_1, A_2, A_3, A_4, B_1, B_2, B_3 and B_4 from the previous substeps we obtain

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim \pi_K \left(\sqrt{s}\delta + \frac{E_{S,\infty}(f)}{\sqrt{k}} + E_{S,2}(f) \right) + E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}$$

and

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \pi_K \sqrt{k} \left(\sqrt{s}\delta + \frac{E_{S,\infty}(f)}{\sqrt{k}} + E_{S,2}(f) \right) + \sqrt{k}(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}).$$

Step 4: Establishing the algebraic rates. Suppose that $\boldsymbol{b} \in \ell^p(\mathbb{N})$. We now apply part (ii) of Theorem 2.4.13 with $\bar{k}/2$ in place of k. Let $q = 1$. Then this guarantees the existence of a set S_1 with $|S_1|_{\boldsymbol{u}} \leq \bar{k}/2$ such that

$$\|\boldsymbol{c} - \boldsymbol{c}_{S_1}\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1-1/p}.$$

Similarly, letting $q = 2$, we obtain a set S_2 with $|S_2|_{\boldsymbol{u}} \leq \bar{k}/2$ such that

$$\|\boldsymbol{c} - \boldsymbol{c}_{S_2}\|_{2;\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1/2-1/p}.$$

Instead of (5.8.45), we now define

$$S = (S_1 \cup S_2) \cap \Lambda, \quad \text{where } \Lambda = \Lambda^{\mathsf{HC}}_{\lceil \bar{k} \rceil, \infty} = \left\{ \boldsymbol{\nu} = (\nu_k)_{k=1}^{\infty} \in \mathcal{F} : \prod_{k:\nu_k \neq 0} (\nu_k + 1) \leq \lceil \bar{k} \rceil \right\}. \tag{5.8.51}$$

Observe that $|S|_{\boldsymbol{u}} \leq |S_1|_{\boldsymbol{u}} + |S_2|_{\boldsymbol{u}} \leq \bar{k}$. Therefore (5.8.37) holds for this choice of S. Once more S is independent of $f \in \mathcal{H}(\boldsymbol{b}, \varepsilon)$ and depends only on $\boldsymbol{b}$, ε.

Having defined S, we now bound

$$E_{S,\infty}(f) = \|f - f_S\|_{L^\infty_\varrho(\mathcal{U};\mathcal{V})} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1-1/p} + \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}}$$

and by Parseval's identity

$$E_{S,2}(f) = \|f - f_S\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \|\boldsymbol{c} - \boldsymbol{c}_S\|_{2;\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1/2-1/p} + \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{2;\mathcal{V}}.$$

Following similar arguments as before, by part (i) of Theorem 2.4.13, we have

$$\|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{2;\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1/2-1/p} \quad \text{and} \quad \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{1,\boldsymbol{u};\mathcal{V}} \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1-1/p}. \tag{5.8.52}$$

Since $k \leq \bar{k}$, we deduce that

$$\frac{E_{S,\infty}(f)}{\sqrt{k}} + E_{S,2}(f) \leq C(\boldsymbol{b}, \varepsilon, p) \cdot \bar{k}^{1/2-1/p}.$$

Substituting this bound into (5.8.42) and (5.8.43) gives

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim \pi_K \left(\sqrt{s}\delta + \bar{k}^{1/2-1/p} \right) + E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}$$

and

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim \pi_K \sqrt{k} \left(\sqrt{s}\delta + \bar{k}^{1/2-1/p} \right) + \sqrt{k}(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}).$$

Arguing in the same way as Step 4, and making a similar choice as in (5.8.46) for δ, we see that

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \lesssim E_{\mathsf{app},\mathsf{KH}} + E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}},$$

and

$$\|f - f_{\hat{\Phi},\Theta}\|_{L^\infty(\mathcal{U};\mathcal{V})} \lesssim E^\infty_{\mathsf{app},\mathsf{KH}} + m^{1/2}(E_{\mathsf{opt}} + E_{\mathsf{disc}} + E_{\mathsf{samp}}),$$

as required.

Step 5 is identical. For Step 6, we replace the definition of S in (5.8.51) with

$$S = (S_1 \cup S_2) \cap \Lambda, \quad \text{where } \Lambda = \Lambda^{\mathsf{HCI}}_{\lceil \bar{k} \rceil}, \tag{5.8.53}$$

and $\Lambda^{\mathsf{HCI}}_{\lceil \bar{k} \rceil}$ is as in (2.4.21). Finally, we note that (5.8.49) also holds for this choice of S. Thus the bounds for the widths and depths of the various DNNs hold in this case as well. □

5.9 Conclusions

The main results in this chapter demonstrate the existence of sample-efficient training procedures for approximating infinite-dimensional, holomorphic functions taking values in Hilbert or Banach spaces using DNNs. They account for all main sources of error in the problem through the approximation error E_{app}, the physical discretization error E_{disc}, the sampling error E_{samp} and the optimization error E_{opt}. Note that the second error E_{disc} is given in terms of the linear operator $\mathcal{P}_K(f)$, where this operator is only used to provide a bound for E_{disc} and it is not used in the training procedure.

From Hilbert-valued to the Banach-valued case

In the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm considering the Hilbert-valued case, our results are optimal up to constants and (poly)logarithmic factors. Specifically, the rate $m^{1/2-1/p}$ is the best achievable for the class of infinite-dimensional functions considered, regardless of sampling strategy or learning procedure.

Keeping in mind our four theorems we can answer Question 5 of §1.6 in the affirmative.

Answer to Question 5
It is possible to learn infinite-dimensional Hilbert-valued or Banach-valued approximations to functions from a limited dataset using DNNs with a complete theoretical understanding of the sample complexity and algebraic approximation rates. Moreover, these approximation error rates overcome the curse of dimensionality and, in the Hilbert-valued case, are optimal up to constatnts and (poly)logarithmic factors in the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm.

In addition, we answer Question 9 of §1.6 for the setting in this chapter.

Answer to Question 9

Banach-valued case

For the unknown anisotropy, the terms E_{samp}, E_{opt} and E_{disc} enter the error multiplied by a factor of $m^{1/2}$ in the the $L^2_\varrho(\mathcal{U};\mathcal{V})$ and $L^\infty(\mathcal{U};\mathcal{V})$-norms. Conversely, for the known anisotropy, these terms enter the error multiplied by a factor of m in the the $L^2_\varrho(\mathcal{U};\mathcal{V})$ and $L^\infty(\mathcal{U};\mathcal{V})$-norms.

Hilbert-valued case

For the unknown anisotropy, the terms E_{samp}, E_{opt} and E_{disc} enter the error linearly in the the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm and are multiplied by a factor of $m^{1/2}$ in the $L^\infty(\mathcal{U};\mathcal{V})$-norm. For the known anisotropy, these terms enter the error linearly in the the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm and are multiplied by a factor of $m^{1/2}$ in the $L^\infty(\mathcal{U};\mathcal{V})$-norm.

5.10 Future works

There are several interesting directions for future research

- First, whether or not the rates of decay of the approximation error can be improved in the Banach-valued case is an open problem. We conjecture that they can, and optimal rates can be shown for Banach-valued function approximation with DNNs. However, this will require a different approach and is the subject of a future work.

  It is worth noting, however, that better rates can be shown for certain spaces $\mathcal{V}_K$. In particular, following a different proof strategy [19, §13.2.1], one could use the basis $\{\varphi_i\}_{i=1}^K$ to assert a wRIP-type bound of the form

  $$\alpha_K(1-\delta)\|\boldsymbol{v}\|_{2;\mathcal{V}} \leq \|\boldsymbol{A}\boldsymbol{v}\|_{2;\mathcal{V}} \leq \beta_K(1+\delta)\|\boldsymbol{v}\|_{2;\mathcal{V}}$$

  for all s-sparse vectors $\boldsymbol{v} \in \mathcal{V}_K^N$, where α_K and β_K depend the space $\mathcal{V}_K$. However, this dependence leads to (typically undesirable) convergence rates of the type $(\varpi_K m)^{1/2-1/p}$, where ϖ_K depends on β_K/α_K. The construction and implementation of suitable discretizations – i.e., those for which $\beta_K/\alpha_K \lesssim 1$ – is nontrivial and go beyond the scope of this work (see, e.g., [263] and references therein for more information on this topic).

- As mentioned several times above, while our results in the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm rate when $\mathcal{V}$ is a Hilbert space are optimal (up to the polylogarithmic factor L), it is currently

unknown whether the rate $(m/L)^{1-1/p}$ is optimal for the $L^\infty_\varrho(\mathcal{U};\mathcal{V})$-norm. Indeed, Chapter 6 only considers optimal approximation rates in the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm and it is unclear if it can be extended to the $L^\infty_\varrho(\mathcal{U};\mathcal{V})$-norm.

- As described in [42], existing DL approaches for parametric DEs are often not robust when the mesh (e.g., a finite difference or FE mesh) used to simulate the training data is refined, since the DNN architecture depends on the mesh size. In our work, the DNN architecture depends on the space $\mathcal{V}_K$. This could correspond to the mesh used to generate the data, leading to a non-mesh invariant approach. But, as in [42], $\mathcal{V}_K$ could also be constructed in a different manner, e.g., via PCA in the case of Hilbert spaces, leading to a mesh-invariant scheme. There are various open problems in this direction; for example, how to compute a reduced dimension space $\mathcal{V}_K$ when $\mathcal{V}$ is a Banach space, or how to perform adaptive mesh refinement (see, e.g., [106, 107]).

- We have not strived to optimize the width and depths of the corresponding DNNs. In the RePU and tanh cases, the depth grows logarithmically in m, which is reasonable in practice. However, the width bounds are $\mathcal{O}\left(m^2\right)$ in the known anisotropy case and $\mathcal{O}\left(m^{3+\log_2(m)}\right)$ in the unknown case. The latter, in particular, grows superalgebraically in m. In the known anisotropy case, it may be possible to reduce this quadratic scaling by finding a more efficient way to emulate polynomials using DNNs. However, the primary reason for the superalgebraic growth in the unknown case stems not from the specific emulation procedure, but from the need to emulate all polynomials in the large index set (2.4.21): recall the cardinality bound (2.4.23). Finding a way to avoid forming all polynomials in this index set would be useful not just for improving the width bounds in these practical existence theorems. It would also be extremely helpful for the underlying compressed sensing-based polynomial approximation schemes, as these schemes suffer from high computational cost for precisely this reason. See Chapter 4.

Chapter 6

Optimal learning of holomorphic functions

The purpose of this chapter is to provide theoretical approximation guarantees for the class of functions described in §2.3.1, demonstrating that the algebraic rates obtained in Chapters 3–5 are close to the best possible rates for approximating infinite-dimensional, Banach-valued holomorphic functions from limited samples. We begin in §6.1 by introducing key preliminary concepts, the setup, and new definitions. In §6.2, we outline our main contributions, and in §6.3, we present the main results. We discuss important aspects of these results in §6.4. In §6.5, we introduce additional concepts and notations for the proofs, which we then provide in §6.6. Finally, we draw some conclusions in §6.7, answer Question 8 of §1.6, and discuss future work in §6.8. This chapter also relies on two appendices (Appendices B and C) that present several technical results needed for the main arguments in this chapter.

The content of this chapter is primarily derived from [18].

6.1 Preliminaries

In Chapters 3–5 we presented a series of convergence rates of the form $\mathcal{O}((m/\text{polylog}(m))^{1/2-1/p})$ for the approximation of Hilbert-valued, $(\boldsymbol{b},\varepsilon)$-holomorphic functions in infinite dimensions. Specifically, these rates are achieved for the unknown anisotropy case in Chapters 3–5 and for the known anisotropy case in Chapter 5. The approximation of such functions from finitely many samples is of particular interest, and motivated by their applications in processes that can be represented by a function $f : \boldsymbol{y} \in \mathcal{U} \mapsto f(\boldsymbol{y}) \in \mathcal{V}$ arising as a solution of a (system of) parametric DEs (see Chapter 1). Consequently, it is crucial for this thesis to establish whether the algebraic decay rates in §2.4.3 represent the optimal approximation rates, with respect to the number of samples m, that the non-adaptive (random) sampling and recovery strategies developed in Chapters 3-5 achieve.

Before stating the problem formally in §6.1.5, we require some setup and definitions.

6.1.1 Setup

In this chapter we consider a Banach space $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$. As in §2.2, we consider a nonnegative Borel probability measure $\varrho^{(1)}$ on the interval $[-1,1]$. We then define ϱ as a tensor product of measures $\varrho^{(1)}$, and we denote this by

$$\varrho = \varrho^{(1)} \times \varrho^{(1)} \times \cdots .$$

The existence of such measure on $[-1,1]^{\mathbb{N}}$ is guaranteed by the Kolmogorov extension theorem (see, e.g. [252, §2.4]). A typical example we consider in this chapter is case where $\varrho^{(1)}$ is the uniform probability measure, i.e., the tensor product of the univariate measure $\mathrm{d}\rho(y) = \frac{1}{2}\,\mathrm{d}y$ as in (2.2.1). Note that in Chapters 3–5 we focus on the uniform or Chebyshev measure only.

Recall that a function $f : [-1,1]^{\mathbb{N}} \to \mathcal{V}$ is $(\boldsymbol{b},\varepsilon)$-*holomorphic* §2.3 if it is holomorphic in the region $\mathcal{R}(\boldsymbol{b},\varepsilon)$ defined in (2.3.2). In addition, recall the definition of the space of holomorphic functions with L^∞-norm at most one $\mathcal{H}(\boldsymbol{b},\varepsilon)$ defined in (2.3.3). In this chapter, for simplicity we consider $\varepsilon = 1$ (see §2.3.1). As in the previous chapters, we refer to $\boldsymbol{b} \geq \boldsymbol{0}$ as the *anisotropy* parameter of a function $f \in \mathcal{H}(\boldsymbol{b})$. For more discussion on how $\boldsymbol{b}$ relates to the anisotropy of f see §2.3.1.

As mentioned in §2.3.2, an important consideration in this thesis is whether $\boldsymbol{b}$ is known or unknown. In the *known anisotropy* setting, a reconstruction procedure can use $\boldsymbol{b}$ to achieve a good approximation uniformly over the class $\mathcal{H}(\boldsymbol{b})$. However, in the *unknown anisotropy* setting, the reconstruction procedure has no access to $\boldsymbol{b}$. In this case, motivated by the best s-term approximation theory described in §2.4.3, we fix a $0 < p < 1$ and consider the classes

$$\mathcal{H}(p) = \bigcup \left\{ \mathcal{H}(\boldsymbol{b}) : \boldsymbol{b} \in \ell^p(\mathbb{N}),\ \boldsymbol{b} \in [0,\infty)^{\mathbb{N}},\ \|\boldsymbol{b}\|_p \leq 1 \right\}$$

and

$$\mathcal{H}(p,\mathsf{M}) = \bigcup \left\{ \mathcal{H}(\boldsymbol{b}) : \boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N}),\ \boldsymbol{b} \in [0,\infty)^{\mathbb{N}},\ \|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1 \right\},$$

where $\ell^p_{\mathsf{M}}(\mathbb{N})$ is the monotone ℓ^p-space defined in §2.1. One of the reasons to introduce these spaces is that, as we will see later in §6.3.1, it is generally impossible to approximate $(\boldsymbol{b},\varepsilon)$-holomorphic functions in infinite dimensions from limited samples without any assumption on $\boldsymbol{b}$. Therefore we must define a space that allows to formally quantify this aspect.

Given this setup, in this chapter we pose the following question: *How well can we uniformly approximate functions in $\mathcal{H}(\boldsymbol{b})$, for fixed $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with $\boldsymbol{b} \geq \boldsymbol{0}$, $\mathcal{H}(p)$ or $\mathcal{H}(p,\mathsf{M})$ from m samples, and how does this reconstructions depend on $\boldsymbol{b}$ (in the first case) and p (in the first, second and third cases)?*

In the following section we provide necessary definitions to formally state the problem in this chapter.

6.1.2 Sampling operators

We now define the concept of adaptive sampling operator. Note that this is commonly referred to as *adaptive information* in the field of information-based complexity [211, §4.1.1]. Since it is generally simpler than the Banach-valued case, we commence with the scalar-valued case $\mathcal{V} = \mathbb{R}$.

Definition 6.1.1 (Adaptive sampling operator; scalar-valued case)**.** Consider a normed vector space $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$. A scalar-valued *adaptive sampling operator* is a map of the form

$$\boldsymbol{\mathcal{S}} : \mathcal{Y} \to \mathbb{R}^m, \quad \boldsymbol{\mathcal{S}}(f) = \begin{bmatrix} S_1(f) \\ S_2(f; S_1(f)) \\ \vdots \\ S_m(f; S_1(f), \dots, S_{m-1}(f)) \end{bmatrix},$$

where $S_1 : \mathcal{Y} \to \mathbb{R}$ is a bounded linear functional and, for $i = 2, \dots, m$, $S_i : \mathcal{Y} \times \mathbb{R}^{i-1} \to \mathbb{R}$ is bounded and linear in its first component.

Notice that any linear map $\boldsymbol{\mathcal{S}} : \mathcal{Y} \to \mathbb{R}^m$ is an adaptive sampling operator. The rationale for considering adaptive sampling operators is to cover approximation methods where each subsequent sample is chosen adaptively in terms of the previous measurements.

Note also that this definition allows for arbitrary adaptive sampling operators. An important special case is that of (adaptive) pointwise samples (so-called *standard information* [211, §4.1.1]). Let $\mathcal{Y} = C(\mathcal{U})$. Then this is defined as

$$\boldsymbol{\mathcal{S}}(f) = (f(\boldsymbol{y}_i))_{i=1}^m \in \mathbb{R}^m, \quad \forall f \in C(\mathcal{U}), \tag{6.1.1}$$

where $\boldsymbol{y}_i$ is the ith sample point, which is potentially chosen adaptively based on the previous measurements $f(\boldsymbol{y}_1), \dots, f(\boldsymbol{y}_{i-1})$.

Next, we consider the Banach-valued case. In the following definition, for any $w \in \mathcal{V}$ and $\boldsymbol{v} = (v_i)_{i=1}^m \in \mathbb{R}^m$, we write $w\boldsymbol{v}$ for the vector $(wv_i)_{i=1}^m \in \mathcal{V}^m$. Note that in this definition, we consider Lebesgue–Bochner spaces only, as opposed to arbitrary Banach spaces.

Definition 6.1.2 (Adaptive sampling operator; Banach-valued case)**.** Consider a vector space $\mathcal{Y} \subseteq L^2_\varrho(\mathcal{U}; \mathcal{V})$ with norm $\|\cdot\|_{\mathcal{Y}}$ and an operator

$$\boldsymbol{\mathcal{S}} : \mathcal{Y} \to \mathcal{V}^m, \quad \boldsymbol{\mathcal{S}}(f) = \begin{bmatrix} S_1(f) \\ S_2(f; S_1(f)) \\ \vdots \\ S_m(f; S_1(f), \dots, S_{m-1}(f)) \end{bmatrix},$$

where $S_1 : \mathcal{Y} \to \mathcal{V}$ is a bounded linear operator and, for $i = 2, \dots, m$, $S_i : \mathcal{Y} \times \mathcal{V}^{i-1} \to \mathcal{V}$ is a bounded linear operator in its first component. Then $\boldsymbol{\mathcal{S}}$ is a *Banach-valued adaptive*

sampling operator if the following condition holds. There exist $v, w \in \mathcal{V}\backslash\{0\}$, a normed vector space $\widetilde{\mathcal{Y}} \subseteq L^2_\varrho(\mathcal{U})$ and a scalar-valued adaptive sampling operator $\widetilde{\boldsymbol{\mathcal{S}}} : \widetilde{\mathcal{Y}} \to \mathbb{R}^m$ (see Definition 6.1.1) such that if $vg \in \mathcal{Y}$ for some $g \in L^2_\varrho(\mathcal{U})$ then $g \in \widetilde{\mathcal{Y}}$ and $\boldsymbol{\mathcal{S}}(vg) = w\widetilde{\boldsymbol{\mathcal{S}}}(g)$.

Note that the condition imposed in Definition 6.1.2 is not a strong one. For example, it trivially holds in the important case of (adaptive) pointwise sampling. Here, we consider $\mathcal{Y} = C(\mathcal{U}; \mathcal{V})$ and

$$\boldsymbol{\mathcal{S}}(f) = (f(\boldsymbol{y}_i))_{i=1}^m \in \mathcal{V}^m, \quad \forall f \in \mathcal{Y},$$

where $\boldsymbol{y}_i \in \mathcal{U}$ is the ith sample point, which is potentially chosen adaptively based on the previous measurements $f(\boldsymbol{y}_1), \dots, f(\boldsymbol{y}_{i-1})$. In this case, we clearly have

$$\boldsymbol{\mathcal{S}}(vg) = v\widetilde{\boldsymbol{\mathcal{S}}}(g), \quad \forall g \in \widetilde{\mathcal{Y}} := C(\mathcal{U}),\ v \in \mathcal{V},$$

where $\widetilde{\boldsymbol{\mathcal{S}}}$ is the scalar-valued (adaptive) pointwise sampling operator (6.1.1).

The condition imposed in Definition 6.1.2 is used to establish our lower bounds – in particular, the reduction to a discrete problem in Lemma 6.6.2. It is an open problem whether these bounds hold in the Banach-valued case without this assumption.

Given the definition of a sampling operator above, we need a way to measure how good the recovery of f can be from its samples. To do so, in this chapter, we study the *(adaptive) m-width* (which is related to the *information complexity* [211, §4.1.4]).

6.1.3 Adaptive m-widths

We now generalize a standard definition (see, e.g., [44]) that measures the worse-case recovery error for a function f from a set $\mathcal{K}$. Given $f \in \mathcal{K} \subseteq \mathcal{Y}$, the adaptive sampling operator $\boldsymbol{\mathcal{S}}$ first yields m measurements belonging to the underlying Banach space $\mathcal{V}$. Then, an arbitrary *reconstruction map* $\boldsymbol{\mathcal{R}}$ takes this vector of m $\mathcal{V}$-valued measurements and produces an approximation in $L^2_\varrho(\mathcal{U}; \mathcal{V})$. Thus, we define the (adaptive) m-width of a subset $\mathcal{K} \subseteq \mathcal{Y}$, which pertains to the optimal sampling and reconstruction maps for minimizing the worst-case recovery error over the set $\mathcal{K}$.

Definition 6.1.3. Let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be a Banach space. The *(adaptive) m-width* of a subset $\mathcal{K} \subseteq \mathcal{Y}$ is a number given by

$$\Theta_m(\mathcal{K}; \mathcal{Y}, \mathcal{X}) = \inf \left\{ \sup_{f \in \mathcal{K}} \|f - \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{S}}(f))\|_{\mathcal{X}} : \boldsymbol{\mathcal{S}} : \mathcal{Y} \to \mathcal{V}^m \text{ adaptive},\ \boldsymbol{\mathcal{R}} : \mathcal{V}^m \to \mathcal{X} \right\}, \tag{6.1.2}$$

where $\mathcal{Y}$ is a normed vector subspace of the Lebesgue–Bochner space $\mathcal{X} = L^2_\varrho(\mathcal{U}; \mathcal{V})$ with $\mathcal{K} \subseteq \mathcal{Y}$, $\boldsymbol{\mathcal{R}}$ is an arbitrary reconstruction map and $\boldsymbol{\mathcal{S}}$ is an adaptive sampling operator, as in Definitions 6.1.1 (for scalar-valued functions) and 6.1.2 (for Banach-valued functions).

Note that the choice of $\mathcal{Y}$ determines the type of allowed sampling operators. For instance, if $\mathcal{Y} = C(\mathcal{U};\mathcal{V})$ is the space of continuous functions from $\mathcal{U}$ to $\mathcal{V}$ with respect to the uniform norm, then the type of allowed sampling operators includes (adaptive) pointwise sampling, whereas if $\mathcal{Y} = \mathcal{X}$ then it does not. Note, however, that $\mathcal{Y}$ plays a quite minor role in our analysis: our lower bounds (see §6.3.1) hold for arbitrary $\mathcal{Y}$, while our upper bounds (see §6.3.2) require the (mild) assumption that $\mathcal{Y}$ is compactly contained in $\mathcal{X}$. We provide some additional discussion on $\mathcal{Y}$ in Remark 6.1.4 below.

Observe that (6.1.2) generalizes standard definitions [44], where the measurements $\boldsymbol{\mathcal{S}}(f) \in \mathbb{R}^m$ are scalar-valued. We introduce this extension to allow for Banach-valued measurements, as this is relevant for sampling-based approximation methods for parametric PDEs.

6.1.4 Adaptive m-widths in the case of known and unknown anisotropy

Known anisotropy

In the definition of the the m-width in (6.1.2) the optimal reconstruction map can (and generally will) depend on the anisotropy parameter $\boldsymbol{b}$. Motivated by the discussion in §6.1.1 and to make the notation simpler, in the case of known anisotropy we define

$$\theta_m(\boldsymbol{b}) = \Theta_m(\mathcal{H}(\boldsymbol{b});\mathcal{Y},L^2_\varrho(\mathcal{U};\mathcal{V})). \tag{6.1.3}$$

In addition, our interest also lies in the case where $\boldsymbol{b} \in \ell^p(\mathbb{N})$ or $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ for some $0 < p < 1$. Thus, we also define

$$\begin{aligned} \overline{\theta_m}(p) &= \sup\left\{\theta_m(\boldsymbol{b}) : \boldsymbol{b} \in \ell^p(\mathbb{N}),\ \boldsymbol{b} \in [0,\infty)^{\mathbb{N}},\ \|\boldsymbol{b}\|_p \leq 1\right\}, \\ \overline{\theta_m}(p,\mathsf{M}) &= \sup\left\{\theta_m(\boldsymbol{b}) : \boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N}),\ \boldsymbol{b} \in [0,\infty)^{\mathbb{N}},\ \|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1\right\}. \end{aligned} \tag{6.1.4}$$

Unknown anisotropy

In the unknown anisotropy setting, we define

$$\theta_m(p) = \Theta_m(\mathcal{H}(p);\mathcal{Y},L^2_\varrho(\mathcal{U};\mathcal{V})), \quad \theta_m(p,\mathsf{M}) = \Theta_m(\mathcal{H}(p,\mathsf{M});\mathcal{Y},L^2_\varrho(\mathcal{U};\mathcal{V})). \tag{6.1.5}$$

Notice that $\theta_m(p)$ is not equivalent to $\overline{\theta_m}(p)$, and likewise for $\theta_m(p,\mathsf{M})$ and $\overline{\theta_m}(p,\mathsf{M})$. The former pertains to reconstruction maps that are not permitted to depend on $\boldsymbol{b}$, whereas the latter pertains to reconstruction maps that can. In particular, we have

$$\theta_m(p) \geq \overline{\theta_m}(p) \geq \overline{\theta_m}(p,\mathsf{M}) \text{ and } \theta_m(p,\mathsf{M}) \geq \overline{\theta_m}(p,\mathsf{M}).$$

Remark 6.1.4 Note that one cannot set $\mathcal{Y} = \mathcal{K}$ in (6.1.3) or (6.1.5), since $\mathcal{K}$ is not a linear subspace of $L^2_\varrho(\mathcal{U};\mathcal{V})$. In the known anisotropy case, a natural choice would be to set $\mathcal{Y} = \mathcal{G}(\boldsymbol{b})$, where $\mathcal{G}(\boldsymbol{b})$ is vector space of $(\boldsymbol{b},1)$-holomorphic functions equipped with the $L^\infty(\mathcal{R}(\boldsymbol{b});\mathcal{V})$-norm. In this case, $\mathcal{H}(\boldsymbol{b})$ is the unit ball of $\mathcal{G}(\boldsymbol{b})$. However, there is no similar such choice in the cases of $\mathcal{H}(p)$ (or $\mathcal{H}(p,\mathsf{M})$). It follows immediately from the definition that $\mathcal{H}(p)$ is a union over $\boldsymbol{b} \in \ell^p(\mathbb{N})$, $\|\boldsymbol{b}\|_p \leq 1$ of unit balls of the subspaces $\mathcal{G}(\boldsymbol{b})$, each equipped with a different norm. Therefore, $\mathcal{H}(p)$ is contained in a union of subspaces

$$\mathcal{H}(p) = \bigcup_{\|\boldsymbol{b}\|_p \leq 1} \mathcal{H}(\boldsymbol{b}) \subset \bigcup_{\|\boldsymbol{b}\|_p \leq 1} \mathcal{G}(\boldsymbol{b}).$$

However, it is possible to show that the right-hand side is not itself a subspace. Therefore, there is no intrinsic choice for $\mathcal{Y}$ in the unknown anisotropy setting.

6.1.5 Problem statement

Keeping these concepts in mind, we now formally define the problem statement for this chapter. Our goal is to study how effective adaptive sampling operators and an arbitrary recovery scheme $\boldsymbol{\mathcal{R}} : \mathcal{V}^m \to L^2_\varrho(\mathcal{U};\mathcal{V})$ can be for both known and unknown anisotropy cases.

First, we examine the known anisotropy case where the arbitrary recovery scheme $\boldsymbol{\mathcal{R}}$ is allowed to depend on $\boldsymbol{b} \in \ell^p(\mathbb{N})$. Specifically, we aim to determine the best performance of a sampling and recovery scheme in this scenario: *Are there lower and upper bounds for the adaptive m-width $\theta_m(\boldsymbol{b})$ in terms of m, and are these bound sharp?*

Additionally, we investigate the performance of a sampling and recovery scheme when considering all possible (unit-norm) $\boldsymbol{b} \in \ell^p(\mathbb{N})$ and $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$. *Are there lower and upper bounds for the adaptive m-widths $\overline{\theta_m}(p,\mathsf{M})$ in terms of m, and are these bounds are sharp.*

Next, we address the unknown anisotropy case where the arbitrary recovery scheme $\boldsymbol{\mathcal{R}}$ is not allowed to depend on $\boldsymbol{b} \in \ell^p(\mathbb{N})$. Here, our objective is to determine the best performance of a sampling and recovery scheme in this context: *Are there lower and upper bounds for the adaptive m-width $\theta_m(p)$ and $\theta_m(p,\mathsf{M})$ in terms of m, and are these bounds are sharp?*

6.2 Contributions

Our main contributions in this work are lower and upper bounds for $\theta_m(\boldsymbol{b})$, $\theta_m(p)$, $\theta_m(p,\mathsf{M})$, $\overline{\theta_m}(p)$ and $\overline{\theta_m}(p,\mathsf{M})$. We first establish lower bounds for $\theta_m(\boldsymbol{b})$ and $\overline{\theta_m}(p)$, which show that no method – i.e., no combination of an arbitrary (adaptive) sampling operator and (potentially nonlinear) reconstruction map – can achieve better rates than $m^{1/2-1/p}$ within the class of $(\boldsymbol{b},1)$-holomorphic functions for $\boldsymbol{b} \in \ell^p(\mathbb{N})$ and $0 < p < 1$. We also show sharp bounds when considering $\overline{\theta_m}(p,\mathsf{M})$. These lower bounds are close to the rates achieved by the algorithms and DL discussed in Chapters 4–5 (see §4.3 and §5.4). Hence, our results indicate that

these algorithms are near optimal (we use the term 'indicate' here since the bounds for these methods are generally nonuniform with respect to the function, whereas our lower bounds are uniform – see §1.4 for further discussion).

In Chapters 4–5 we provided upper bounds in the context of Hilbert-valued function approximation using pointwise samples (see the discussion in §6.3.2 for more details). In this chapter, we show that sharper uniform upper bounds in the Hilbert-valued case can be attained. Specifically, we show that the rate $m^{1/2-1/p}$ can be achieved without log factors in the known anisotropy case, and up to a possible small log factor by using a suitable random sampling operator and reconstruction map.

A key conclusion of this chapter is that in the unknown anisotropy setting the term $\theta_m(p)$ does not decay as $m \to \infty$. In other words, approximation from finite samples is impossible without some inherent ordering of the variables, even if the samples are chosen adaptively.

6.3 Main results

6.3.1 Lower bounds

We first consider lower bounds. Note that these bounds hold for any choice of the normed vector space $\mathcal{Y}$ appearing in (6.1.2) that contains $\mathcal{K} = \mathcal{H}(\boldsymbol{b})$ (known anisotropy case) or $\mathcal{K} = \mathcal{H}(p), \mathcal{H}(p, \mathsf{M})$ (unknown anisotropy case).

To state the corresponding result in the known anisotropy case, we now recall the definition of the ℓ^2-norm *best s-term approximation error* of a sequence $\boldsymbol{c} \in \ell^2(\mathbb{N}; \mathbb{R})$ from Definition 2.4.2. This is given by

$$\sigma_s(\boldsymbol{c})_2 = \min\{\|\boldsymbol{c} - \boldsymbol{z}\|_2 : \boldsymbol{z} \in \ell^2(\mathbb{N}),\ |\mathrm{supp}(\boldsymbol{z})| \leq s\},$$

where $\mathrm{supp}(\boldsymbol{z}) = \{i : z_i \neq 0\}$ for $\boldsymbol{z} = (z_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$.

Theorem 6.3.1 (Known anisotropy; lower bounds and the rate $m^{1/2-1/p}$)**.** *Let $m \geq 1$ and ϱ be a tensor-product probability measure on $\mathcal{U}$. Then the following hold.*

(a) For every $\boldsymbol{b} \in [0,\infty)^{\mathbb{N}}$ with $\boldsymbol{b} \in \ell^1(\mathbb{N})$, the m-width (6.1.3) satisfies

$$\theta_m(\boldsymbol{b}) \geq c \cdot \sigma_m(\boldsymbol{b})_2,$$

where $c > 0$ depends on the measure ϱ and $\|\boldsymbol{b}\|_1$ only.

(b) For every $0 < p < 1$, the m-widths (6.1.4) satisfy

$$\overline{\theta_m}(p) \geq \overline{\theta_m}(p, \mathsf{M}) \geq c \cdot 2^{-1/p} \cdot m^{1/2-1/p}, \tag{6.3.1}$$

where $c > 0$ depends on the measure ϱ only.

(c) Let $(g(n))_{n\in\mathbb{N}}$ be a positive nondecreasing sequence such that

$$\sum_{n\in\mathbb{N}}(ng(n))^{-1}<\infty.$$

Then, for every $0<p<1$, there exists a $\boldsymbol{b}\in\ell^p_{\mathsf{M}}(\mathbb{N})$, $\boldsymbol{b}\in[0,\infty)^{\mathbb{N}}$ such that

$$\overline{\theta_m}(p)\geq\overline{\theta_m}(p,\mathsf{M})\geq\theta_m(\boldsymbol{b})\geq c'\cdot g(2m)^{-1/p}\cdot m^{1/2-1/p}, \tag{6.3.2}$$

where $c'=c\cdot 2^{-1/p}\cdot\left(\sum_{n\in\mathbb{N}}(ng(n))^{-1}\right)^{-1/p}$ and c is the constant in (b).

Part (a) of this result provides a lower bound for the m-width $\theta_m(\boldsymbol{b})$ in terms of the best m-term approximation error $\sigma_m(\boldsymbol{b})_2$. The inequality often referred to as Stechkin's inequality (see, e.g., the historical note [12, Rem. 3.4] on the origins and naming of this inequality and [12, Lem. 3.5]) shows that

$$\sigma_m(\boldsymbol{b})_2\leq\|\boldsymbol{b}\|_p(m+1)^{1/2-1/p},$$

whenever $\boldsymbol{b}\in\ell^p(\mathbb{N})$. Hence, the main contribution of part (b) is to show that the rate $m^{1/2-1/p}$ is, in effect, sharp when considering all possible (unit-norm) $\boldsymbol{b}\in\ell^p(\mathbb{N})$ or $\boldsymbol{b}\in\ell^p_{\mathsf{M}}(\mathbb{N})$. However, it is notable that, based in the proof technique, the $\boldsymbol{b}$ that achieves this bound depends on m and has equal entries, i.e., it corresponds to a class of the functions $\mathcal{H}(\boldsymbol{b})$ that is completely isotropic. In part (c), we prove that a nearly sharp lower bound of the form $g(2m)^{-1/p}\cdot m^{1/2-1/p}$ can be obtained for a fixed $\boldsymbol{b}$, which is independent of m and determined only by a nondecreasing function g. Specifically, the function g can be chosen to grow very slowly, such as $g(n)=\log^2(n+1)$ or even $g(n)=\log(n+1)(\log(\log(n+1)))^2$. For further examples, we refer to [231, Ch. 3].

We now consider lower bounds in the unknown anisotropy setting.

Theorem 6.3.2 (Unknown anisotropy; lower bounds)**.** *Let $m\geq 1$, ϱ be a tensor-product probability measure on $\mathcal{U}$ and $0<p<1$. Then the following hold.*

(a) The m-width $\theta_m(p)$ in (6.1.5) satisfies

$$\theta_m(p)\geq c\cdot 2^{1/2-2/p}, \tag{6.3.3}$$

where $c>0$ depends on the measure ϱ only.

(b) The m-width $\theta_m(p,\mathsf{M})$ in (6.1.5) satisfies

$$\theta_m(p,\mathsf{M})\geq\overline{\theta_m}(p,\mathsf{M})\geq c\cdot 2^{-1/p}\cdot m^{1/2-1/p}, \tag{6.3.4}$$

where $c>0$ depends on the measure ϱ only.

Part (a) of this theorem, more precisely the factor $2^{1/2-2/p}$, shows that approximation from finite samples is, in fact, impossible in the space $\mathcal{H}(p)$, since the m-width does not decay as $m \to \infty$. Note that this bound holds for any sampling operator, i.e., not just the random pointwise samples considered in Chapters 3–5. This is perhaps unsurprising. Functions in $\mathcal{H}(p)$ are anisotropic, but their (infinitely-many) variables can be ordered in terms of importance in arbitrary and infinitely-many different ways. It seems implausible that one could approximate such functions from a finite set of data. This result confirms this intuition.

However, part (b) of this theorem reveals that the situation changes completely when we restrict to the monotone space $\mathcal{H}(p, \mathsf{M})$. Here the lower bound for the m-width is once more $m^{1/2-1/p}$. Note that this change holds for any sample strategy. Thus, Chapters 3–5 suggest that pointwise samples are nearly optimal when we consider the aforementioned monotone space.

6.3.2 Upper bounds

We now present a series of upper bounds for the various m-widths. In addition to the discussion in §6.2, we recall that in Chapters 4–5 we showed nonuniform upper bounds (see Remark 3.3.4 and the discussion after Theorem 5.3.4) of the form $(m/\text{polylog}(m))^{1/2-1/p}$ using pointwise samples. These polylogarithmic factors were of the order of $\mathcal{O}(\log^4(m))$ in the case of unknown anisotropy (see Theorem 4.3.3 and Theorem 5.3.2) and of the order of $\mathcal{O}(\log(m))$ in the known anisotropy case (see Theorem 5.3.4). Below, we show that the rate $m^{1/2-1/p}$ can be achieved without log factors in the known anisotropy case, and up to a possible small log factor by using a suitable random sampling operator and reconstruction map. Our main upper bounds below are uniform since they consider a worst-case error over all f.

In the following results, we make the additional (mild) assumption that the normed vector space $\mathcal{Y}$ appearing in (6.1.2) is compactly contained in $L^2_\varrho(\mathcal{U}; \mathcal{V})$, i.e., $\mathcal{Y} \hookrightarrow L^2_\varrho(\mathcal{U}; \mathcal{V})$.

Theorem 6.3.3 (Known anisotropy; upper bounds)**.** *Let $\mathcal{V}$ be a Hilbert space, $m \geq 1$, ϱ be the uniform probability measure on $\mathcal{U}$ and $0 < p < 1$. Then the following hold.*

(a) The m-width (6.1.3) satisfies

$$\theta_m(\boldsymbol{b}) \leq c \cdot m^{1/2-1/p}, \quad \forall \boldsymbol{b} \in \ell^p(\mathbb{N}),\ \boldsymbol{b} \in [0,\infty)^{\mathbb{N}}, \tag{6.3.5}$$

where $c > 0$ depends on $\boldsymbol{b}$ and p only. Moreover, this bound is attained by a bounded linear (nonadaptive) sampling map $\boldsymbol{\mathcal{S}}$ and a bounded linear reconstruction operator $\boldsymbol{\mathcal{R}}$.

(b) In addition, for any $q \in (p,1)$ the m-width (6.1.4) satisfies

$$\overline{\theta_m}(p,\mathsf{M}) \leq c \cdot m^{1/2-1/q}, \tag{6.3.6}$$

where $c > 0$ depends on p and q only.

Part (a) shows that in the case of known anisotropy for a fixed $\boldsymbol{b}$ we can achieve a rate $m^{1/2-1/p}$ as an upper bound with a constant depending on $\boldsymbol{b} \in \ell^p(\mathbb{N})$. In contrast, the proof of part (b) provides a uniform bound over all $\boldsymbol{b}$ belonging to the unit ball of the monotone ℓ^p space $\ell^p_{\mathsf{M}}(\mathbb{N})$. Notice also that part (b) only considers the monotone space $\ell^p_{\mathsf{M}}(\mathbb{N})$. We suspect that achieving a uniform bound for all $\boldsymbol{b}$ belonging to the unit ball of the standard ℓ^p space $\ell^p(\mathbb{N})$ may be possible. However, this is a problem for future research.

The key difference between parts (a) and (b) is the algebraic rate, which can be arbitrarily close to $m^{1/2-1/p}$, but not equal to it. Currently, we do not have an explicit expression for how the constant c in part (b) depends on p and q and, in particular, how it behaves as $q \to p^+$, besides the knowledge that it must blow up. In particular, the question of whether (6.3.6) may in fact hold with $q = p$ is an open problem and, if true, its proof would require a different technique. See Remark 6.6.4 for some additional discussion.

Both Theorem 6.3.3 and the next result are proved using Legendre polynomials. In Theorem 6.3.3, the sampling operator $\boldsymbol{\mathcal{S}}$ computes m Legendre coefficients of f from a suitable index set (depending on $\boldsymbol{b}$ only). However, unlike in Chapters 4–5, the sampling operator $\boldsymbol{S}$ in our upper bounds does not compute pointwise samples of the target function f in either result. Consequently, while our theorems in this chapter provide upper bounds for the m-widths, they do not address the practical scenario of pointwise samples.

In the unknown anisotropy setting, we resort to a different approach based on compressed sensing (see §3.6). For a more detailed discussion see §6.4. In summary the idea is to choose the sampling operator $\boldsymbol{\mathcal{S}}$ so that the recovery problem for the coefficients reduces to the problem of recovering a sparse vector from random Gaussian measurements. In order to use this approach, we need to carefully restrict the infinite vector of polynomial coefficients to a finite vector of length N. As in Chapters 3–5 we use the hyperbolic cross index set in §2.4.6. Specifically, we do this so that no large coefficients are excluded from the resulting finite vector. This is where we use the additional assumption $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ and the properties of anchored sets (see §2.5.2). See the proof in §6.6.4 which follows a similar idea to the previous chapters. For notational convenience, in the following result we introduce a normal random vector $\boldsymbol{r} \sim \mathcal{N}(0, I_\ell)$, $\ell = mN$, which contains the entries of the aforementioned Gaussian random matrix.

Theorem 6.3.4 (Unknown anisotropy; upper bounds)**.** *Let $\mathcal{V}$ be a Hilbert space, $m \geq 3$ and ϱ be the uniform probability measure on $\mathcal{U} = [-1,1]^{\mathbb{N}}$. Then there exists an $\ell = \ell(m)$, a (nonlinear) reconstruction map $\boldsymbol{\mathcal{R}}_{\boldsymbol{r}} : \mathcal{V}^m \to L^2_\varrho(\mathcal{U};\mathcal{V})$ and a (nonadaptive) bounded linear*

sampling operator $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}} : \mathcal{Y} \to \mathcal{V}^m$ *depending on a random vector* $\boldsymbol{r} \sim \mathcal{N}(0, I_\ell)$*, where* I_ℓ *is the* $\ell \times \ell$ *identity matrix, such that the following holds.*

(a) We have

$$\mathbb{E}_{\boldsymbol{r}\sim\mathcal{N}(0,I_\ell)} \sup_{f\in\mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c \cdot \left(\frac{m}{\log^2(m)}\right)^{1/2-1/p},$$

for all $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$*,* $\boldsymbol{b} \in [0,\infty)^{\mathbb{N}}$*, and* $0 < p < 1$*, where* $c > 0$ *depends on* $\boldsymbol{b}$ *and* p *only.*

(b) For $0 < p < q < 1$*, the* m*-width* (6.1.5) *satisfies*

$$\theta_m(p,\mathsf{M}) \leq \mathbb{E}_{\boldsymbol{r}\sim\mathcal{N}(0,I_\ell)} \sup_{f\in\mathcal{H}(p,\mathsf{M})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq c \cdot m^{1/2-1/q},$$

where $c > 0$ *depends on* p *and* q *only.*

Part (b) of this theorem shows that the lower bounds in Theorem 6.3.2(b) can be nearly achieved by a (nonadaptive) random sampling operator and reconstruction map. Here, "nearly" means with a rate that can be arbitrarily close to $1/2 - 1/p$, but not equal to it. As in the previous result, we also have no explicit expression for the constant c in part (b). This is for much the same reasons as those discussed in Remark 6.6.4. By contrast, Theorem 6.3.4(a) shows that the algebraic rate $1/2 - 1/p$ can be achieved, but with a constant depending on $\boldsymbol{b}$. Crucially, however, the sampling operator and reconstruction map in Theorem 6.3.4(a) are both independent of $\boldsymbol{b}$ and p and therefore part (a) is also a result about the unknown anisotropy setting, even though the supremum is taken over $\mathcal{H}(\boldsymbol{b})$.

Theorems 6.3.3 and 6.3.4 consider only the uniform probability measure on $\mathcal{U}$. We anticipate they hold for more general tensor-product probability measures, such as Jacobi measures. See the preprint [17].

Remark 6.3.5 We also note in passing that Theorems 6.3.3–6.3.4 assume that $\mathcal{V}$ is a Hilbert space. In the case of Theorem 6.3.3, the reason to use a Hilbert spaces comes from a technical step in the proof that uses Parseval's identity. We encounter the same issue in Chapter 5 (see the discussion after Theorem 5.3.1). Whether these rates hold in the Banach-valued case remains an open problem.

6.4 Discussion

In addition to the discussion in §6.3, we now comment on two important aspects of the main results. These are crucial to understand their proofs.

Lower bounds

The proofs of our lower bounds rely on the reduction of the continuous problem to a discrete one. First, we prove a holomorphy result of order-one polynomials. This allows us to lower bound the m-width $\theta_m(\boldsymbol{b})$ in terms of a certain discrete problem. Having done this, we then use certain results on the so-called *Gelfand* and *Kolmogorov* width [111, 220] of certain weighted ℓ^p balls to get the desired bounds.

Theorem 6.3.3: Known anisotropy; upper bounds

As mentioned after Theorem 6.3.3, the sampling operator $\boldsymbol{\mathcal{S}}$ computes m Legendre coefficients of f from a suitable index set (depending on $\boldsymbol{b}$ only). The reconstruction map then simply forms the corresponding Legendre polynomial expansion. In particular, the reconstruction map is linear. This approach is possible in the setting of known anisotropy, since we can choose the index set in terms of $\boldsymbol{b}$.

Theorem 6.3.4: Unknown anisotropy; upper bounds

The main difficulty in this theorem comes from choosing a suitable sampling operator and reconstruction map that do not depend on an specific choice of index set. As we did in the previous Chapters 3–5, we see the polynomial coefficients as an (approximately) sparse vector, whose significant entries are unknown (since $\boldsymbol{b}$ is unknown). Then we choose the sampling operator $\boldsymbol{\mathcal{S}}$ so that the recovery problem for the coefficients reduces to the problem of recovering a sparse vector from random Gaussian measurements. Specifically, $\boldsymbol{\mathcal{S}}$ computes m random weighted sums of the Legendre coefficients of f where the weights are i.i.d. normal random variables (a similar idea was also used in a different context in [11]). Finally, we formulate the (nonlinear) reconstruction mapin terms of a convex ℓ^1-minimization problem.

6.5 Proofs setup

Before diving into the details of the proofs, we now provide further setup and introduce important notations.

6.5.1 Notation

Let $0 < p \leq \infty$ and $\boldsymbol{w} = (w_i)_{i\in\mathbb{N}} > \boldsymbol{0}$ be a sequence of positive weights. We define the weighted ℓ^p-space $\ell^p(\boldsymbol{w})$ to be the set of all sequences $\boldsymbol{z} = (z_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ for which the weighted (quasi-) norm $\|\boldsymbol{z}\|_{p,\boldsymbol{w}}$ is finite. That is

$$\|\boldsymbol{z}\|_{p,\boldsymbol{w}} := \begin{cases} \left(\sum_{i\in\mathbb{N}} w_i^{-p}|z_i|^p\right)^{1/p} < \infty, & p < \infty, \\ \sup_{i\in\mathbb{N}} \left\{w_i^{-1}|z_i|\right\} < \infty, & p = \infty. \end{cases}$$

When $\boldsymbol{w} = \mathbf{1}$ is the vector of ones, we just write $(\ell^p, \|\cdot\|_p)$. For $N \in \mathbb{N}$ and $\boldsymbol{w} = (w_i)_{i\in[N]}$, we use $\ell^p_N(\boldsymbol{w})$ to denote the finite dimensional space $(\mathbb{R}^N, \|\cdot\|_{p,\boldsymbol{w}})$ of vectors of length N. When $\boldsymbol{w} = \mathbf{1}$, we just write $(\ell^p_N, \|\cdot\|_p)$.

Next, we write

$$B^p_N(\boldsymbol{w}) = \begin{cases} \left\{\boldsymbol{x} = (x_i)_{i=1}^N \in \mathbb{R}^N : \left(\sum_{i=1}^N w_i{}^{-p}|x_i|^p\right)^{1/p} \leq 1\right\} & p < \infty, \\ \left\{\boldsymbol{x} = (x_i)_{i=1}^N \in \mathbb{R}^N : \max_{i=1,\dots,N}\left\{w_i^{-1}|x_i|\right\} \leq 1\right\} & p = \infty, \end{cases} \tag{6.5.1}$$

for the weighted ℓ^p-norm (quasi-norm) unit ball when $p \geq 1$ ($0 < p < 1$). When $\boldsymbol{w} = \mathbf{1}$, we simply write B^p_N.

6.5.2 Widths and standard results on widths

We commence with a brief overview of Gelfand widths and their properties, since these will be crucial in the proofs of the lower bounds. See [220] or [112, Ch. 10] for more details. Let $\mathcal{K}$ be a subset of a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$. Then its *Gelfand m-width* is

$$d^m(\mathcal{K}, \mathcal{X}) = \inf\left\{\sup_{x\in\mathcal{K}\cap L^m} \|x\|_{\mathcal{X}},\ L^m \text{ a subspace of } \mathcal{X} \text{ with } \operatorname{codim}(L^m) \leq m\right\}. \tag{6.5.2}$$

An equivalent representation is

$$d^m(\mathcal{K}, \mathcal{X}) = \inf\left\{\sup_{x\in\mathcal{K}\cap\operatorname{Ker}(A)} \|x\|_{\mathcal{X}},\ A : \mathcal{X} \to \mathbb{R}^m \text{ linear}\right\}.$$

The Gelfand width is related to the following quantity:

$$E^m_{\mathsf{ada}}(\mathcal{K}, \mathcal{X}) = \inf\left\{\sup_{x\in\mathcal{K}} \|x - \boldsymbol{\Delta}(\boldsymbol{\Gamma}(x))\|_{\mathcal{X}},\ \boldsymbol{\Gamma} : \mathcal{X} \to \mathbb{R}^m \text{ adaptive},\ \boldsymbol{\Delta} : \mathbb{R}^m \to \mathcal{X}\right\}, \tag{6.5.3}$$

where $\boldsymbol{\Gamma}$ is an adaptive sampling operator as in Definition 6.1.1. This is referred to as the *adaptive compressive m-width* of a subset $\mathcal{K}$ of a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ in [112, Ch. 10]. Note that, if $\mathcal{X} = L^2_\varrho(\mathcal{U}; \mathbb{R})$ and $\mathcal{V} = \mathbb{R}$, then $E^m_{\mathsf{ada}}(\mathcal{K}, \mathcal{X})$ coincides with $\Theta_m(\mathcal{K}; \mathcal{X}, \mathcal{X})$ in (6.1.2).

The following result is standard and can be found in [112, Thm. 10.4].

Theorem 6.5.1. *Let $\mathcal{K} \subseteq \mathcal{X}$, where $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ is a normed space. If $-\mathcal{K} = \mathcal{K}$ then*

$$d^m(\mathcal{K}, \mathcal{X}) \leq E^m_{\mathsf{ada}}(\mathcal{K}, \mathcal{X}).$$

Finally, we also define the *Kolmogorov m-width* of a subset $\mathcal{K}$ of a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ as

$$d_m(\mathcal{K}, \mathcal{X}) = \inf \left\{ \sup_{x \in \mathcal{K}} \inf_{z \in \mathcal{X}_m} \|x - z\|_{\mathcal{X}}, \ \mathcal{X}_m \text{ a subspace of } \mathcal{X} \text{ with } \dim(\mathcal{X}_m) \leq m \right\}.$$

6.6 Proof of main results: Theorems 6.3.1–6.3.4

6.6.1 Proof of Theorem 6.3.1: known anisotropy, lower bound

As mentioned in §6.4, we first prove a holomorphy result of order-one polynomials that allows us to lower bound the m-width $\theta_m(\boldsymbol{b})$ in terms of a certain discrete problem.

We start by making the following observation. Let ρ be a probability measure as in §2.2 and recall that ϱ is the tensor-product probability measure defined in (2.2.1) . Since ρ is a nonnegative Borel probability measure on the interval $[-1, 1]$ its moments exist [264, §6.1]. Now let

$$\tau = \int_{-1}^{1} y \, \mathrm{d}\rho(y), \quad \text{and} \quad \sigma = \sqrt{\int_{-1}^{1} (y - \tau)^2 \, \mathrm{d}\rho(y)}, \tag{6.6.1}$$

be the first and second moment of ρ and notice that $\tau, \sigma < \infty$. Then the functions

$$\psi_i(\boldsymbol{y}) = \frac{y_i - \tau}{\sigma}, \quad \boldsymbol{y} = (y_j)_{j \in \mathbb{N}} \in \mathcal{U}, \ i \in \mathbb{N},$$

form an orthonormal set $\{\psi_i\}_{i \in \mathbb{N}} \subset L^2_{\varrho}(\mathcal{U})$ (but not a basis).

Lemma 6.6.1 (Holomorphy of order one polynomials)**.** *Let ϱ be a tensor-product probability measure as in (2.2.4), $p \in (0, 1]$, $\boldsymbol{b} \in [0, \infty)^{\mathbb{N}}$ with $\boldsymbol{b} \in \ell^p(\mathbb{N})$, $v \in \mathcal{V} \setminus \{0\}$ and consider a sequence $\boldsymbol{c} = (c_i)_{i \in \mathbb{N}} \subset \mathbb{R}^{\mathbb{N}}$ with $|c_i| \leq b_i$ for all $i \in \mathbb{N}$. Define the function*

$$f = \sum_{i=1}^{\infty} c_i v \psi_i. \tag{6.6.2}$$

Then f is $(\boldsymbol{b}, 1)$-holomorphic with

$$\|f\|_{L^\infty(\mathcal{R}(\boldsymbol{b}); \mathcal{V})} \leq \frac{\|v\|_{\mathcal{V}}}{\sigma} \left(1 + (|\tau| + 1)\|\boldsymbol{c}\|_p\right).$$

Proof. Notice that $(|c_i|)_{i \in \mathbb{N}} \in \ell^1(\mathbb{N})$ and that f is holomorphic at any $\boldsymbol{y} \in \mathcal{U}$ for which the series $\sum_{i=1}^{\infty} c_i v \psi_i(\boldsymbol{y})$ converges absolutely. Now suppose that $\boldsymbol{y} \in \mathcal{E}_{\boldsymbol{\rho}}$, where $\boldsymbol{\rho}$ satisfies the

condition in (2.3.2) with $\varepsilon = 1$. Then $|y_i| \leq (\rho_i + \rho_i^{-1})/2$ and therefore

$$\begin{aligned}\left\|\sum_{i=1}^{\infty} c_i v \psi_i(\boldsymbol{y})\right\|_{\mathcal{V}} &\leq \sum_{i=1}^{\infty} \|c_i v\|_{\mathcal{V}} \left(\frac{|y_i| + |\tau|}{\sigma}\right) \\ &\leq \frac{\|v\|_{\mathcal{V}}}{\sigma} \left(\sum_{i=1}^{\infty} |c_i| \left((\rho_i + \rho_i^{-1})/2 - 1\right) + \|\boldsymbol{c}\|_1 + |\tau| \|\boldsymbol{c}\|_1\right) \\ &\leq \frac{\|v\|_{\mathcal{V}}}{\sigma} \left(1 + (|\tau| + 1)\|\boldsymbol{c}\|_p\right),\end{aligned}$$

as required. □

In the next result, we relate the m-width $\theta_m(\boldsymbol{b})$ in (6.1.3) to the Gelfand m-width of a certain finite-dimensional unit ball.

Lemma 6.6.2 (Reduction to a discrete problem; known anisotropy case)**.** *Let ϱ be a tensor-product probability measure as in (2.2.4), $\boldsymbol{b} \in [0,\infty)^{\mathbb{N}}$ with $\boldsymbol{b} \in \ell^1(\mathbb{N})$. Let $N \in \mathbb{N}$ and $I \subset \mathbb{N}$ be an index set with $|I| = N$. Then the m-width (6.1.3) satisfies*

$$\theta_m(\boldsymbol{b}) \geq C(\boldsymbol{b}, \tau, \sigma) \cdot d^m(B_N^{\infty}(\boldsymbol{b}_I), \ell_N^2), \tag{6.6.3}$$

where $B_N^{\infty}(\boldsymbol{b}_I)$ is as in (6.5.1) and

$$C(\boldsymbol{b}, \tau, \sigma) = \frac{\sigma}{1 + (1 + |\tau|)\|\boldsymbol{b}\|_1}. \tag{6.6.4}$$

Observe that the bound (6.6.3) holds trivially when $N \leq m$, since $d^m(B_N^{\infty}(\boldsymbol{b}_I), \ell_N^2) = 0$ in this case (see [112, §10.1]).

Proof. Let $\boldsymbol{\mathcal{S}}$ be as in Definition 6.1.2. Then there are $v \in \mathcal{V}\backslash\{0\}$, $w \in \mathcal{V}\backslash\{0\}$ and a normed vector space $\widetilde{\mathcal{Y}} \subseteq L_{\varrho}^2(\mathcal{U})$ such that $\boldsymbol{\mathcal{S}}(vg) = w\widetilde{\boldsymbol{\mathcal{S}}}(g)$ whenever $vg \in \mathcal{Y}$ and $g \in L_{\varrho}^2(\mathcal{U})$, where $\widetilde{\boldsymbol{\mathcal{S}}} : \widetilde{\mathcal{Y}} \to \mathbb{R}^m$ is as in Definition 6.1.1. Now let $\boldsymbol{c} = (c_i)_{i\in I} \subset \mathbb{R}^{\mathbb{N}}$ be any sequence supported in I with $|c_i| \leq b_i$, $\forall i \in I$, and define $f : \mathcal{U} \to \mathcal{V}$ by

$$f = cv \sum_{i\in I} c_i \psi_i, \qquad c = \frac{\sigma}{\|v\|_{\mathcal{V}}(1 + (|\tau| + 1)\|\boldsymbol{b}\|_1)}. \tag{6.6.5}$$

Since $\|\boldsymbol{c}\|_1 \leq \|\boldsymbol{b}\|_1$, Lemma 6.6.1 implies that $f \in \mathcal{H}(\boldsymbol{b}) \subseteq \mathcal{Y}$. By definition, we have

$$\boldsymbol{\mathcal{S}}(f) = w\widetilde{\boldsymbol{\mathcal{S}}}(g), \qquad g = c\sum_{i\in I} c_i \psi_i \in \widetilde{\mathcal{Y}}.$$

Now let $\boldsymbol{\Gamma} : \mathbb{R}^N \to \mathbb{R}^m$ be the scalar-valued adaptive sampling operator defined by

$$\boldsymbol{\Gamma}(\boldsymbol{d}) = \widetilde{\boldsymbol{\mathcal{S}}}\left(c \cdot \sum_{i\in I} d_i \psi_i\right), \quad \boldsymbol{d} = (d_i)_{i\in I} \in \mathbb{R}^N.$$

Here, for convenience, we index vectors in $\mathbb{R}^N$ using the index set I. We need to show that this operator is well defined. Since $\widetilde{\boldsymbol{\mathcal{S}}}$ has domain $\widetilde{\mathcal{Y}}$, this is equivalent to showing that $\sum_{i\in I} d_i\psi_i \in \widetilde{\mathcal{Y}}$ for all $\boldsymbol{d} = (d_i)_{i\in I} \in \mathbb{R}^N$. To see this, recall that $\mathcal{H}(\boldsymbol{b}) \subset \mathcal{Y}$ and therefore

$$\mathcal{H}_0(\boldsymbol{b}) := \{f : \mathcal{U} \to \mathcal{V}, (\boldsymbol{b}, 1)\text{-holomorphic}\} \subseteq \mathcal{Y},$$

since $\mathcal{Y}$ is a vector space. Now, for any $\boldsymbol{d} = (d_i)_{i\in I} \in \mathbb{R}^N$, the function $\sum_{i\in I} d_i\psi_i$ is entire, therefore

$$v \cdot \sum_{i\in I} d_i\psi_i \in \mathcal{H}_0(\boldsymbol{b}).$$

Hence, $\sum_{i\in I} d_i\psi_i \in \widetilde{\mathcal{Y}}$ due to Definition 6.1.2, and therefore $\boldsymbol{\Gamma}$ is well defined.

With this in hand, recall that $\boldsymbol{c} \in \mathbb{R}^{\mathbb{N}}$ is zero outside of the index set I. Hence we may consider it as an element of $\mathbb{R}^N$ indexed over I. Using this and the definition of $\boldsymbol{\Gamma}$, we have $\boldsymbol{\mathcal{S}}(f) = w\widetilde{\boldsymbol{\mathcal{S}}}(g) = w\boldsymbol{\Gamma}(\boldsymbol{c}) = (w(\boldsymbol{\Gamma}(\boldsymbol{c}))_i)_{i=1}^m \in \mathcal{V}^m$.

Now let $\boldsymbol{\mathcal{R}} : \mathcal{V}^m \to L^2_\varrho(\mathcal{U};\mathcal{V})$ be an arbitrary reconstruction map and let $\widetilde{\boldsymbol{\mathcal{R}}} : \mathbb{R}^m \to L^2_\varrho(\mathcal{U};\mathcal{V})$ be defined by

$$\widetilde{\boldsymbol{\mathcal{R}}}(\boldsymbol{z}) = \boldsymbol{\mathcal{R}}(w \cdot \boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathbb{R}^m.$$

Observe that

$$\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{S}}(f)) = \boldsymbol{\mathcal{R}}(w \cdot \boldsymbol{\Gamma}(\boldsymbol{c})) = \widetilde{\boldsymbol{\mathcal{R}}}(\boldsymbol{\Gamma}(\boldsymbol{c})). \tag{6.6.6}$$

Let $\mathcal{V}^*$ be the dual space of $\mathcal{V}$. From the Hahn–Banach theorem (see [232, Thm. 3.3]) there exists a linear bounded functional $\phi_v^* \in \mathcal{V}^*$ with unit norm such that $\phi_v^*(v) = \|v\|_{\mathcal{V}}$. Using this fact and the definition of a norm in Banach spaces, for every $\boldsymbol{y} \in \mathcal{U}$, we get

$$\begin{aligned}
\|f(\boldsymbol{y}) - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)(\boldsymbol{y})\|_{\mathcal{V}} &= \sup_{\phi^* \in \mathcal{V}^*, \|\phi^*\|_{\mathcal{V}^*}=1} |\langle \phi^*, f(\boldsymbol{y}) - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)(\boldsymbol{y})\rangle_{\mathcal{V}^*\times\mathcal{V}}| \\
&\geq |\langle \phi_v^*, f(\boldsymbol{y}) - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)(\boldsymbol{y})\rangle_{\mathcal{V}^*\times\mathcal{V}}| \\
&= \left| c' \sum_{i\in I} c_i\psi_i(\boldsymbol{y}) - \langle \phi_v^*, \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)(\boldsymbol{y})\rangle_{\mathcal{V}^*\times\mathcal{V}} \right|,
\end{aligned}$$

where $c' = c\|v\|_{\mathcal{V}} = \sigma/(1 + (|\tau| + 1)\|\boldsymbol{b}\|_1)$. Then, squaring and integrating over $\mathcal{U}$, we can use Bessel's inequality on the rightmost term to obtain

$$\begin{aligned}
\|f - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)\|^2_{L^2_\varrho(\mathcal{U};\mathcal{V})} &\geq \|c' \sum_{i\in I} c_i\psi_i - \phi_v^* \circ \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)\|^2_{L^2_\varrho(\mathcal{U})} \\
&\geq \sum_{j\in I} |\langle \psi_j, c' \sum_{i\in I} c_i\psi_i - \phi_v^* \circ \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)\rangle_{L^2_\varrho(\mathcal{U})}|^2.
\end{aligned}$$

Combining this with (6.6.6) we obtain

$$\begin{aligned}\|f-\boldsymbol{\mathcal{R}}\circ\boldsymbol{\mathcal{S}}(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}^2 &\geq (c')^2\sum_{j\in I}|c_j-\langle\psi_j,\phi_v^*\circ\boldsymbol{\mathcal{R}}\circ\boldsymbol{\mathcal{S}}(f)\rangle_{L^2_\varrho(\mathcal{U})}/c'|^2\\ &= (c')^2\sum_{j\in I}|c_j-\langle\psi_j,\phi_v^*\circ\widetilde{\boldsymbol{\mathcal{R}}}(\boldsymbol{\Gamma}(\boldsymbol{c}))\rangle_{L^2_\varrho(\mathcal{U})}/c'|^2.\end{aligned}$$

Notice from the last term that every map $\widetilde{\boldsymbol{\mathcal{R}}}:\mathbb{R}^m\to L^2_\varrho(\mathcal{U};\mathcal{V})$ gives rise to a mapping $\boldsymbol{z}\mapsto\boldsymbol{\Delta}(\boldsymbol{z})$ with $\boldsymbol{\Delta}:\mathbb{R}^m\to\mathbb{R}^N$ via

$$\boldsymbol{\Delta}(\boldsymbol{z})=\left(\langle\psi_i,\phi_v^*\circ\widetilde{\boldsymbol{\mathcal{R}}}(\boldsymbol{z})\rangle_{L^2_\varrho(\mathcal{U})}/c'\right)_{i\in I}=\left(\int_{\mathcal{U}}\psi_i(\boldsymbol{y})\cdot\phi_v^*\left(\widetilde{\boldsymbol{\mathcal{R}}}(\boldsymbol{z})(\boldsymbol{y})\right)\mathrm{d}\varrho(\boldsymbol{y})/c'\right)_{i\in I}. \tag{6.6.7}$$

Hence we obtain

$$\|f-\boldsymbol{\mathcal{R}}\circ\boldsymbol{\mathcal{S}}(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}^2\geq(c')^2\sum_{i\in I}|c_i-(\boldsymbol{\Delta}(\boldsymbol{\Gamma}(\boldsymbol{c})))_i|^2=(c')^2\|\boldsymbol{c}-\boldsymbol{\Delta}(\boldsymbol{\Gamma}(\boldsymbol{c}))\|_2^2.$$

Thus, we have shown that for any pair $(\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{R}})$ and any f of the form (6.6.5), the error $\|f-\boldsymbol{\mathcal{R}}\circ\boldsymbol{\mathcal{S}}(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}$ can be bounded below by a constant times the error $\|\boldsymbol{c}-\boldsymbol{\Delta}(\boldsymbol{\Gamma}(\boldsymbol{c}))\|_2$ for some pair $(\boldsymbol{\Gamma},\boldsymbol{\Delta})$. Using this, we deduce that

$$\begin{aligned}\theta_m(\boldsymbol{b}) &= \inf\left\{\sup_{f\in\mathcal{H}(\boldsymbol{b})}\|f-\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{S}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}:\boldsymbol{\mathcal{S}}:\mathcal{Y}\to\mathcal{V}^m\text{ adaptive},\ \boldsymbol{\mathcal{R}}:\mathcal{V}^m\to L^2_\varrho(\mathcal{U};\mathcal{V})\right\}\\ &\geq c'\inf\left\{\sup_{\substack{\boldsymbol{c}\in\mathbb{R}^N,\boldsymbol{c}\neq\boldsymbol{0}\\ |c_i|\leq b_i,\forall i\in I}}\|\boldsymbol{c}-\boldsymbol{\Delta}(\boldsymbol{\Gamma}(\boldsymbol{c}))\|_2:\boldsymbol{\Gamma}:\mathbb{R}^N\to\mathbb{R}^m\text{ adaptive},\ \boldsymbol{\Delta}:\mathbb{R}^m\to\mathbb{R}^N\right\}\\ &\geq c'E^m_{\mathsf{ada}}(B^\infty_N(\boldsymbol{b}_I),\ell^2_N),\end{aligned}$$

where in the final inequality we recall (6.5.3). The result now follows from Theorem 6.5.1. □

Next, we give a lower bound for the right-hand side of (6.6.3). For this, we make use of Theorem C.0.2, which is due to Stesin [241].

Proof of Theorem 6.3.1. We first prove part (a). Let $N\in\mathbb{N}$ with $N>m$ and $I\subset\mathbb{N}$, $|I|=N$ be the index set corresponding to the largest N entries of $\boldsymbol{b}$. First, using the duality result Theorem C.0.3 and then Lemma C.0.4 we obtain

$$d^m(B^\infty_N(\boldsymbol{b}_I),\ell^2_N)=d_m(B^2_N,\ell^1_N(1/\boldsymbol{b}_I))=d_m(B^2_N(\boldsymbol{b}_I),\ell^1_N). \tag{6.6.8}$$

Here $1/\boldsymbol{b}_I$ is the vector with entries $1/\boldsymbol{b}_I = (1/b_i)_{i\in I}$. Next, applying Theorem C.0.2 with $p = 2$ and $q = 1$ we see that the right-hand side of the previous equation satisfies

$$d_m(B_N^2(\boldsymbol{b}_I), \ell_N^1) = \left(\max_{\substack{i_1,\dots,i_{N-m}\in I \\ i_k\neq i_j}} \left(\sum_{j=1}^{N-m} (b_{i_j})^2 \right)^{-1/2} \right)^{-1}. \tag{6.6.9}$$

Now, let $\pi : \mathbb{N} \mapsto \mathbb{N}$ be a bijection whose entries are a nonincreasing rearrangement of the vector $\boldsymbol{b}$ in nonincreasing order. That is $\boldsymbol{b}_\pi = (b_{\pi(j)})_{j\in\mathbb{N}}$ is such that

$$b_{\pi(1)} \geq b_{\pi(2)} \geq \cdots b_{\pi(N)} \geq \ldots \geq 0.$$

Observe that $\boldsymbol{b}_I$ has a one-to-one relation with the first N terms of $\boldsymbol{b}_\pi$. Hence, the maximum in (6.6.9) is achieved by the last $N - m$ terms of $(\boldsymbol{b}_{\pi(i)})_{i=1}^N$. Thus,

$$d_m(B_N^2(\boldsymbol{b}_I), \ell_N^1) = \min_{\substack{i_1,\dots,i_{N-m}\in I \\ i_k\neq i_j}} \left(\sum_{j=1}^{N-m} (b_{i_j})^2 \right)^{1/2} = \left(\sum_{j=m+1}^{N} b_{\pi(j)}^2 \right)^{1/2}.$$

Moreover, we know that $\sigma_m(\boldsymbol{b})_2^2 = \sum_{j=m+1}^{\infty} b_{\pi(j)}^2$, which implies that

$$\sigma_m(\boldsymbol{b})_2^2 = \sum_{j=m+1}^{N} b_{\pi(j)}^2 + \sum_{j=N+1}^{\infty} b_{\pi(j)}^2 = (d_m(B_N^2(\boldsymbol{b}_I), \ell_N^1))^2 + \sum_{j=N+1}^{\infty} b_{\pi(j)}^2.$$

Then, from (6.6.8) and Lemma 6.6.2 we obtain

$$\sigma_m(\boldsymbol{b})_2^2 \leq C(\boldsymbol{b}, \tau, \sigma)^{-2} \cdot \theta_m^2(\boldsymbol{b}) + \sum_{j=N+1}^{\infty} b_{\pi(j)}^2,$$

where $C(\boldsymbol{b}, \tau, \sigma)$ is as in (6.6.4). Taking limit when $N \to \infty$ we obtain the result.

Next we prove part (b). Consider the sequence $\boldsymbol{b} = (b_i)_{i=1}^\infty$ with

$$b_i = (2m)^{-1/p}, \quad i = 1, \dots, 2m, \quad b_i = 0,\ i > 2m. \tag{6.6.10}$$

Observe that $\boldsymbol{b} \in \ell_{\mathsf{M}}^p(\mathbb{N})$ and that $\|\boldsymbol{b}\|_{p,\mathsf{M}} = \|\boldsymbol{b}\|_p = 1$ by construction. Also, note that

$$\sigma_m(\boldsymbol{b})_2 = \sqrt{\sum_{i=m+1}^{2m} (2m)^{-2/p}} = 2^{-1/p} m^{1/2-1/p}.$$

Thus using this and part (a) we get $\overline{\theta_m}(p, \mathsf{M}) \geq C(\boldsymbol{b}, \tau, \sigma) \cdot \sigma_m(\boldsymbol{b})_2 \geq C(\tau, \sigma) \cdot 2^{-1/p} m^{1/2-1/p}$, where

$$C(\tau, \sigma) = \frac{\sigma}{1 + (1 + |\tau|)} = \frac{\sigma}{1 + (1 + |\tau|)\|\boldsymbol{b}\|_p} \leq C(\boldsymbol{b}, \tau, \sigma). \tag{6.6.11}$$

Finally, we prove part (c). Let $c_{p,g} = \left(\sum_{n\in\mathbb{N}}(ng(n))^{-1}\right)^{-1/p}$ and consider the sequence $\boldsymbol{b} = (b_i)_{i=1}^{\infty}$ defined by $b_i = c_{p,g}(ig(i))^{-1/p}$. Observe that $\|\boldsymbol{b}\|_p = 1$ by construction. Recall that $\ell^p_{\mathsf{M}}(\mathbb{N})$ is the space of sequences whose minimal monotone majorant is in $\ell^p(\mathbb{N})$, with norm defined as the ℓ^p-norm of the majorant. Since the constructed $\boldsymbol{b}$ is mononotonically nonincreasing, it is equal to its minimal monotone majorant. Therefore, $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ and $\|\boldsymbol{b}\|_{p,\mathsf{M}} = \|\boldsymbol{b}\|_p = 1$. Then, using monotonicity once more, we get that

$$(\sigma_m(\boldsymbol{b})_2)^2 = c_{p,g}^2 \sum_{i=m+1}^{\infty} (ig(i))^{-2/p} \geq c_{p,g}^2 \sum_{i=m+1}^{2m} (ig(i))^{-2/p} \geq c_{p,g}^2 2^{-2/p} \cdot (g(2m))^{-2/p} m^{1-2/p}.$$

Hence, using part (a) we get

$$\overline{\theta_m}(p, \mathsf{M}) \geq c \cdot c_{p,g} 2^{-1/p} \cdot (g(2m))^{-1/p} \cdot m^{1/2-1/p},$$

where $c = C(\tau, \sigma)$ is the constant depending on ϱ in (6.6.11), as required. □

6.6.2 Proof of Theorem 6.3.2: unknown anisotropy, lower bound

We first proceed as in the proof of Theorem 6.3.1. The following lemma does for the m-width $\theta_m(p)$ what Lemma 6.6.2 did for $\theta_m(\boldsymbol{b})$.

Lemma 6.6.3 (Reduction to a discrete problem; unknown anisotropy case)**.** *Let $p \in (0, 1]$, $N \in \mathbb{N}$, ϱ be a tensor-product probability measure as in (2.2.4) and τ, σ be as in (6.6.1). Then the m-width (6.1.5) satisfies*

$$\theta_m(p) \geq C(\tau, \sigma) \cdot d^m(B_N^p, \ell_N^2),$$

where B_N^p is as in (6.5.1) with $\boldsymbol{w} = \mathbf{1}$ and

$$C(\tau, \sigma) = \frac{\sigma}{2 + |\tau|}. \tag{6.6.12}$$

Proof. Recall that $\theta_m(p)$ is defined by

$$\theta_m(p) = \inf\left\{ \sup_{f\in\mathcal{H}(p)} \|f - \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{S}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} : \boldsymbol{\mathcal{S}} : \mathcal{Y} \to \mathcal{V}^m \text{ adaptive}, \ \boldsymbol{\mathcal{R}} : \mathcal{V}^m \to L^2_\varrho(\mathcal{U};\mathcal{V}) \right\}.$$

Let $\boldsymbol{\mathcal{S}} : \mathcal{Y} \subset L^2_\varrho(\mathcal{U};\mathcal{V}) \to \mathcal{V}^m$ be a general adaptive sampling operator as in Definition 6.1.2 and v be the corresponding nonzero element of $\mathcal{V}$. Consider $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with $\boldsymbol{b} \in [0, \infty)^{\mathbb{N}}$ and $\|\boldsymbol{b}\|_p \leq 1$, and let $\boldsymbol{c} = (c_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be any sequence supported in $[N]$ with $|c_i| = b_i$ for $i \in [N]$. Define the function

$$f = \frac{\sigma}{(2 + |\tau|)\|v\|_{\mathcal{V}}} v \sum_{i=1}^{N} c_i \psi_i. \tag{6.6.13}$$

Lemma 6.6.1 implies that $f \in \mathcal{H}(\boldsymbol{b})$. We now use the same arguments as in the proof of Lemma 6.6.2 to obtain

$$\|f - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \geq c' \|\boldsymbol{c} - \boldsymbol{\Delta}(\boldsymbol{\Gamma}(\boldsymbol{c}))\|_2,$$

where $c' = \sigma/(2+|\tau|)$ and $\boldsymbol{\Gamma}$, $\boldsymbol{\Delta}$ are as before. We next take the supremum over $\boldsymbol{b} \geq \boldsymbol{0}$ with $\boldsymbol{b} \in \ell^p(\mathbb{N})$ and $\|\boldsymbol{b}\|_p \leq 1$ and all sequences $\boldsymbol{c}$ of the above form. Then we get

$$\sup_{f \in \mathcal{H}(p)} \|f - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \geq c' \sup_{\substack{\boldsymbol{c} \in \ell^p(\mathbb{N}),\ \|\boldsymbol{c}\|_p \leq 1 \\ \operatorname{supp}(\boldsymbol{c}) \subseteq [N]}} \|\boldsymbol{c} - \boldsymbol{\Delta}(\boldsymbol{\Gamma}(\boldsymbol{c}))\|_2.$$

Hence

$$\theta_m(p) \geq c' \inf \left\{ \sup_{\substack{\boldsymbol{c} \in \ell^p(\mathbb{N}),\ \|\boldsymbol{c}\|_p \leq 1 \\ \operatorname{supp}(\boldsymbol{c}) \subseteq [N]}} \|\boldsymbol{c} - \boldsymbol{\Delta}(\boldsymbol{\Gamma}(\boldsymbol{c}))\|_2 : \boldsymbol{\Gamma} : \mathbb{R}^N \to \mathbb{R}^m \text{ adaptive},\ \boldsymbol{\Delta} : \mathbb{R}^m \to \mathbb{R}^N \right\}.$$

Using this and (6.5.3) we see that

$$\theta_m(p) \geq c' E^m_{\mathsf{ada}}(B^p_N, \ell^2_N).$$

We now apply Theorem 6.5.1 with $\mathcal{K} = B^p_N$ and $\mathcal{X} = \ell^2_N$ to get the result. □

Proof of Theorem 6.3.2. We first prove part (a). To do so, we use the bound obtained in Lemma 6.6.3. Let $N \in \mathbb{N}$ be such that

$$N \geq m \mathrm{e}^{\frac{\log(3^8 \mathrm{e})}{2p} m - 1} \Leftrightarrow \frac{\frac{2p}{\log(3^8 \mathrm{e})} \log(\mathrm{e}N/m)}{m} \geq 1. \tag{6.6.14}$$

Then, from Proposition C.0.1 with $q = 2$ we obtain

$$d^m(B^p_N, \ell^2_N) \geq \left(\frac{1}{2}\right)^{2/p - 1/2}.$$

Thus,

$$\theta_m(p) \geq \frac{\sigma}{2+|\tau|} \left(\frac{1}{2}\right)^{2/p - 1/2},$$

as required. Part (b) follows immediately from part (b) of Theorem 6.3.1 and the inequality $\theta_m(p, \mathsf{M}) \geq \overline{\theta_m}(p, \mathsf{M})$. □

6.6.3 Proof of Theorem 6.3.3: known anisotropy, upper bound

Here we will employ polynomial techniques to establish the two upper bounds presented in Theorems 6.3.3 and 6.3.4. See Appendix B for further details on the Legendre coeffi-

cients summability and relevant polynomial approximation theory. We commence in this subsection with the proof of Theorem 6.3.3.

Proof of Theorem 6.3.3. The proof is divided into three parts. We first construct a sampling operator $\boldsymbol{\mathcal{S}}$ and show that it is a well-defined sampling operator in the sense of Definition 6.1.2. Then we prove parts (a) and (b), respectively.

Consider $\mathcal{F}$ to be the set of multi-indices with at most finitely-many zero entries and $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$ be the orthonormal Legendre basis of $L^2_\varrho(\mathcal{U})$. Let $S \subset \mathcal{F}$ be a finite index of size $|S| = m$ that will be chosen later in the proof and $\mathcal{Y} \supseteq \mathcal{H}(\boldsymbol{b})$ with $\mathcal{Y} \hookrightarrow L^2_\varrho(\mathcal{U};\mathcal{V})$. We now define $\boldsymbol{\mathcal{S}} : \mathcal{Y} \to \mathcal{V}^m$ and $\boldsymbol{\mathcal{R}} : \mathcal{V}^m \to L^2_\varrho(\mathcal{U};\mathcal{V})$ by

$$\boldsymbol{\mathcal{S}}(f) = \left(\langle f, \Psi_{\boldsymbol{\nu}}\rangle_{L^2_\varrho(\mathcal{U})}\right)_{\boldsymbol{\nu}\in S} \text{ and } \boldsymbol{\mathcal{R}}(\boldsymbol{v}) = \sum_{\boldsymbol{\nu}\in S} v_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}, \tag{6.6.15}$$

for any $f \in \mathcal{Y}$ and $\boldsymbol{v}= (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in S} \in \mathcal{V}^m$, respectively. Observe that $\langle f, \Psi_{\boldsymbol{\nu}}\rangle_{L^2_\varrho(\mathcal{U})} \in \mathcal{V}$ are precisely the coefficients of the expansion of f in (2.5.2) with $\Lambda = S$. However, we keep this notation to emphasize that $\boldsymbol{\mathcal{S}}$ is a linear operator.

We first prove that $\boldsymbol{\mathcal{S}}$ is well defined and that it satisfies the conditions of Definition 6.1.2. By construction and the fact that $\mathcal{Y} \hookrightarrow L^2_\varrho(\mathcal{U};\mathcal{V})$ we readily see that $\boldsymbol{\mathcal{S}}$ is a bounded linear operator. Therefore, it suffices to show there exists a normed vector space $\widetilde{\mathcal{Y}} \hookrightarrow L^2_\varrho(\mathcal{U})$, a nonzero $v \in \mathcal{V}$, and a bounded, linear scalar-valued sampling operator $\widetilde{\boldsymbol{\mathcal{S}}} : \widetilde{\mathcal{Y}} \to \mathbb{R}^m$ such that, if $vg \in \mathcal{Y}$ for $g \in L^2_\varrho(\mathcal{U})$, then $g \in \widetilde{\mathcal{Y}}$ and $\boldsymbol{\mathcal{S}}(vg) = v\widetilde{\boldsymbol{\mathcal{S}}}(g)$. To this end, let $v \in \mathcal{V}$, $\|v\|_{\mathcal{V}} = 1$, be arbitrary and define the space

$$\widetilde{\mathcal{Y}} = \{g \in L^2_\varrho(\mathcal{U}) : vg \in \mathcal{Y}\}.$$

It is easily seen that this is a vector space and that the quantity $\|g\|_{\widetilde{\mathcal{Y}}} = \|vg\|_{\mathcal{Y}}$, $\forall g \in \widetilde{\mathcal{Y}}$, defines a norm on $\widetilde{\mathcal{Y}}$. Moreover, using the fact that $\|v\|_{\mathcal{V}} = 1$ and $\mathcal{Y} \hookrightarrow L^2_\varrho(\mathcal{U};\mathcal{V})$, there exists a constant $C > 0$ such that

$$\|g\|_{L^2_\varrho(\mathcal{U})} = \|vg\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq C\|vg\|_{\mathcal{Y}} = C\|g\|_{\widetilde{\mathcal{Y}}}, \quad \forall g \in \widetilde{\mathcal{Y}}.$$

Hence $\widetilde{\mathcal{Y}} \hookrightarrow L^2_\varrho(\mathcal{U})$. We now define the scalar-valued sampling operator

$$\widetilde{\boldsymbol{\mathcal{S}}} : \widetilde{\mathcal{Y}} \to \mathbb{R}^m, \quad \widetilde{\boldsymbol{\mathcal{S}}}(g) = \left(\langle g, \Psi_{\boldsymbol{\nu}}\rangle_{L^2_\varrho(\mathcal{U})}\right)_{\boldsymbol{\nu}\in S}, \quad \forall g \in \widetilde{\mathcal{Y}}.$$

Note that $\widetilde{\boldsymbol{\mathcal{S}}}$ is closely related to the operator $\boldsymbol{\mathcal{S}}$, except that it is defined over a space $\widetilde{\mathcal{Y}}$ consisting of scalar-valued functions. Observe that $\widetilde{\boldsymbol{\mathcal{S}}}$ is linear and bounded, with the latter property due to the fact that $\widetilde{\mathcal{Y}} \hookrightarrow L^2_\varrho(\mathcal{U})$. Moreover, if $vg \in \mathcal{Y}$ then $g \in \widetilde{\mathcal{Y}}$ and $\boldsymbol{\mathcal{S}}(vg) = v\widetilde{\boldsymbol{\mathcal{S}}}(g)$ by construction. Therefore $\widetilde{\boldsymbol{\mathcal{S}}}$ satisfies the desired properties. We deduce that $\boldsymbol{\mathcal{S}}$ is a well defined operator that satisfies the conditions of Definition 6.1.2, as required.

We now prove part (a) of Theorem 6.3.3. Let $f \in \mathcal{H}(\boldsymbol{b})$. By (6.6.15) and Parseval's identity (see Remark 6.3.5) we get

$$\|f - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)\|^2_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \sum_{\boldsymbol{\nu} \notin S} \|c_{\boldsymbol{\nu}}\|^2_{\mathcal{V}},$$

where $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{F}}$ is as in (2.4.2) (see §B.1 for further information). Now we choose the specific set S. Let $\boldsymbol{b} \in \ell^p(\mathbb{N}), \boldsymbol{b} \in [0,\infty)^{\mathbb{N}}$. From Corollary B.2.3, there exists a set $S \subset \mathcal{F}$ of size $|S| = m$ depending on $\boldsymbol{b}$ and p only such that

$$\left(\sum_{\boldsymbol{\nu} \notin S} \|c_{\boldsymbol{\nu}}\|^2_{\mathcal{V}} \right)^{1/2} \leq C(\boldsymbol{b}, p) \cdot m^{1/2-1/p},$$

where $C(\boldsymbol{b}, p) > 0$ is the constant in Lemma B.2.1. Now, taking supremum over $f \in \mathcal{H}(\boldsymbol{b})$ we get

$$\sup_{f \in \mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}} \circ \boldsymbol{\mathcal{S}}(f)\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq C(\boldsymbol{b}, p) \cdot m^{1/2-1/p}.$$

Since $\boldsymbol{b}$ was arbitrary, this completes the proof of part (a).

Next we prove part (b). Let $q \in (p, 1)$. Then any $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ satisfies $\boldsymbol{b} \in \ell^q(\mathbb{N})$ and therefore

$$\theta_m(\boldsymbol{b}) \leq C(\boldsymbol{b}, q) \cdot m^{1/2-1/q}$$

by part (a). Using this we get

$$\overline{\theta_m}(p, \mathsf{M}) = \sup_{\|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1} \theta_m(\boldsymbol{b}) \leq \sup_{\|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1} C(\boldsymbol{b}, q) \cdot m^{1/2-1/q}.$$

The result now follows from Lemma B.2.2. □

Remark 6.6.4 As shown in this proof, the constant $c = c_{p,q}$ in part (b) of Theorem 6.3.3 comes from Lemma B.2.2. Unfortunately, the dependence of this constant on p and q is unknown, since it depends on an abstract summability criterion (see, e.g., [12, Lem. 3.29]) that does not give explicit upper bounds on the norm of the relevant sequence (in this case, the term $\|\tilde{\boldsymbol{g}}(p)\|_q$ in (B.2.7)). A first step towards understanding the constant $c_{p,q}$ would involve modifying the proof of this result to provide explicit bounds. However, one can already deduce from the proof of Lemma B.2.2 that $c_{p,q}$ must blow up as $q \to p^+$, since the sequence $\tilde{\boldsymbol{h}}(p) \notin \ell^p(\mathbb{N})$ (see (B.2.4)). Therefore, if part (b) of Theorem 6.3.3 were to hold with $q = p$, its proof would require a different technique.

6.6.4 Proof of Theorem 6.3.4: unknown anisotropy, upper bound

As in the previous proof, we rely on Legendre polynomial expansions. Consider the setup of §2.5 once more. In the previous proof, we made a judicious choice of index set $S \subset \mathcal{F}$

depending on the term $\boldsymbol{b}$ and used it to define the sampling and reconstruction map that gave the desired bound. In Theorem 6.3.4 we consider the case of unknown $\boldsymbol{b}$. Recalling the discussion in §6.4 (see also §2.4.5), we must proceed in a different way, in which the sampling operator $\boldsymbol{\mathcal{S}}$ and the reconstruction map $\boldsymbol{\mathcal{R}}$ do not rely on a specific choice of S. Fortunately, we can circumvent this issue by formulating a compressed sensing problem (see, e.g., §3.6). Here, by using a suitable set of (random) measurements and a recovery procedure, we can construct an approximation that gives us the desired error bounds.

First we restrict the search space to one that contains the union of all anchored sets, which has the desirable property that it is itself a finite index set. Given $n \in \mathbb{N}$, let $\Lambda \subset \mathcal{F}$ be the index set defined in (2.4.21). This set contains all anchored sets of size at most n [12, Prop. 2.18]. Recall from (5.5.4) that the size of this index set is bounded by

$$N := |\Lambda_n^{\mathsf{HCI}}| \leq e n^{2+\log(n-1)/\log(2)}, \quad \forall n \in \mathbb{N}.$$

Given Λ and f with expansion (2.4.2), consider the truncated expansion of f based on the index set Λ and its corresponding vector coefficients in $\mathcal{V}^N$ as (2.5.2). Here, $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N$ is some ordering of the multi-indices in Λ.

In this proof we also use the $\ell^p(\mathcal{F}; \mathcal{V})$- and $\ell^p([N]; \mathcal{V})$-norms, as defined in §2.2.

Setup and the vector recovery problem

We now describe the sampling operator and reconstruction map that are used to establish Theorem 6.3.4. Let $\boldsymbol{A} = 1/\sqrt{m}(a_{i,j})_{i,j=1}^{m,N} \in \mathbb{R}^{m \times N}$, where the $a_{i,j}$ are independent Gaussian random variables with zero mean and variance one. Note that the entries of $\boldsymbol{A}$ define the random vector $\boldsymbol{r}$ by allocating each entry of $\boldsymbol{A}$ to an entry of $\boldsymbol{r}$. Observe that this gives

$$\boldsymbol{r} \sim \mathcal{N}(0, I_{mN}). \tag{6.6.16}$$

Now let $\Phi_{i,\boldsymbol{r}} : \mathcal{U} \to \mathbb{R}$, $\Phi_{i,\boldsymbol{r}} = 1/\sqrt{m} \sum_{j=1}^{N} a_{i,j} \Psi_{\boldsymbol{\nu}_j}$ for $i = 1, \dots, m$, where $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \mathcal{F}}$ are the Legendre polynomials (see §2.4). We next define the sampling operator $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}} : \mathcal{Y} \subset L^2_{\varrho}(\mathcal{U}; \mathcal{V}) \to \mathcal{V}^m$ by

$$\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f) = \left(\langle f, \Phi_{i,\boldsymbol{r}} \rangle_{L^2_{\varrho}(\mathcal{U})} \right)_{i=1}^{m} = \boldsymbol{f}. \tag{6.6.17}$$

Using the same argument as in the proof of Theorem 6.3.3 with $\Phi_{i,\boldsymbol{r}}$ in place of $\Psi_{\boldsymbol{\nu}}$, we deduce that $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}$ is well-defined and satisfies Definition 6.1.2. Now, let $\boldsymbol{c}_\Lambda \in \mathcal{V}^N$ be the vector of coefficients of f_Λ in (2.4.2). Then

$$\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f) = \boldsymbol{A}\boldsymbol{c}_\Lambda.$$

Next, we define $\boldsymbol{\mathcal{R}}_{\boldsymbol{r}} : \mathcal{V}^m \to L^2_{\varrho}(\mathcal{U}; \mathcal{V})$ as the reconstruction map

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{v}) = \begin{cases} \sum_{\boldsymbol{\nu} \in \Lambda} \hat{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}} & \boldsymbol{v} \in \mathrm{Ran}(\boldsymbol{A}), \\ \boldsymbol{0} & \boldsymbol{v} \notin \mathrm{Ran}(\boldsymbol{A}), \end{cases} \tag{6.6.18}$$

where $\mathrm{Ran}(\boldsymbol{A})$ is the range of $\boldsymbol{A}$, the vector $\hat{\boldsymbol{c}}_\Lambda = (\hat{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ is defined by

$$\hat{\boldsymbol{c}}_\Lambda = \operatorname{argmin} \left\{ \|\boldsymbol{z}\|_{2;\mathcal{V}} : \boldsymbol{z} \in M(\boldsymbol{v}, \boldsymbol{A}) \right\},$$

and $M(\boldsymbol{v}, \boldsymbol{A})$ is the set of minimizers

$$M(\boldsymbol{v}, \boldsymbol{A}) = \operatorname*{argmin}_{\boldsymbol{z} \in \mathcal{V}^N} \|\boldsymbol{z}\|_{1;\mathcal{V}} \text{ subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{v}.$$

Note that we make this slightly awkward-looking definition for $\boldsymbol{\mathcal{R}}_{\boldsymbol{r}}$ to ensure that it is a well-defined (i.e., single-valued) map. In general, the minimization problem defined above does not have a unique solution. Thus, we choose the solution with the minimal ℓ^2-norm to enforce uniqueness. Further, the problem has no solution if $\boldsymbol{v} \notin \mathrm{Ran}(\boldsymbol{A})$. Hence in this case, we simply set $\boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{v}) = \boldsymbol{0}$.

Note that the composition of this reconstruction map with the sampling operator $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}$ gives that

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f)) = \sum_{\boldsymbol{\nu} \in \Lambda} \hat{c}_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}, \quad \text{where } \hat{\boldsymbol{c}}_\Lambda \in \operatorname{argmin} \|\boldsymbol{z}\|_{1;\mathcal{V}} \text{ subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{f}, \tag{6.6.19}$$

and $\boldsymbol{f}$ is as in (6.6.17). Because of this setup, in order to prove Theorem 6.3.4, we first need to consider properties of matrices $\boldsymbol{A} \in \mathbb{R}^{m \times N}$, and consequently operators in $\mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$, to recover (approximately) sparse vectors by solving the (Hilbert-valued) basis pursuit (BP) problem

$$\min_{\boldsymbol{z} \in \mathcal{V}^N} \|\boldsymbol{z}\|_{1;\mathcal{V}} \text{ subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{f}, \tag{6.6.20}$$

with $\boldsymbol{f} \in \mathcal{V}^m$. Specifically, we shall make use of the robust Null Space Property (rNSP). Note that this is different to the weighted robust Null Space Property (wrNSP) defined in §3.6.

Definition 6.6.5. A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ satisfies the *robust Null Space Property (rNSP)* of order $1 \leq s \leq N$ over $\mathcal{V}^N$ with constants $0 < \rho < 1$ and $\gamma > 0$ if

$$\|\boldsymbol{x}_S\|_{2;\mathcal{V}} \leq \frac{\rho \|\boldsymbol{x}_{S^c}\|_{1;\mathcal{V}}}{\sqrt{s}} + \gamma \|\boldsymbol{A}\boldsymbol{x}\|_{2;\mathcal{V}}, \quad \forall \boldsymbol{x} \in \mathcal{V}^N,$$

for any $S \subseteq [N]$ with $|S| \leq s$.

Recall that $\boldsymbol{x}_S$ is the vector in $\mathcal{V}^N$ with ith entry equal to x_i if $i \in S$ and zero otherwise. The following result is a straightforward application of [95, Prop. 4.2]. It shows that the rNSP is sufficient to provide an error bound for minimizers of the BP problem (6.6.20).

Theorem 6.6.6. *Suppose that $\boldsymbol{A} \in \mathbb{R}^{m\times N}$ has the rNSP over $\mathcal{V}$ of order $s \in [N]$ with constants $0 < \rho < 1$ and $\gamma > 0$. Let $\boldsymbol{x} \in \mathcal{V}^N$, $\boldsymbol{f} = \boldsymbol{A}\boldsymbol{x} \in \mathcal{V}^m$. Then every minimizer $\hat{\boldsymbol{x}} \in \mathcal{V}^N$ of the BP problem (6.6.20) satisfies*

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_{1;\mathcal{V}} \leq C_1 \sigma_s(\boldsymbol{x})_{1;\mathcal{V}} \text{ and } \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_{2;\mathcal{V}} \leq C_2 \frac{\sigma_s(\boldsymbol{x})_{1;\mathcal{V}}}{\sqrt{s}}, \tag{6.6.21}$$

where $C_1 = 2(1+\rho)/(1-\rho)$, $C_2 = 2(1+\rho)^2/(1-\rho)$.

Here $\sigma_s(\boldsymbol{x})_{p;\mathcal{V}}$ is the ℓ^p-norm best s-term approximation error, as defined in (2.4.8). Now consider a matrix $\boldsymbol{A} = 1/\sqrt{m}(a_{i,j})_{i,j=1}^{m,N}$, whose entries $a_{i,j}$ are independent Gaussian random variables with zero mean and variance one. Then, following similar arguments to those in Lemma 3.7.1 (or [12, Ch. 6] for the scalar case), one can show that if

$$m \geq c \cdot \left(s \cdot \log(2N/s) + \log(2\epsilon^{-1}) \right), \tag{6.6.22}$$

for some universal constant $c > 0$, then $\boldsymbol{A}$ satisfies rNSP over $\mathcal{V}$ of order s with constants $\rho \leq 1/2$ and $\gamma \leq 3/2$ with probability at least $1 - \epsilon$ (see [12, Thm. 6.11–6.12] for more details).

Remark 6.6.7 We note in passing that the universal constant c in (6.6.22) can be estimated. However, its precise value plays a minor role in what follows. The discussion in [12, §6.3.2] and the estimates in [112, Rmk. 9.28] suggests that in our case this constant can be taken as $c \approx 80.098 \cdot (2\sqrt{2}+1)^2$.

Error bounds in probability

To prove Theorem 6.3.4, we need to show an error bound in expectation. The first step towards doing this is establishing an error bound in probability. This is given by the following theorem.

Theorem 6.6.8. *Let $\mathcal{V}$ be a Hilbert space, $m \geq 3$, $0 < \epsilon < 1$, ϱ be the uniform probability measure on $\mathcal{U}$, $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}} : \mathcal{Y} \to \mathcal{V}^m$ and $\boldsymbol{\mathcal{R}}_{\boldsymbol{r}} : \mathcal{V}^m \to L^2_\varrho(\mathcal{U};\mathcal{V})$ be defined as in (6.6.17) and (6.6.18), respectively, where $\Lambda = \Lambda_n^{\mathsf{HCI}}$ is the index set in (2.4.21) with $n = \lceil m/\log^2(m) \rceil$, and $L = L(m,\epsilon) = \log^2(m) + \log(2\epsilon^{-1})$. Then the following hold.*

a) With probability at least $1 - \epsilon$,

$$\sup_{f \in \mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \widetilde{C}(\boldsymbol{b},p) \cdot \left(\frac{m}{L}\right)^{1/2-1/p},$$

for all $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$, $\boldsymbol{b} \geq \boldsymbol{0}$, *and* $0 < p < 1$.

b) With probability one,

$$\sup_{f \in \mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \widetilde{C}(\boldsymbol{b}, p),$$

for all $\boldsymbol{b} \in \ell^p(\mathbb{N})$, $\boldsymbol{b} \geq \boldsymbol{0}$, *and* $0 < p < 1$.

Here the constant $\widetilde{C}(\boldsymbol{b}, p)$ *depends on* $\boldsymbol{b}$ *and* p *only.*

Proof. We first prove part (b). Using triangle inequality and Parseval's identity we obtain

$$\|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|\boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \|\boldsymbol{c}\|_{2;\mathcal{V}} + \|\hat{\boldsymbol{c}}_\Lambda\|_{2;\mathcal{V}}, \quad \forall f \in \mathcal{H}(\boldsymbol{b}).$$

Using the inequality $\|\boldsymbol{c}\|_{2;\mathcal{V}} \leq \|\boldsymbol{c}\|_{1;\mathcal{V}}$ and the fact that $\hat{\boldsymbol{c}}_\Lambda$ is a minimizer of (6.6.19), we get

$$\|\hat{\boldsymbol{c}}_\Lambda\|_{2;\mathcal{V}} \leq \|\hat{\boldsymbol{c}}_\Lambda\|_{1;\mathcal{V}} \leq \|\boldsymbol{c}_\Lambda\|_{1;\mathcal{V}} \leq \|\boldsymbol{c}\|_{1;\mathcal{V}}.$$

Let $\boldsymbol{b} \in \ell^p(\mathbb{N})$, $\boldsymbol{b} \geq \boldsymbol{0}$, and $f \in \mathcal{H}(\boldsymbol{b})$. Lemma B.2.1 implies that the Legendre coefficients satisfy

$$\|\boldsymbol{c}\|_{1;\mathcal{V}} \leq \|\boldsymbol{c}\|_{p;\mathcal{V}} \leq C(\boldsymbol{b}, p),$$

where $C(\boldsymbol{b}, p)$ is the constant of this lemma. Therefore, taking the supremum over $f \in \mathcal{H}(\boldsymbol{b})$ we get

$$\sup_{f \in \mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq 2C(\boldsymbol{b}, p). \tag{6.6.23}$$

We give the final bound for part (b) after defining $\widetilde{C}(\boldsymbol{b}, p)$ in (6.6.25).

We now prove part (a). Let $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ be the random matrix used in the construction of $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}$ and recall that $\boldsymbol{A}$ extends to a bounded linear operator $\boldsymbol{A} : \mathcal{V}^N \to \mathcal{V}^m$ given by (2.5.3). Let

$$s = \left\lfloor \frac{m}{18c \cdot L(m, \epsilon)} \right\rfloor,$$

where $c > 0$ is the constant in (6.6.22). We now assume that $s \geq 1$, which is equivalent to $m/(18c \cdot L(m, \epsilon)) \geq 1$. We show part (a) with the assumption $s < 1$ at the end of the proof. From the estimate for N in (2.4.23) we get

$$\log(2N/s) \leq \log(\mathrm{e}N) \leq \log(\mathrm{e}^2 n^{2+\log(n)/\log(2)}) \leq \left(2 + \frac{\log(n)}{\log(2)}\right) \log(\mathrm{e}n) \leq 2(\log(\mathrm{e}n))^2.$$

Hence

$$\begin{aligned} s \cdot \log(2N/s) + \log(2\epsilon^{-1}) &\leq s \cdot (2(\log(en))^2 + \log(2\epsilon^{-1})) \\ &\leq 18s \cdot (\log^2(m) + \log(2\epsilon^{-1})) \\ &\leq 18s \cdot L(m, \epsilon). \end{aligned}$$

Here we use $n \leq 2m$ in the second inequality combined with $\log(2em) \leq 3\log(m)$ (since $m \geq 3$) in the second inequality. Now, using (6.6.22) and the definition of s, we deduce that, with probability $1 - \epsilon$ the matrix $\boldsymbol{A}$ has the rNSP over $\mathcal{V}$ of order s with constants $\rho \leq 1/2$ and $\gamma \leq 3/2$.

Next, we derive a bound for the approximation error. Using the triangle inequality we get

$$\|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \|f - f_\Lambda\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} + \|f_\Lambda - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}.$$

Now, Parseval's identity gives

$$\|f_\Lambda - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \|\boldsymbol{c}_\Lambda - \hat{\boldsymbol{c}}_\Lambda\|_{2;\mathcal{V}},$$

where $\hat{\boldsymbol{c}}_\Lambda$ is as in (6.6.19). Since $\boldsymbol{A}$ has the rNSP over $\mathcal{V}$, we may apply Theorem 6.6.6 to get

$$\|f_\Lambda - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq C_2 \frac{\sigma_s(\boldsymbol{c}_\Lambda)_{1;\mathcal{V}}}{\sqrt{s}} \leq 9 \frac{\sigma_s(\boldsymbol{c}_\Lambda)_{1;\mathcal{V}}}{\sqrt{s}}.$$

This last inequality is due to the fact that $\rho \leq 1/2$ in Theorem 6.6.6. Then, using Corollary B.2.3 with $q = 1 > p$ we get

$$\frac{\sigma_s(\boldsymbol{c}_\Lambda)_{1;\mathcal{V}}}{\sqrt{s}} \leq \frac{\sigma_s(\boldsymbol{c})_{1;\mathcal{V}}}{\sqrt{s}} \leq C(\boldsymbol{b}, p) \cdot s^{1/2-1/p}, \quad \forall f \in \mathcal{H}(\boldsymbol{b}),\ \boldsymbol{b} \in \ell^p(\mathbb{N}),\ \boldsymbol{b} \geq \boldsymbol{0}.$$

Now let $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$, $\boldsymbol{b} \geq \boldsymbol{0}$, $f \in \mathcal{H}(\boldsymbol{b})$ and consider the term $\|f - f_\Lambda\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}$. By applying Corollary B.3.3 with $q = 2 > p$, we get that there exists an anchored set $S_{\mathsf{A}} \subset \mathcal{F}$ with $|S_{\mathsf{A}}| \leq n$ such that

$$\|\boldsymbol{c} - \boldsymbol{c}_{S_{\mathsf{A}}}\|_{2;\mathcal{V}} \leq C_{\mathsf{A}}(\boldsymbol{b}, p) \cdot n^{1/2-1/p},$$

where the constant $C_{\mathsf{A}}(\boldsymbol{b}, p)$ is as in (B.3.3). Using the definition of Λ given in (2.4.21), we know that it contains the union of all anchored sets of size n and therefore $S_{\mathsf{A}} \subseteq \Lambda$. This yields

$$\|f - f_\Lambda\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \|\boldsymbol{c} - \boldsymbol{c}_\Lambda\|_{2;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_{S_{\mathsf{A}}}\|_{2;\mathcal{V}} \leq C_{\mathsf{A}}(\boldsymbol{b}, p) \cdot n^{1/2-1/p}, \quad \forall f \in \mathcal{H}(\boldsymbol{b}).$$

Combining these two results, noticing that $s \leq n$ and using the definition of s we get

$$\|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \widetilde{C}(\boldsymbol{b}, p) \cdot \left(\frac{m}{L}\right)^{1/2-1/p}, \quad \forall f \in \mathcal{H}(\boldsymbol{b}) \tag{6.6.24}$$

and all $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$, $\boldsymbol{b} \geq \mathbf{0}$, as required. Here the constant

$$\widetilde{C}(\boldsymbol{b}, p) = 9(18c)^{1/p-1/2} C(\boldsymbol{b}, p) + C_{\mathsf{A}}(\boldsymbol{b}, p), \tag{6.6.25}$$

depends on $\boldsymbol{b}$ and p only. Notice that, the inequality $2C \leq \widetilde{C}$ and (6.6.23) give the final bound for part (b).

Finally, we prove the case $s < 1$, i.e., $m < 18c \cdot L(m, \epsilon)$. Since $p < 1$, we have

$$1 < \left(\frac{m}{18c \cdot L}\right)^{1/2-1/p}.$$

From part (b) we deduce that

$$\sup_{f \in \mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq 2C(\boldsymbol{b}, p) \left(\frac{m}{18c \cdot L}\right)^{1/2-1/p} = (18c)^{1/p-1/2} 2C(\boldsymbol{b}, p) \left(\frac{m}{L}\right)^{1/2-1/p},$$

where C is the constant from Lemma B.2.1 and $c > 0$ the constant in (6.6.22). The result now follows immediately by noticing that $(18c)^{1/p-1/2} \cdot 2C \leq \widetilde{C}$, where $\widetilde{C}$ is as in (6.6.25). □

Proof of Theorem 6.3.4. Let $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}$, $\boldsymbol{\mathcal{R}}_{\boldsymbol{r}}$ be as in Theorem 6.6.8 and set

$$\epsilon = \left(\frac{m}{\log^2(m)}\right)^{1/2-1/p} < 1. \tag{6.6.26}$$

Recall that the random vector $\boldsymbol{r}$ is defined by (6.6.16), where $N = |\Lambda|$. Since $\Lambda = \Lambda_n^{\mathsf{HCI}}$ and $n = \lceil m/\log^2(m) \rceil$, we deduce that there exists an $\ell = \ell(m)$ depending on m only such that $\boldsymbol{r} \sim \mathcal{N}(0, I_\ell)$. In particular, $\ell(m) = m \cdot |\Lambda_n^{\mathsf{HCI}}|$. Next, using the definition of $\theta_m(p, \mathsf{M})$ in (6.1.5), notice that the following holds:

$$\theta_m(p, \mathsf{M}) \leq \mathbb{E}_{\boldsymbol{r} \sim \mathcal{N}(0, I_\ell)} \sup_{f \in \mathcal{H}(p, \mathsf{M})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})}.$$

We now prove part (a). Let $L(m, \epsilon)$ be as in Theorem 6.6.8. Since $m \geq 3$, we have $\log(m) > 1$. Therefore

$$L(m, \epsilon) \leq \log^2(m) + \log(2) + (1/p - 1/2)\log(m) \leq (1/2 + \log(2) + 1/p)\log^2(m) =: c_p \log^2(m),$$

where ϵ is as in (6.6.26). Now let E be the event

$$E = \left\{ \sup_{f \in \mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} \leq \overline{C}(\boldsymbol{b}, p) \cdot \left(\frac{m}{\log^2(m)}\right)^{1/2-1/p}, \ \forall \boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N}), \boldsymbol{b} \geq \mathbf{0}, 0 < p < 1 \right\},$$

where $\overline{C}(\boldsymbol{b},p) = (c_p)^{1/p-1/2} \cdot \widetilde{C}(\boldsymbol{b},p)$ and $\widetilde{C}(\boldsymbol{b},p)$ is the constant in Theorem 6.6.8. Then Theorem 6.6.8 implies that $\mathbb{P}(E^c) \leq \epsilon$. Hence

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{r}\sim\mathcal{N}(0,I_\ell)} \sup_{f\in\mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \\
&= \mathbb{E}\left(\sup_{f\in\mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \Big| E \right) + \mathbb{E}\left(\sup_{f\in\mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \Big| E^c \right) \\
&\leq \overline{C}(\boldsymbol{b},p) \cdot \left(\frac{m}{\log^2(m)} \right)^{1/2-1/p} + \widetilde{C}(\boldsymbol{b},p) \cdot \epsilon \\
&\leq 2\overline{C}(\boldsymbol{b},p) \cdot \left(\frac{m}{\log^2(m)} \right)^{1/2-1/p}.
\end{aligned}$$

This completes the proof of part (a).

We now prove part (b). Let $\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}$, $\boldsymbol{\mathcal{R}}_{\boldsymbol{r}}$ be as in Theorem 6.6.8 with ϵ as in (6.6.26). Using Theorem 6.6.8 part (a) we get that $\mathbb{P}(\widetilde{E}) \geq 1-\epsilon$, where

$$\widetilde{E} = \bigcap_{\substack{\boldsymbol{b}\in\ell^q_{\mathsf{M}}(\mathbb{N}),\boldsymbol{b}\geq\boldsymbol{0} \\ 0<q<1}} \left\{ \sup_{f\in\mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \leq \overline{C}(\boldsymbol{b},q) \cdot \left(\frac{m}{\log^2(m)} \right)^{1/2-1/q} \right\},$$

and

$$\overline{C}(\boldsymbol{b},q) = (c_q)^{1/q-1/2} \cdot \widetilde{C}(\boldsymbol{b},q) = 9\left(18c\left(\frac{1}{2} + \log(2) + \frac{1}{q} \right) \right)^{1/q-1/2} \left(C(\boldsymbol{b},q) + C_{\mathsf{A}}(\boldsymbol{b},q) \right),$$

where c is the constant in (6.6.22).

Note that $C(\boldsymbol{b},q)$ is the constant in Lemma B.2.1 and $C_{\mathsf{A}}(\boldsymbol{b},q)$ is the constant in Lemma B.3.1, both with q instead of p. Then, since $q > p$, applying Lemma B.2.2 and Lemma B.3.2 we obtain the following uniform bound for $\overline{C}(\boldsymbol{b},q)$:

$$\sup_{\|\boldsymbol{b}\|_{p,\mathsf{M}}\leq 1} \overline{C}(\boldsymbol{b},q) \leq 9\left(18c\left(\frac{1}{2} + \log(2) + \frac{1}{q} \right) \right)^{1/q-1/2} \left(\sup_{\|\boldsymbol{b}\|_{p,\mathsf{M}}\leq 1} C(\boldsymbol{b},q) + \sup_{\|\boldsymbol{b}\|_{p,\mathsf{M}}\leq 1} C_{\mathsf{A}}(\boldsymbol{b},q) \right) \leq c_{p,q},$$

where $c_{p,q}$ is a positive constant depending on p and q only. Next, recall that

$$\mathcal{H}(p,\mathsf{M}) = \bigcup \{ \mathcal{H}(\boldsymbol{b}) : \boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N}),\ \boldsymbol{b} \geq \boldsymbol{0},\ \|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1 \}.$$

Since $q > p$, any $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ satisfies $\boldsymbol{b} \in \ell^q_{\mathsf{M}}(\mathbb{N})$. Therefore

$$\begin{aligned}\widetilde{E} &\subseteq \bigcap_{\substack{\boldsymbol{b}\in\ell^p_{\mathsf{M}}(\mathbb{N}),\boldsymbol{b}\geq\boldsymbol{0}\\ \|\boldsymbol{b}\|_{p,\mathsf{M}}\leq 1\\ 0<p<q<1}} \left\{ \sup_{f\in\mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \leq \overline{C}(\boldsymbol{b},q)\cdot\left(\frac{m}{\log^2(m)}\right)^{1/2-1/q} \right\} \\ &\subseteq \bigcap_{\substack{\boldsymbol{b}\in\ell^p_{\mathsf{M}}(\mathbb{N}),\boldsymbol{b}\geq\boldsymbol{0}\\ \|\boldsymbol{b}\|_{p,\mathsf{M}}\leq 1\\ 0<p<q<1}} \left\{ \sup_{f\in\mathcal{H}(\boldsymbol{b})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \leq c_{p,q}\cdot\left(\frac{m}{\log^2(m)}\right)^{1/2-1/q} \right\} =: E.\end{aligned}$$

Therefore $\mathbb{P}(E) \geq \mathbb{P}(\widetilde{E}) \geq 1-\epsilon$ and by using a similar argument to that of part (a) we obtain

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{r}\sim\mathcal{N}(0,I_\ell)} \sup_{f\in\mathcal{H}(p,\mathsf{M})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \\ &= \mathbb{E}\left(\sup_{f\in\mathcal{H}(p,\mathsf{M})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \middle| E \right) + \mathbb{E}\left(\sup_{f\in\mathcal{H}(p,\mathsf{M})} \|f - \boldsymbol{\mathcal{R}}_{\boldsymbol{r}}(\boldsymbol{\mathcal{S}}_{\boldsymbol{r}}(f))\|_{L^2_{\varrho}(\mathcal{U};\mathcal{V})} \middle| E^c \right) \\ &\leq c_{p,q}\cdot\left(\frac{m}{\log^2(m)}\right)^{1/2-1/q} + c_{p,q}\cdot\epsilon \quad \leq \quad 2c_{p,q}\cdot\left(\frac{m}{\log^2(m)}\right)^{1/2-1/q},\end{aligned}$$

where $c_{p,q}$ is a constant depending on p and q only. To complete the proof, we need to remove the $\log^2(m)$ factor from the error bound. However, this follows immediately (up to a possible change in the constant) since $q \in (p,1)$ was arbitrary. □

6.7 Conclusions

This chapter introduced new theoretical guarantees for the limits of approximating Banach-valued, $(\boldsymbol{b},1)$-holomorphic functions in infinite dimensions from finite data, where $\boldsymbol{b} \in \ell^p(\mathbb{N})$, $0 < p < 1$. Specifically, in Theorems 6.3.1–6.3.4 we established new lower and upper bounds for various (adaptive) m-widths in both the known and unknown anisotropy cases. As we showed, when these bounds are decaying, they do so at a rate close to $m^{1/2-1/p}$.

Before answering Question 8 of §1.6 we provide a summary of the results.

1. We demonstrated that optimal recovery is attainable for functions in the monotone case, i.e., $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with an algebraic decay with respect to the number of samples m.

2. For functions in $\mathcal{H}(\boldsymbol{b})$, the best decay rate of the m-width $\theta_m(\boldsymbol{b})$ in terms of the number of samples m is the same as that of the best m-term approximation error $\sigma_m(\boldsymbol{b})_2$.

3. The rate $m^{1/2-1/p}$ is sharp when considering all possible (unit-norm) $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$. However, the type of $\boldsymbol{b}$ that achieves this bound corresponds to the class of functions in $\mathcal{H}(\boldsymbol{b})$ that is completely isotropic.

4. A nearly sharp lower bound of $m^{1/2-1/p}$ can be achieved for a fixed and m-independent $\boldsymbol{b} \in \ell^p(\mathbb{N})$, but dependent on a nondecreasing function

5. Approximating a Banach-valued, $(\boldsymbol{b}, 1)$-holomorphic function from limited data is impossible without any prior information on the variables, i.e., in the space $\mathcal{H}(p)$, in which $\boldsymbol{b}$ is unknown and $\boldsymbol{b} \in \ell^p(\mathbb{N})$ only. Specifically, the m-width $\theta_m(p)$ does not decay as $m \to \infty$.

6. In contrast, when considering functions in $\mathcal{H}(p, \mathsf{M})$, the best possible rate of decay of the corresponding m-width $\theta_m(p, \mathsf{M})$ is $m^{1/2-1/p}$.

7. In the case of a known fixed sequence $\boldsymbol{b} \in \ell^p(\mathbb{N})$, the rate $m^{1/2-1/p}$ can be achieved by a specific sampling-recovery pair $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{R}})$ up to a constant depending on $\boldsymbol{b}$.

8. Moreover, the same sampling-recovery pair $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{R}})$ yields a bound that holds uniformly for all $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ and achieves a rate $m^{1/2-1/t}$ for arbitrary $t > p$ in the case of known anisotropy.

9. There is a random sampling-recovery pair $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{R}})$ that is independent of $\boldsymbol{b}$ and p that achieves the upper bound $(m/\log^2(m))^{1/2-1/p}$ but with a constant depending on $\boldsymbol{b}$ in the case of unknown anisotropy.

10. The same sampling-recovery pair $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{R}})$ yields a bound that holds uniformly for all $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ and achieves a rate $m^{1/2-1/t}$ for arbitrary $t > p$ in the case of unknown anisotropy.

We now answer Question 8 of §1.6.

Answer to Question 8
Yes, the approximation rates derived in answering Questions 1–5 are near-optimal.

Note that this chapter does not consider algorithms achieving these rates, but 'reconstruction maps' involving minimizers of certain convex optimization problems. For a more detailed exploration using algorithms or DL to achieve rates of the form $(m/\text{polylog}(m))^{1/2-1/p}$ we refer to Chapter 4 and Chapter 5.

6.8 Future work

There are several other promising directions for future research.

- As mentioned there are a variety of methods that can achieve an approximation error decay rate of $(m/\text{polylog}(m))^{1/2-1/p}$ for $(\boldsymbol{b},\epsilon)$-holomorphic functions. For instance, in Chapters 3–5 we achieve an upper bound for the approximation error in the Hilbert-valued case for each $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ using i.i.d. pointwise samples, i.e., *standard information* as it is commonly termed [211, 212]. As discussed therein, the results shown in Chapters 3–5 are nonuniform guarantees, since the decay rates obtained from i.i.d. samples hold for a fixed function f. On the other hand, the results shown in this chapter are uniform, and therefore stronger, since they hold for any function belonging to the given class. However, our upper bounds do not consider pointwise samples. Ongoing work involves showing uniform guarantees (and therefore optimal approximation) for infinite-dimensional, holomorphic functions from i.i.d. pointwise samples. See [17] for recent work in this direction.

- As shown in part (b) of Theorems 6.3.3 and 6.3.4, our upper bounds for $\overline{\theta_m}(p,\mathsf{M})$ and $\theta_m(p,\mathsf{M})$ are non-sharp in comparison to the lower bounds shown in Theorems 6.3.1 and 6.3.2 by an algebraic factor that can be made arbitrarily small. We conjecture that this gap can be closed. A possible route towards doing so involves showing that the constant $C(\boldsymbol{b},p)$ in Lemma B.2.2 and $C_{\mathsf{A}}(\boldsymbol{b},p)$ in Lemma B.3.2 can be bounded uniformly for $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1$.

- In the case of known anisotropy, Theorem 6.3.1 part (b) establishes that approximation from finite data is possible, even for the worst case of $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with unit norm, i.e., there is a lower bound for $\overline{\theta_m}(p)$ with a rate of $m^{1/2-1/p}$. Nonetheless, Theorem 6.3.3 part (b) does not provide an analogous upper bound for $\overline{\theta_m}(p)$. As a result, it is interesting to bridge this gap. Ongoing work involves showing a uniform upper bounds for all $\boldsymbol{b}$ belonging to the unit ball of the space $\ell^p(\mathbb{N})$. See [17] for recent work in this direction.

- It is an open problem to extend Theorem 6.3.3 and Theorem 6.3.4 to the Banach-valued case. While the analysis in Chapter 5 indicates that such an extension is possible, using it would result in a worse exponent $1/2(1-1/p)$ (or $1/2(1-1/q)$) in the unknown anisotropy case or $2-1/p$ (or $2-1/q$) in the known anisotropy case, in place of $1/2-1/p$ (or $1/2-1/q$), which is suboptimal. Proving that the same rates can be achieved in the Banach-valued case would close a key theoretical gap observed in Chapter 5 between approximating holomorphic Hilbert- and Banach-valued functions.

- Finally, it remains an open problem to derive bounds in the $L^\infty(\mathcal{U};\mathcal{V})$-norm instead of the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm. The approximation theory of parametric PDEs is well studied in the $L^\infty(\mathcal{U};\mathcal{V})$-norm. Yet our proof strategy for the lower bounds on the m-widths relies on using the $L^2_\varrho(\mathcal{U};\mathcal{V})$-norm. We note that the best s-term polynomial approximation attains rates of the form $s^{1-1/p}$ in this norm. We therefore conjecture that versions of

our main theorems hold in this norm, except with an exponent of $1 - 1/p$ (or $1 - 1/q$) in place of $1/2 - 1/p$ (or $1/2 - 1/q$).

Chapter 7

Numerical approximation to parametric PDEs using DNNs

Up to this point in this thesis, we have conducted an in-depth theoretical study of approximating Banach-valued, holomorphic functions from pointwise samples. In particular Chapter 5 provides practical existence theorems that offer a theoretical justification for using DL in scientific computing, and Chapter 6 shows that the rates attained are essentially optimal. These theorems set a theoretical goal to achieve in practice, in terms of approximation errors, by demonstrating the existence of potentially effective architectures and training procedures for learning a holomorphic function from a limited number of sample points m. This chapter focuses on the practical aspects of DNNs for approximating smooth functions. Specifically, we implement various DNN architectures to approximate Hilbert- and Banach-valued functions that arise as solutions to parametric PDEs. Additionally, we study how the approximation, with respect to the given sample values, improves as we increase the number of samples. Section §7.1 introduces the key preliminary considerations for this chapter. In §7.2, we present the setup for the DNN approximation, including a detailed description of the methodology and the problem statement. In §7.3 we present the contributions of this chapter. Section §7.4 introduces the parametric PDE problems and provides a description of the parametric coefficients used. In §7.5, we describe the numerical results obtained, and in §7.6 , we provide additional general discussion on these results. Finally, in §7.7, we answer Questions 6–7 of §1.6.

7.1 Preliminaries

In Chapter 5, we constructed DL procedures that, in light of the main results from Chapter 6, are near-optimal for Hilbert-valued functions. However, as discussed, these procedures are based on emulating polynomial-based methods. These are of limited practical utility since DL aims to surpass the performance of these traditional methods, thereby making traditional methods less desirable for practical applications.

Specifically, the methods in Chapter 5 rely on handcrafted architectures where only the final layer is trained, and they utilize nonstandard loss functions involving explicit regularization. Therefore, evaluating how standard DL procedures–where all layers are trained, and the loss function is the standard ℓ^2-loss—perform in practice is key. Theoretical analysis of this case is beyond the scope of this work. Instead, we focus on the practical implementation and performance of DL on a series of challenging parametric DE problems.

7.2 Setup

It is important to note that we will use the standard notation u to refer to the function to be approximated (the solution) when specifically discussing the solution of parametric DEs, instead of f. Now we briefly recall the setup from §2.6. Let $K, d, L \in \mathbb{N}$. The goal is to approximate a given function $u : \mathcal{U} \to \mathcal{V}$, where $\mathcal{U}$ is the parameter space and $\mathcal{V}$ is a Banach or Hilbert space. Here $\{\varphi_k\}_{k=1}^K$ represents a basis of the finite-dimensional space $\mathcal{V}_K \subset \mathcal{V}$, e.g., a FE basis.

7.2.1 Deep neural network approximation

The approximation is achieved by constructing a fully connected feedforward DNN $\Phi : \mathbb{R}^d \to \mathbb{R}^K$ with L hidden layers, each with a constant width $N \in \mathbb{N}$, such that

$$u(\boldsymbol{y}) = \sum_{k=1}^{K} c_k(\boldsymbol{y})\varphi_k \approx u_\Phi(\boldsymbol{y}) = \sum_{k=1}^{K} (\Phi(\boldsymbol{y}))_k \varphi_k, \tag{7.2.1}$$

where the coefficients $c_k : \mathcal{U} \to \mathbb{R}$ are scalar-valued functions as in (5.1.2).

Given u, we compute a DNN approximation from the samples $\{(\boldsymbol{y}_i, u(\boldsymbol{y}_i))\}_{i=1}^m$ by minimizing a loss function $\mathcal{L} : \mathcal{N} \to \mathbb{R}$, where $\mathcal{N}$ is a family of DNNs of a chosen architecture. A common practice choice is to choose the ℓ^2-loss function

$$\mathcal{L}(\Phi) := \frac{1}{m} \sum_{j=1}^{m} \left(\sum_{k=1}^{K} (\Phi(\boldsymbol{y}_j)_k - c_k(\boldsymbol{y}_j))^2 \right), \quad \Phi \in \mathcal{N}. \tag{7.2.2}$$

Here, c_k are the coefficients defined in (7.2.1).

7.2.2 Methodology

In this section, we outline the general implementation details to ensure reproducibility and clarity and detail the methodology, including the choice of DNN architecture, training procedure, and evaluation metrics. We highlight any modifications made to the standard DNN setup to suit the numerical approximation of each parametric DE. Our experiments consist of three different PDEs: (i) a Poisson problem, (ii) a Navier-Stokes-Brinkman (NSB) problem, and (iii) a Boussinesq problem. We describe these in detail in §7.4.

We now summarize the main methodology:

(i) **Choice of architectures and initialization**. Based on the strategies in [16], we fixed the number of nodes per layer N and depth L such that the ratio $\beta := L/N$ is $\beta = 0.5$. We restrict our analysis to the Rectified Linear Unit (ReLU)

$$\sigma_1(z) := \max\{0, z\},$$

hyperbolic tangent (tanh)

$$\sigma_2(z) := \frac{e^z - e^{-z}}{e^z + e^{-z}},$$

or Exponential Linear Unit (ELU)

$$\sigma_3(z) = \begin{cases} z & z > 0, \\ e^z - 1 & z \leq 0 \end{cases}$$

activation function. Additionally, we use TensorFlow's He uniform initializer, which initializes the weights by drawing samples from a uniform distribution within the range determined by the He initialization scheme. We also set the seed for the random number generator. By providing a seed, the initialization process becomes deterministic, meaning that each time the code runs with the same seed value, the same initial weights will be generated.

(ii) **Training data and design of experiments**. First, we define a *trial* as a complete training run for a DNN approximating a specific function, initialized as mentioned above.

We run several trials solving the problem:

Given the measurements $\{(\boldsymbol{y}_i, u(\boldsymbol{y}_i))\}_{i=1}^m$, approximate a smooth function u.

Each of our architectures is trained across a range of datasets with increasing sizes. This involves utilizing a set of training data consisting of values $\{(\boldsymbol{y}_i, u(\boldsymbol{y}_i))\}_{i=1}^{m_k}$, where m_k denotes the size of the training data and belongs to the set $\{m_1, m_2, \ldots, m_{\text{final}}\}$. Here, each $u(\boldsymbol{y}_i)$ is generated by computing the solution of the parametric PDE at points drawn randomly and independently from the uniform measure $\{\boldsymbol{y}_i\}_{i=1}^{m_{\text{final}}} \subset \mathcal{U}$, and m_{final} represents a large number of total training points available for each trial, e.g., $m_{\text{final}} = 500$. Subsequently, for each trial we calculate the minimum testing error and run statistics across all trials for each dataset.

To be more specific, for each experiment we consider training with 14 sets of points of size $m_i \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500\}$, and for 9 different architectures (4×40, 5×50 and 10×100 with ReLU, ELU and tanh activation functions) over two dimensions ($d = 4$ and $d = 8$) and two PDE coefficients (a_1 from

(7.4.1) and a_2 from (7.4.2)). This setup results in 504 DNNs to be trained for each trial.

In this thesis we run experiments over 8 or 12 trials per solution depending on the PDE. That is, for 12 trials we trained and tested a total of 6,048 NNs per component of the solution of a PDE. For 8 trials we run a total 4,032 NNs per component of the solution of a PDE.

(iii) **Implementation**. The software employed for conducting our numerical experiments depends on the implementation of the open-source FE library `FEniCS`, specifically version 2019.1.0 [25], and Google's `Tensorflow` version 2.12.0. More information about `Tensorflow` at https://www.tensorflow.org/. The main code and instructions for running it on a local machine are accessible on GitHub via https://github.com/Sebanthalas/MLSPDE.

(iv) **Hardware**. We conducted most of our computations in single precision, utilizing the Intel Xenon Processor E5-2683 v4 CPU with either 125GB or 250GB of RAM per node available on the Compute Canada Cedar compute cluster provided by Simon Fraser University as a member of the Digital Research Alliance of Canada (https://alliancecan.ca).

The procedure for training and testing the DNN in the case of 12 trials is as follows: We first set up the architecture, the sample size m_i, dimension d and the PDE. Once these parameters are established, we conduct a total of 12 trials per experiment. We then submit a job array to the cluster (also known as a task array or an array job), dividing the 12 trials into threads of 3 trials each. That is, each thread is responsible for training and testing, linearly, 3 trials on a single core. Tables 7.1 and 7.2 illustrate this distribution for the experiments: Node 1, Thread 1 runs trials 1, 5, and 9; Node 1, Thread 2 runs trials 2, 6, and 10; Node 1, Thread 3 runs trials 3, 7, and 11; and Node 1, Thread 4 runs trials 4, 8, and 12. These tables also details the maximum memory used and the wall time for the thread. For the experiments over 8 trials we follow a similar idea, but only with 8 trials divided into threads of 2 trials. See Table 7.3 for an illustration.

Under normal circumstances, the training and testing would be carried out on faster, specialized GPUs, i.e., Cedar's Tesla P100 GPUs. However, during the experiments, we used Cedar's CPUs due to usage limitations and undergoing infrastructure upgrades on its Cedar clusters. This meant that the training and testing of each DNN took longer. However, one can take advantage of the large amount of CPUs available on Cedar's clusters and run the experiments in parallel reducing the total training and testing time. For more details on this incident, see https://status.alliancecan.ca/view_incident?incident=1100.

(v) **Optimizers for training and parametrization**. To train the DNN, we utilize the `Adam` optimizer [164], incorporating an exponentially decay learning rate. We train our model for 60,000 epochs or until converging to a tolerance level of $\epsilon_{\text{tol}} = 5 \cdot 10^{-7}$ in single precision. In light of the nonmonotonic convergence behaviour observed during the minimization of the nonconvex loss (see, e.g., [7, 16]), we implement an early stopping strategy. More precisely, we save the weights and biases of the partially trained DNN once the ratio between the current loss and the last checkpoint loss is reduced below 1/8 or if the current weights and biases produce the best loss value observed in training. We then restore these weights after training only if the loss value of the current weights is larger than that of the saved checkpoint.

(vi) **Testing data and error metric**. The testing data is generated similarly to the training data, obtaining solutions at different points $\boldsymbol{y}_i \subset \mathcal{U}$ for $i = 1, \ldots, m_{\text{test}}$. However, the testing data is generated by using a deterministic high-order sparse grid collocation method [209]. In particular, we use sparse grid quadrature rules to compute approximations to the Bochner norms

$$\|u\|_{L^2_\varrho(\mathcal{U};\mathcal{V})} = \left(\int_{\mathcal{U}} \|u(\boldsymbol{y})\|^2_{\mathcal{V}} \, \mathrm{d}\varrho(\boldsymbol{y}) \right)^{1/2} \approx \left(\sum_{i=1}^{m_{\text{test}}} \|u(\boldsymbol{y}_i)\|^2_{\mathcal{V}} w_i \right)^{1/2},$$

where ϱ is the measure defined in (2.2.4), using the approximation formula to the $L^2_\varrho(\mathcal{U};\mathcal{V})$ relative error

$$e_u^{\text{test}} = \frac{\left(\sum_{i=1}^{m_{\text{test}}} \|u_h(\boldsymbol{y}_i) - u_\Phi(\boldsymbol{y}_i)\|^2_{\mathcal{V}} w_i \right)^{1/2}}{\left(\sum_{i=1}^{m_{\text{test}}} \|u_h(\boldsymbol{y}_i)\|^2_{\mathcal{V}} w_i \right)^{1/2}},$$

where w_i are the quadrature weights associated with the sparse grid rule. As mentioned in [7] we use a high order isotropic Clenshaw Curtis sparse grid quadrature, since this method shows a superior convergence over Monte Carlo integration to evaluate the global error in Bochner spaces. The sparse grid rule gives m_{test} points at a level ℓ for d dimensions. We rely on the `TASMANIAN` sparse grid toolkit [242–244] for the generation of the isotropic rule to study the generalization performance of the DNN.

(vii) **Visualization**. The graphs in Figs. 7.5–7.22 show the geometric mean (the main curve) and plus/minus one (geometric) standard deviation (the shaded region). We use the geometric mean because our errors are plotted in logarithmic scale on the y-axis. See [12, Appx. A.1] for further discussion about this choice.

(viii) **Resources**. For the Poisson and Navier-Stokes-Brinkman PDEs, we run 12 trials. For the Boussinesq problem, we run 8 trials due to the larger problem size. In Tables 7.1–7.3 we show the maximum computational resources used per node for the most

demanding architecture in each problem. That is, $m_i = 500$ (samples), $d = 8$, $a = a_1$ (coefficient), $\sigma = \sigma_2$ (activation function) and $L = 10$ (number of layers).

- For the Poisson problem we used 504 nodes with 1×32 core CPUs (totaling 2016 threads, 6.8 GB RAM per node) each running 3 trials, approximately 4 hours and 15 minutes to complete.
- For the Navier-Stokes-Brinkman PDE we use the same setup, allocating 9.88 GB of RAM per node. The runs take approximately 9 hours and 13 minutes for each of the two components of the solution to complete.
- For the Boussinesq PDE, we allocate 504 nodes with 1×32 core CPUs running 4 threads per node (totaling 2016 threads, 10 GB of RAM per node) each running 2 trials, approximately 12 hours and 32 minutes for each of the 3 components to complete.

Given this, the total time required to reproduce the results in parallel having 504 nodes available with the above setup is approximately 60 hours or 2.5 days. The results were stored locally on the cluster and the estimated total space used to store the data for testing and training and results from computation is approximately 70 GB of memory. The trained models were not retained due to space limitations on the cluster.

Poisson equation - Maximum resources requested - per **data point**				
Node	Trials	RAM (GB)	CPU (hrs:min)	wall time (hrs:min)
Thread 1	1,5,9	1.734	4:00	4:14
Thread 2	2,6,10	1.737	4:04	4:15
Thread 3	3,7,11	1.748	3:47	4:00
Thread 4	4,8,12	1.672	3:43	4:00

Table 7.1: Shows the maximum requested and used resources for the Poisson problem for a single data point trought 12 different trials.

NSB problem - Maximum resources requested - per **data point**				
Node	Trials	RAM (GB)	CPU (hrs:min)	wall time (hrs:min)
Thread 1	1,5,9	2.53	9:01	9:11
Thread 2	2,6,10	2.57	9:04	9:13
Thread 3	3,7,11	2.53	9:01	9:12
Thread 4	4,8,12	2.49	9:02	9:12

Table 7.2: Shows the maximum requested and used resources for the NSB problem for a single data point trought 12 different trials.

Boussinesq problem - Maximum resources requested - per **data point**				
Node	Trials	RAM (GB)	CPU (hrs:min)	wall time (hrs:min)
Thread 1	1,5	1.933	11:44	11:51
Thread 2	2,6	1.922	12:17	12:23
Thread 3	3,7	1.918	12:24	12:31
Thread 4	4,8	1.932	12:25	12:32

Table 7.3: Shows the maximum requested and used resources for the Boussinesq problem for a single data point trought 8 different trials.

7.3 Contributions

Keeping this methodology in mind, we now formally define the contributions of this chapter. We divide this into three key constributions.

1. *We use DL to approximate the solutions of three specific parametric PDEs. First, based on a mixed formulation, we approximate the solution to a diffusion equation whose solution u is $(\boldsymbol{b}, \epsilon)$-holomorphic and Hilbert-valued. Second, we approximate the solution to a NSB problem in two dimensions, whose weak solution is Banach-valued. Third, we test DNNs approximating a coupled problem in 3D based on a Boussinesq approximation whose solution is also Banach-valued.*

2. *We provide insight into the practical approximation capabilities of DL with respect to the number of samples m for those parametric DEs. Specifically, we compare different architectures approximating these solutions and plot their results on a* log-log *plot to compare their approximation rate of decay as we increase the number of samples.*

3. *Differing from more standard variational formulations in Hilbert spaces, we show that mixed variational formulations (which offer several advantages to problems that are naturally posed in Banach spaces) can be combined with DL to provide practical implementations for approximating challenging PDEs.*

7.4 Main formulations

In this section, we introduce three parametric PDE problems whose solutions we aim to approximate using DL. As mentioned in §7.3, the first problem involves approximating a Hilbert-valued solution of a parametric diffusion equation with a physical domain in $\mathbb{R}^2$ using a mixed formulation, to which the theory developed in Chapter 5 applies. The second problem involves learning the solution to a parametric PDE with a Banach-valued solution, which serves to explore the performance of DL on parametric PDEs with mixed boundary conditions. Finally, we use DNNs to approximate the solution of a parametric PDE from

fluid mechanics with a physical domain in $\mathbb{R}^3$ and with three functions of interest, two of which are Banach-valued. These examples are particularly relevant to problems with stochastic coefficients, such as unexpected variations in the diffusion coefficient or randomly changing viscosity in fluid dynamics, which are among our main motivating problems (see §1.2). In this section we provide a brief summary of these formulations.

7.4.1 The parametric coefficients

As mentioned in §2.3.1, we study scenarios where the uncertainty arises as stochastic coefficients in a given parametric DE, which can be effectively modeled by using affine representations. Let Ω as the physical domain of a given PDE. Our experiments first consider the following affine diffusion coefficient

$$a_1(\boldsymbol{x}, \boldsymbol{y}) = 2.62 + \sum_{j=1}^{d} y_j \frac{\sin(\pi x_1 j)}{j^{3/2}}, \quad \forall \boldsymbol{x} \in \Omega, \forall \boldsymbol{y} \in [-1,1]^d. \tag{7.4.1}$$

Next, we consider a coefficient that arises from a Karhunen-Loève expansion (see §2.3.1), which is a modification of the example from [209] of a diffusion coefficient with one-dimensional (layered) spatial dependence given by

$$a_2(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(1 + y_1 \left(\frac{\sqrt{\pi}\beta}{2}\right)^{1/2} + \sum_{i=2}^{d} \zeta_i \, \vartheta_i(\boldsymbol{x}) \, y_i\right) \tag{7.4.2}$$

$$\zeta_i := (\sqrt{\pi}\beta)^{1/2} \exp\left(\frac{-\left(\left\lfloor \frac{i}{2} \right\rfloor \pi\beta\right)^2}{8}\right), \quad \vartheta_i(\boldsymbol{x}) := \begin{cases} \sin\left(\left\lfloor \frac{i}{2} \right\rfloor \pi x_1/\beta_p\right), & \text{if } i \text{ is even,} \\ \cos\left(\left\lfloor \frac{i}{2} \right\rfloor \pi x_1/\beta_p\right), & \text{if } i \text{ is odd,} \end{cases}$$

for $\boldsymbol{x} \in \Omega$ and $\boldsymbol{y} \in [-1,1]^d$. Here we let $\beta_c = 1/8$, and $\beta_p = \max\{1, 2\beta_c\}$, $\beta = \beta_c/\beta_p$. In order to implement the coefficients defined above we truncate them to a parameter dimension $d = 4, 8$.

7.4.2 The parametric diffusion equation

We commence with a well known elliptic parametric boundary value problem. Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain, $\partial\Omega$ be the boundary of Ω, $f \in \mathrm{L}^2(\Omega; \mathbb{R})$ and $g \in \mathrm{H}^{1/2}(\partial\Omega)$. See §A.2 for further details on the space $\mathrm{H}^{1/2}(\partial\Omega)$.

Given $\boldsymbol{y} \in \mathcal{U}$ consider the linear elliptic equation with Dirichlet boundary conditions

$$\begin{aligned} -\mathrm{div}(a(\boldsymbol{x}, \boldsymbol{y}) \nabla u(\boldsymbol{x}, \boldsymbol{y})) &= f(\boldsymbol{x}), \quad \text{in } \Omega, \\ u(\boldsymbol{x}, \boldsymbol{y}) &= g(\boldsymbol{x}), \quad \text{on } \partial\Omega. \end{aligned}$$

Here, the parameters are in $\mathcal{U} = [-1,1]^d$, the coefficient $a(\boldsymbol{x}, \boldsymbol{y})$ is parametric, the term $f(\boldsymbol{x})$ is nonparametric and $g(\boldsymbol{x})$ is nonparametric as well. With a slight abuse of notation, we

sometimes switch between the notation $u(\cdot, \boldsymbol{y})$ and $u(\boldsymbol{y})$ when referring to the parametric solution map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$.

The mixed variational formulation

Our first step in precisely defining the problem is to identify sufficient conditions on $\boldsymbol{y} \mapsto a(\boldsymbol{y})$ for the map $\boldsymbol{y} \mapsto u(\boldsymbol{y})$ to be well-defined. To do this, we turn our attention to the mixed variational formulation of the elliptic problem in (A.2.3) (see also (7.4.4)–(7.4.5) below). Using a mixed formulation to study the solution of PDEs offers several benefits. One key advantage is that it allows us to introduce additional variables that can be of physical interest. Additionally, mixed formulations can naturally accommodate different types of boundary conditions and introduce Dirichlet boundary conditions directly into the formulation rather than imposing them on the search space. For further details on mixed formulations, we refer to [114] and references within.

Assume that there exists $r, M > 0$, independent of $\boldsymbol{y}$ such that

$$0 < r \leq \operatorname{ess\,inf}_{\boldsymbol{x} \in \Omega} a(\boldsymbol{x}, \boldsymbol{y}) = a_{\min}(\boldsymbol{y}), \text{ and } a_{\max}(\boldsymbol{y}) = \operatorname{ess\,sup}_{\boldsymbol{x} \in \Omega} a(\boldsymbol{x}, \boldsymbol{y}) \leq M, \quad \forall \boldsymbol{y} \in \mathcal{U}. \tag{7.4.3}$$

Then, based on the analysis in §A.2, the problem can be stated as a first-order system: given $\boldsymbol{y} \in \mathcal{U}$, find $(\boldsymbol{\sigma}, u)(\boldsymbol{y}) \in \mathbf{H}(\operatorname{div}; \Omega) \times \mathrm{L}^2(\Omega)$ such that

$$d_{a(\boldsymbol{y})}(\boldsymbol{\sigma}(\boldsymbol{y}), \boldsymbol{\tau}) + b(\boldsymbol{\tau}, u(\boldsymbol{y})) = G(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in \mathbf{H}(\operatorname{div}; \Omega), \tag{7.4.4}$$

$$b(\boldsymbol{\sigma}, v) = F(v), \quad \forall v \in \mathrm{L}^2(\Omega). \tag{7.4.5}$$

Here d and b are the bilinear forms defined by

$$d_{a(\boldsymbol{y})}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \int_\Omega a^{-1}(\boldsymbol{y}) \boldsymbol{\sigma} \cdot \boldsymbol{\tau}, \quad \forall (\boldsymbol{\tau}, \boldsymbol{\sigma}) \in \mathbf{H}(\operatorname{div}; \Omega) \times \mathbf{H}(\operatorname{div}; \Omega)$$

$$b(\boldsymbol{\tau}, v) = \int_\Omega \operatorname{div}(\boldsymbol{\tau}) v, \quad \forall (\boldsymbol{\tau}, v) \in \mathbf{H}(\operatorname{div}; \Omega) \times \mathrm{L}^2(\Omega)$$

and the functionals $G \in (\mathbf{H}(\operatorname{div}; \Omega))'$ and $F \in \mathrm{L}^2(\Omega)$ are defined by

$$F(v) = -\int_\Omega f v, \quad \forall v \in \mathrm{L}^2(\Omega), \text{ and } G(\boldsymbol{\tau}) = \langle \gamma(\boldsymbol{\tau}) \cdot \boldsymbol{n}, g \rangle_{1/2, \partial\Omega}, \quad \forall \boldsymbol{\tau} \in \mathbf{H}(\operatorname{div}; \Omega). \tag{7.4.6}$$

Well-posedness and holomorphy

Given $\boldsymbol{y} \in \mathcal{U}$, $f \in \mathrm{L}^2(\Omega; \mathbb{R})$ and $g \in \mathrm{H}^{1/2}(\partial\Omega)$, consider a parametric coefficient $a \in L^\infty(\mathcal{U}; \mathrm{L}^\infty(\Omega; \mathbb{R}))$ as in (2.3.4) such that (7.4.3) holds. Then Theorem A.2.1 shows that the mapping $\boldsymbol{y} \mapsto (\boldsymbol{\sigma}, u)(\boldsymbol{y}) \in \mathbf{H}(\operatorname{div}; \Omega) \times \mathrm{L}^2(\Omega)$ is well-defined. Moreover, Proposition A.3.2 shows that it admits a holomorphic extension to a certain complex region.

For intance, consider the convergent affine representation (7.4.1). Observe that the condition (A.3.3) in Proposition A.3.3 holds with $M = 2.7$ and

$$\Big|\sum_{j\in\mathbb{N}} y_j \frac{\sin(\pi x_1 j)}{j^{3/5}}\Big| \leq \sum_{j\in\mathbb{N}} \frac{1}{j^{3/5}} \approx 2.61238 = 2.62 - r, \tag{7.4.7}$$

for some $r < 0.00762$. Then, the Hilbert-valued Poisson problem with coefficient a_1, $f \in \mathrm{L}^2(\Omega;\mathbb{R})$ and $g \in \mathrm{H}^{1/2}(\partial\Omega)$ analyzed via the mixed formulation (7.4.4)–(7.4.5) is well defined, and for each $\boldsymbol{y} \in \mathcal{U}$, there exists a unique solution $(\boldsymbol{\sigma}, u)(\boldsymbol{y}) \in \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{L}^2(\Omega)$. Furthermore, let $\boldsymbol{\rho} \geq \mathbf{1}$ and $0 < \epsilon < 0.00762$ be such that (A.3.4) holds with $\psi_j = \sin(\pi x_1 j)/j^{3/2}$. Then, the mapping $\boldsymbol{y} \mapsto (\boldsymbol{\sigma}, u)(\boldsymbol{y})$ admits a holomorphic extension to an open set containing $\mathcal{E}_{\boldsymbol{\rho}}$, where $\mathcal{E}_{\boldsymbol{\rho}}$ is the filled-in Bernstein polyellipse defined in (A.3.4). In other words, the mapping $\boldsymbol{y} \mapsto (\boldsymbol{\sigma}, u)(\boldsymbol{y})$ is $(\boldsymbol{b}, \varepsilon)$-holomorphic for $0 < \epsilon < 0.00762$ and $\boldsymbol{b} = (b_i)_{i\in\mathbb{N}}$ given by $b_j = \|\sin(\pi j\cdot)/j^{3/2}\|_{\mathrm{L}^\infty(\Omega)} = j^{-3/2}$.

Remark 7.4.1 Observe that when solving the Poisson problem with affine coefficient a_1 we have that $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ for every for $p < 2/3 \approx 0.666$. Thus, we expect a theoretical rate of convergence with respect to the amount of samples that is arbitrarily close to $m^{1/2-3/2} = m^{-1}$. This holomorphy result applies to the affine diffusion (7.4.1), not to the log-transformed diffusion coefficient (7.4.2). However, we consider the latter in our numerical experiments since it has been widely used in various similar works [7, 14, 16, 209] and we expect that it is possible to extend Proposition A.3.2 to this case.

The specific problem conditions

We now formally define the specific parametric diffusion equation considered in this thesis. This is as follows: given $\boldsymbol{y} \in [-1,1]^d$, find $u(\boldsymbol{y})$ satisfying

$$-\mathrm{div}(a_i(\boldsymbol{x},\boldsymbol{y})\nabla u(\boldsymbol{x},\boldsymbol{y})) = 10, \qquad \text{in } \Omega,$$

for $i = 1, 2$, where $\Omega = (0,1)^2$. See Fig. 7.1 for an illustration and a typical FEM discretization used in this work. For the Dirichlet boundary condition, we consider a constant value $u(\boldsymbol{x},\boldsymbol{y}) = 0.5$ on bottom of the boundary $(0,1)\times\{0\}$, and as well as a zero boundary conditions on the rest of the boundary.

The FE discretization

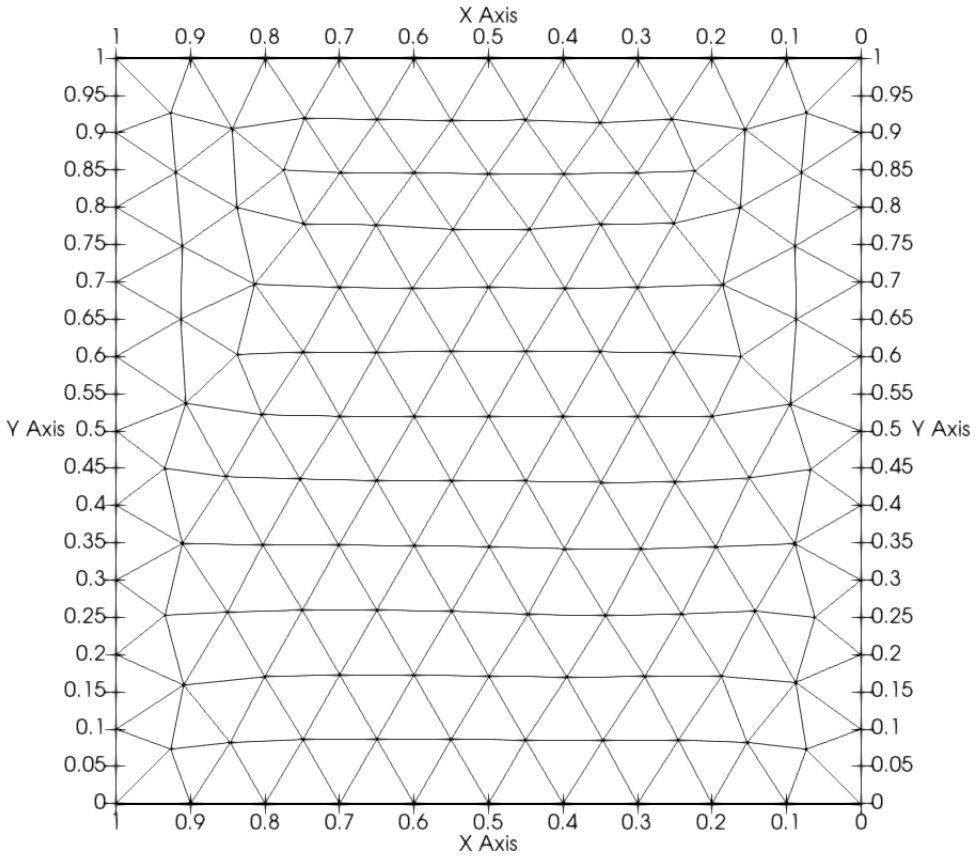

Figure 7.1: Shows the domain for the parametric difussion equation.

Notice that we consider approximations in the function spaces $\mathbf{H}(\mathrm{div};\Omega), \mathrm{L}^2(\Omega)$, which are infinite-dimensional spaces. We consider discretizations [67, Chp. 3] (in the sense of FEs) of the form $H_K \subseteq \mathbf{H}(\mathrm{div};\Omega)$ and $Q_K \subseteq \mathrm{L}^2(\Omega)$ for some $K \in \mathbb{N}$.

We let $\mathcal{T}_K$ be a regular triangulation of $\overline{\Omega}$ made up of triangles T of minimum diameter $h_{\min} = 0.0844$ and maximum diameter $h_{\max} = 0.1146$ with a total number of degrees of freedom $K = 2622$. The value of $h > 0$ represents a discretization parameter, i.e., the mesh size in this context. Therefore, we write the FE approximation of u in terms of the FE basis $\{\varphi_k\}_{k=1}^K$ as

$$u(\boldsymbol{y}) \approx u_h(\boldsymbol{y}) = \sum_{k=1}^{K} c_k(\boldsymbol{y})\varphi_k \in Q_K, \tag{7.4.8}$$

and likewise for $\boldsymbol{\sigma}(\boldsymbol{y})$.

Remark 7.4.2 Note that we mention the FEM in various parts of this thesis. As a disclaimer, this is not a thesis about specific Galerkin methods, and our theory and implementation apply to more general settings. Recall that we use these methods as black-box solvers to obtain the sample points and sample values. For instance, one may also consider using finite difference methods, spectral methods, finite volume methods, virtual elements

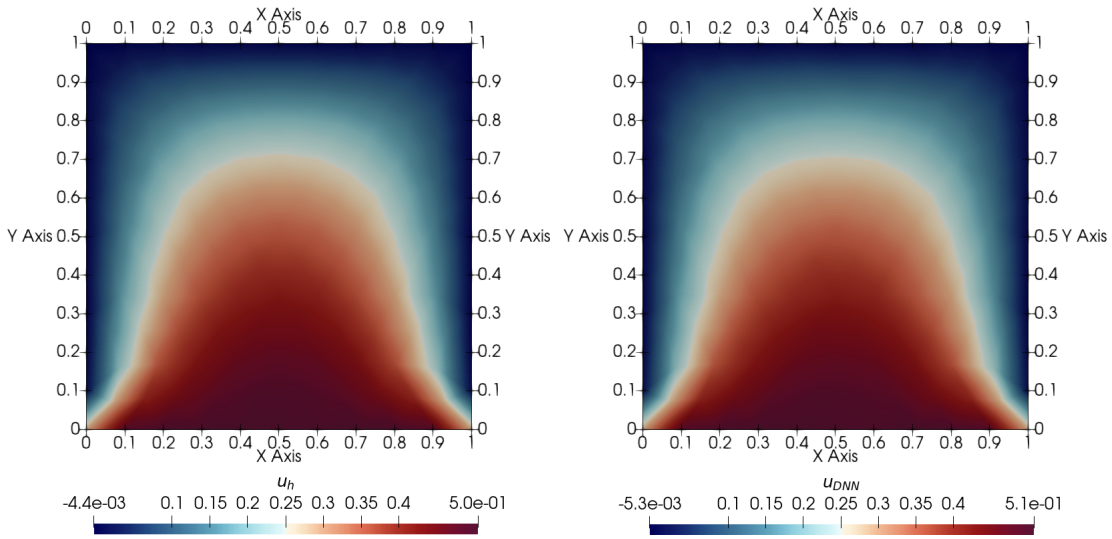

Figure 7.2: Shows the solution $(\boldsymbol{u})(\boldsymbol{y})$ to the parametric Poisson problem in (7.4.4)–(7.4.5) for a given parameter $\boldsymbol{y} = (1, 0, 0, 0)^\top$ with affine coefficient a_1, utilizing a total of 732 degrees of freedom (DoF) for $\boldsymbol{u}$. The left displays the solution given by the FEM solver, while the right column shows the 4×40 ELU DNN approximation after 60000 epochs of training with $m = 500$ sample points.

methods, etc. Here, we use the FEM for its versatility and relatively easy software implementation. We do not aim to discuss specific FEs to keep our analysis as general as possible. Therefore, we rely on the works in [67, 108] and references therein showing a detailed analysis of specific choices of FEs and their approximation properties to the problems presented in this chapter. See, e.g., [49, 67, 108, 114, 120] for further detailed introduction to the FEM and other Galerkin-type methods.

7.4.3 The Navier-Stokes-Brinkman equations

We now consider a parametric model describing the dynamics of a viscous fluid through porous media. Consider a bounded and Lipschitz physical domain $\Omega \subseteq \mathbb{R}^2$. Given $\boldsymbol{y} \in \mathcal{U}$, we consider a manufactured modelling of a fluid in a porous medium with random viscocity within Ω that can be described by the incompressible nonlinear stationary Navier-Stokes-

Brinkman (NSB) equations: find $\boldsymbol{u} : \mathcal{U} \times \Omega \to \mathbb{R}^2$ and $p : \mathcal{U} \times \Omega \to \mathbb{R}$ such that

$$\begin{aligned} \eta \boldsymbol{u} - \lambda \mathbf{div}(a(\boldsymbol{y})\boldsymbol{e}(\boldsymbol{u}(\boldsymbol{y}))) + (\boldsymbol{u}(\boldsymbol{y}) \cdot \nabla)\boldsymbol{u}(\boldsymbol{y}) + \nabla p(\boldsymbol{y}) &= f, \quad \text{in } \Omega \\ \operatorname{div}(\boldsymbol{u}(\boldsymbol{y})) &= 0, \quad \text{in } \Omega \\ \boldsymbol{u} &= \begin{cases} \boldsymbol{u}_D, & \text{on } \partial\Omega_{\text{in}} \\ 0, & \text{on } \partial\Omega_{\text{wall}} \end{cases} \\ (a\nabla \boldsymbol{e}(\boldsymbol{u}) - p\mathbb{I})\nu &= 0, \quad \text{on } \partial\Omega_{\text{out}} \\ \int_\Omega p &= 0, \end{aligned} \tag{7.4.9}$$

where $\lambda = \mathrm{Re}^{-1}$ and Re is the Reynolds number, $a : \mathcal{U} \times \Omega \to \mathbb{R}_+$ is the random viscosity of the fluid, $\eta \in \mathbb{R}_+$ is the scaled inverse permeability of the porous media, $\boldsymbol{u}$ is the velocity of the fluid, $\boldsymbol{e}(\boldsymbol{u}) = \frac{1}{2}(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^t)$ is the symmetric part of the gradient and p is the pressure of the fluid, and $f : \Omega \to \mathbb{R}$ is an external force independent of the parameters. Here, the fourth condition imposes a zero normal Cauchy stress

$$(a\nabla \boldsymbol{e}(\boldsymbol{u}) - p\mathbb{I})\nu = 0$$

for the output boundary on $\partial\Omega_{\text{out}}$. Note as well that the incompressibility of the fluid imposes on $\boldsymbol{u}_D$ the compability condition

$$\int_{\partial\Omega} \boldsymbol{u}_D \cdot \boldsymbol{n} = 0,$$

on $\partial\Omega_{\text{in}}$. The third condition imposes a no-slip condition on the walls Ω_{wall} [115, eq.(2.3)].

The analysis of the detailed mixed formulation used for this problem in the nonparametric case is given in [115]. Over the last decade, many works have employed a mixed formulation using a Banach space framework, solving different PDEs in continuum mechanics in suitable Banach spaces. The advantage of this formulation is that no augmentation is required, the spaces are simpler and closer to the original model, and it allows for obtaining more direct approximations of variables of physical interest [115, §1].

The mixed variational formulation

Based on the analysis in [115], the mixed variational formulation of the parametric NSB equations in (7.4.9) becomes: given $\boldsymbol{y} \in \mathcal{U}$, find $(\boldsymbol{u}, \boldsymbol{t}, \boldsymbol{\sigma}, \boldsymbol{\gamma})(\boldsymbol{y}) \in \mathbf{L}^4(\Omega) \times \mathbb{L}^2_{\text{tr}}(\Omega) \times$

$\mathbb{H}_0(\mathbf{div}_{4/3};\Omega)\times\mathbb{L}^2_{\mathrm{skew}}(\Omega)$ such that

$$\begin{aligned}\lambda\int_\Omega a_i(\boldsymbol{y})\boldsymbol{t}(\boldsymbol{y}):\boldsymbol{s}\;-\;\int_\Omega \boldsymbol{s}:\boldsymbol{\sigma}(\boldsymbol{y})\;-\;\int_\Omega(\boldsymbol{u}\otimes\boldsymbol{u})(\boldsymbol{y}):\boldsymbol{s}\;&=\;0,\\ \int_\Omega \boldsymbol{t}(\boldsymbol{y}):\boldsymbol{\tau}\;+\;\int_\Omega\boldsymbol{\gamma}(\boldsymbol{y}):\boldsymbol{\tau}\;+\;\int_\Omega\boldsymbol{u}(\boldsymbol{y})\cdot\mathbf{div}(\boldsymbol{\tau})\;&=\;\langle\boldsymbol{\tau}\boldsymbol{n},\boldsymbol{u}_D\rangle_{\partial\Omega_{\mathrm{in}}},\\ \int_\Omega\boldsymbol{\delta}:\boldsymbol{\sigma}(\boldsymbol{y})\;+\;\int_\Omega\boldsymbol{v}\cdot\mathbf{div}(\boldsymbol{\sigma}(\boldsymbol{y}))\;-\;\int_\Omega\eta\boldsymbol{u}(\boldsymbol{y})\cdot\boldsymbol{v}\;&=\;\int_\Omega\boldsymbol{f}\cdot\boldsymbol{v},\end{aligned}\tag{7.4.10}$$

for all $(\boldsymbol{v},\boldsymbol{s},\boldsymbol{\tau},\boldsymbol{\delta})\in\mathbf{L}^4(\Omega)\times\mathbb{L}^2_{\mathrm{tr}}(\Omega)\times\mathbb{H}_0(\mathbf{div}_{4/3};\Omega)\times\mathbb{L}^2_{\mathrm{skew}}(\Omega)$. Moreover, we impose the Neumann boundary condition via a Nietsche method as in [115, §5.2] adding

$$\kappa\langle(\boldsymbol{\sigma}+\boldsymbol{u}\otimes\boldsymbol{u})\boldsymbol{n}),\boldsymbol{\tau}\boldsymbol{n}\rangle_{\partial\Omega_{\mathrm{out}}}=0$$

to in (7.4.10) where $\kappa\gg1$ is a large constant (e.g., $\kappa=10^4$). As usual in this formulations, the pressure $p\in\mathrm{L}^2_0(\Omega)$ can be computed according to the postprocessing formula

$$p=-\frac{1}{2}\mathrm{tr}(\boldsymbol{\sigma}+(\boldsymbol{u}\otimes\boldsymbol{u})).$$

Note that above we omitted the term $\boldsymbol{y}$ for simplicity.

The specific problem conditions

In particular, we consider to approximate solutions to the parametric NSB problem with $\lambda=0.1$, a scaled inverse permeability of $\eta=10+x_1^2+x_2^2$, an external force $\boldsymbol{f}=(0,-1)^\top$, and random viscosity a_i as in (7.4.1)–(7.4.2) with $i=1,2$.

We once more consider the unit square as the domain $\Omega=(0,1)^2$. We consider an inlet boundary defined by $\partial\Omega_{\mathrm{in}}=(0,1)\times\{1\}$, an outlet boundary $\partial\Omega_{\mathrm{out}}=\{1\}\times(0,1)$ and walls defined by $\partial\Omega_{\mathrm{wall}}=\{0\}\times(0,1)\cup(0,1)\times\{0\}$. For simplicity, we use the same mesh as that of the previous example. See Fig. 7.1.

On the Neumann boundary $\partial\Omega_{\mathrm{out}}$ we consider a zero normal Cauchy stress. We consider a Dirichlet condition given by $\boldsymbol{u}_D=(0.0625)^{-1}((x_2-0.5)(1-x_2),0)$ on $\partial\Omega_{\mathrm{in}}$ and a no-slip velocity on $\partial\Omega_{\mathrm{wall}}$.

Remark 7.4.3 (Other auxiliary variables) As we describe later in §7.5, we report the performance of the DNNs approximating $(\boldsymbol{u},p)(\boldsymbol{y})\in\mathbf{L}^4(\Omega)\times\mathrm{L}^2(\Omega)$. Note that any solver based on the above formulation outputs several other variables, e.g., $(\boldsymbol{t},\boldsymbol{\sigma},\boldsymbol{\gamma})(\boldsymbol{y})\in\mathbb{L}^2_{\mathrm{tr}}(\Omega)\times\mathbb{H}_0(\mathbf{div}_{4/3};\Omega)\times\mathbb{L}^2_{\mathrm{skew}}(\Omega)$. One could also approximate these auxiliary variables using DNNs. However, we restrict our experiments to $(\boldsymbol{u},p)$ as these are the primary variables of interest in the problem.

Testing the approximation capabilities of DNNs applied to more complex PDEs, such as the NSB problem, reveals that one can practically implement DL to solve parametric

DEs whose solution is Banach-valued. However, it is important to mention that we do not analyze the holomorphy of the map $\boldsymbol{y} \mapsto (\boldsymbol{u}, p)(\boldsymbol{y})$ in this thesis. Doing so would require extending the theory in Appendix A to the Banach-valued case, which requires further investigation and goes beyond the scope of this thesis. We leave the study of holomorphic extensions of the map $\boldsymbol{y} \mapsto (\boldsymbol{u}, p)(\boldsymbol{y})$ as future work.

7.4.4 The stationary parametric Boussinesq equations

In our first two-dimensional parametric diffusion example, the primary challenge is ensuring the $(\boldsymbol{b}, \varepsilon)$-holomorphy of the map $\boldsymbol{y} \mapsto u(\boldsymbol{y}) \in \mathrm{L}^2(\Omega)$ for the mixed formulation. Additionally, we consider a nonzero Dirichlet boundary condition, unlike previous works [16, 71, 95] that focus on the more restrictive homogeneous Dirichlet case with $u \in \mathrm{H}^1_0(\Omega)$.

In our second example, we use DL to approximate a solution with two components of a PDEs with mixed boundary conditions. Specifically, we use DL to approximate the solution of a PDE with Dirichlet and Neumann boundary conditions, as well as Banach-valued components in the solution, e.g., $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$. While this example currently lacks a holomorphy guarantee, we observe a convergence rate that aligns with what we expect (see §7.5). We conjecture this rate holds for more challenging problems.

To illustrate this claim with an example, we now consider a parametric coupled PDE in three dimensions ($\Omega \subset \mathbb{R}^3$) with two random coefficients affecting different parts of the coupled problem. The nonparametric version of this problem is taken from [80].

The Boussinesq model arises in a variety of engineering and fluid dynamics problems where changes in temperature affect the velocity of the fluid. Here, we consider a modification of the Boussinesq formulation in [80] that combines a parametric incompressible Navier–Stokes equation with a parametric heat equation. The parametric dependence affects both: the Navier–Stokes equation is affected by a parametric variable multiplying the temperature-dependent viscosity, and the equation for heat flow is affected directly by the thermal conductivity of the fluid. To be more precise, given $\boldsymbol{y} \in \mathcal{U}$, our goal is to find the velocity $\boldsymbol{u} : \mathcal{U} \times \Omega \to \mathbb{R}^2$, pressure $p : \mathcal{U} \times \Omega \to \mathbb{R}$ and temperature $\varphi : \mathcal{U} \times \Omega \to \mathbb{R}$ of a fluid such that

$$
\begin{aligned}
-\mathbf{div}(2a(\boldsymbol{y})\mu(\varphi(\boldsymbol{y}))\boldsymbol{e}(\boldsymbol{u}(\boldsymbol{y}))) + (\boldsymbol{u}(\boldsymbol{y}) \cdot \nabla)\boldsymbol{u}(\boldsymbol{y}) + \nabla p(\boldsymbol{y}) &= \varphi(\boldsymbol{y})\boldsymbol{g}, \quad \text{in } \Omega, \\
\operatorname{div}(\boldsymbol{u}(\boldsymbol{y})) &= 0, \quad \text{in } \Omega, \\
-\operatorname{div}(\mathbb{K}(\boldsymbol{y})\nabla\varphi(\boldsymbol{y})) + \boldsymbol{u}(\boldsymbol{y}) \cdot \nabla\varphi(\boldsymbol{y}) &= 0, \quad \text{in } \Omega, \\
\boldsymbol{u} &= \boldsymbol{u}_D, \text{ on } \partial\Omega, \\
\varphi &= \varphi_D, \text{ on } \partial\Omega, \\
\int_\Omega p(\boldsymbol{y}) &= 0,
\end{aligned}
\tag{7.4.11}
$$

where $\boldsymbol{g} = (0, 0, -1)^\top$ is a gravitational force, $\mathbb{K} : \mathcal{U} \times \Omega \to \mathbb{R}^{3\times 3}$ is a parametric uniformly positive tensor describing the thermal conductivity of the fluid, given by

$$\mathbb{K}(\boldsymbol{x}, \boldsymbol{y}) = \left(1.89 + \sum_{j\in\mathbb{N}} y_j \frac{\sin(\pi x_3 j)}{j^{9/5}}\right) \begin{bmatrix} \exp(-x_1) & 0 & 0 \\ 0 & \exp(-x_2) & 0 \\ 0 & 0 & \exp(-x_3) \end{bmatrix}, \tag{7.4.12}$$

for all $x \in \Omega$, and $\boldsymbol{y} \in [-1, 1]^{\mathbb{N}}$, $\mu : \mathbb{R} \to \mathbb{R}_+$ is the temperature dependent viscosity given by $\mu(\varphi) = 0.1 + \exp(-\varphi)$, and $a : \mathcal{U} \times \Omega \to \mathbb{R}$ is a parametric variable affecting the viscosity of the fluid assumed to be in $\mathrm{L}^\infty(\Omega)$. Then, in this case we have $(a(\boldsymbol{y}), \mathbb{K}(\boldsymbol{y})) \in \mathrm{L}^\infty(\Omega)\times\mathbb{L}^\infty(\Omega)$. As in the previous example $\boldsymbol{e}(\boldsymbol{u})$ is the symmetric part of $\nabla\boldsymbol{u}$.

The fully-mixed variational formulation

The complete derivation of a fully-mixed variational formulation for the non-parametric Boussinesq equation in Banach spaces can be found in [80, §3.1]. We now rewrite it for the parametric case. Given $\boldsymbol{y} \in \mathcal{U}$, find $(\boldsymbol{u}, \boldsymbol{t}, \boldsymbol{\sigma}, \varphi, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\sigma}})(\boldsymbol{y}) \in \mathbf{L}^4(\Omega) \times \mathbb{L}^2_{\mathrm{tr}}(\Omega) \times \mathbb{H}_0(\mathbf{div}_{4/3}; \Omega) \times \mathrm{L}^4(\Omega) \times \mathbf{L}^2(\Omega) \times \mathbf{H}(\mathrm{div}_{4/3}; \Omega)$ such that

$$\begin{aligned}
-\int_\Omega \boldsymbol{v}\cdot\mathbf{div}(\boldsymbol{\sigma}(\boldsymbol{y})) + \frac{1}{2}\int_\Omega \boldsymbol{t}(\boldsymbol{y})\boldsymbol{u}(\boldsymbol{y})\cdot\boldsymbol{v} - \int_\Omega \varphi(\boldsymbol{y})\boldsymbol{g}\cdot\boldsymbol{v} &= 0 && \forall\, \boldsymbol{v} \in \mathbf{L}^4(\Omega)\,,\\
\int_\Omega 2a(\boldsymbol{y})\mu(\varphi(\boldsymbol{y}))\boldsymbol{t}_{\mathrm{sym}}(\boldsymbol{y}) : \boldsymbol{s} - \frac{1}{2}\int_\Omega (\boldsymbol{u}\otimes\boldsymbol{u})(\boldsymbol{y}) : \boldsymbol{s} &= \int_\Omega \boldsymbol{\sigma}(\boldsymbol{y}) : \boldsymbol{s} && \forall\, \boldsymbol{s} \in \mathbb{L}^2_{\mathrm{tr}}(\Omega)\,,\\
\int_\Omega \boldsymbol{\tau} : \boldsymbol{t}(\boldsymbol{y}) + \int_\Omega \boldsymbol{u}(\boldsymbol{y})\cdot\mathbf{div}(\boldsymbol{\tau}) &= \langle\boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_D\rangle_{\partial\Omega} && \forall\, \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div}_{4/3}; \Omega)\,,\\
-\int_\Omega \psi\,\mathrm{div}(\tilde{\boldsymbol{\sigma}}(\boldsymbol{y})) + \frac{1}{2}\int_\Omega \psi(\boldsymbol{y})\boldsymbol{u}(\boldsymbol{y})\cdot\tilde{\boldsymbol{t}} &= 0 && \forall\, \psi \in \mathrm{L}^4(\Omega)\,,\\
\int_\Omega \mathbb{K}(\boldsymbol{y})\tilde{\boldsymbol{t}}(\boldsymbol{y})\cdot\tilde{\boldsymbol{s}} - \frac{1}{2}\int_\Omega \varphi(\boldsymbol{y})\boldsymbol{u}(\boldsymbol{y})\cdot\tilde{\boldsymbol{s}} &= \int_\Omega \tilde{\boldsymbol{\sigma}}(\boldsymbol{y})\cdot\tilde{\boldsymbol{s}} && \forall\, \tilde{\boldsymbol{\sigma}} \in \mathbf{L}^2(\Omega)\,,\\
\int_\Omega \tilde{\boldsymbol{\tau}}\cdot\tilde{\boldsymbol{t}}(\boldsymbol{y}) + \int_\Omega \varphi(\boldsymbol{y})\,\mathrm{div}(\tilde{\boldsymbol{\tau}}) &= \langle\tilde{\boldsymbol{\tau}}\cdot\boldsymbol{\nu}, \varphi_D\rangle_{\partial\Omega} && \forall\, \tilde{\boldsymbol{\tau}} \in \mathbf{H}(\mathrm{div}_{4/3}; \Omega)\,,\\
\int_\Omega \mathrm{tr}(2\boldsymbol{\sigma} + \boldsymbol{u}\otimes\boldsymbol{u})(\boldsymbol{y}) &= 0, &&
\end{aligned} \tag{7.4.13}$$

where $p \in L^2_0(\Omega)$ can be recovered by using

$$p = -\frac{1}{6}\mathrm{tr}(2\sigma + 2c\mathbb{I} + \boldsymbol{u}\otimes\boldsymbol{u}), \text{ with } c = -\frac{1}{6|\Omega|}\int_\Omega \mathrm{tr}(\boldsymbol{u}\otimes\boldsymbol{u}). \tag{7.4.14}$$

As in the previous example, we omitted the term $\boldsymbol{y}$ for simplicity from this equation. For further details on this formulation we refer to [80] and references within.

Here, given $\boldsymbol{y} \in \mathcal{U}$, we consider to approximate

$$\boldsymbol{y} \in \mathcal{U} \mapsto (\boldsymbol{u}, p, \varphi)(\boldsymbol{y}) \in (\mathbf{L}^4(\Omega) \times L_0^2(\Omega) \times \mathrm{L}^4(\Omega)) \tag{7.4.15}$$

of (7.4.13) by using DNNs and study the approximation capabilities as we increase the number of training samples m. As in the previous example, we do not aim to approximate the other variables $(\boldsymbol{t}, \boldsymbol{\sigma}, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\sigma}})(\boldsymbol{y}) \in \mathbb{L}^2_{\mathrm{tr}}(\Omega) \times \mathbb{H}_0(\mathbf{div}_{4/3}; \Omega) \times \mathbf{L}^2(\Omega) \times \mathbf{H}(\mathrm{div}_{4/3}; \Omega)$ (see, Remark 7.4.3 for further details).

The specific problem conditions

In particular, we consider the unit cube $\Omega = (0,1)^3$ as the domain in $\mathbb{R}^3$. In addition, we consider a nonzero boundary condition $u_D = (1,1,0)$ on the bottom face of the cube $\partial\Omega_{\mathrm{bottom}} = (0,1) \times (0,1) \times \{0\}$, and zero on the rest of the faces. In addition we consider $\varphi_D = \exp(4(-(x_1 - 0.5)^2 - (x_2 - 0.5)^2))$ on $\partial\Omega_{\mathrm{bottom}}$ and zero otherwise. For simplicity, approximate the solution to the problem using the parametric coefficients a_1 and a_2 given by (7.4.1) and (7.4.2) respectively. See Fig. 7.4 for an example of the solution $(\boldsymbol{u}, p, \varphi)(\boldsymbol{y})$ for a given $\boldsymbol{y} \in [-1,1]^4$.

7.5 Numerical results

We now present numerical result for the three parametric PDEs defined in the previous section.

Parametric affine diffusion equation

First we introduce further details about the experiments and figures below:

i) In Figures 7.5–7.7 we show the average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation errors versus the number of samples m for various DNN architectures solving the Hilbert-valued diffusion equation in (7.4.4)–(7.4.5). The testing involved computing the geometric mean performance over 12 trials, represented by the colored lines, and the (geometric) standard deviation, shown in shaded colors. We refer to [12, Appendix A.1.3] for more details. We also represent in all the figures the decay rate m^{-1} for comparison.

ii) To approximate the Bochner norms, we used a sparse grid rule of level 5 with 1105 points when $d = 4$ and level 4 with 3937 points when $d = 8$ to compute the relative testing error. Thus, the figures are divided into two parts: the **(top)** part displays results for $d = 4$, while the **(bottom)** part shows results when $d = 8$. Additionally, the left side of each figure presents results for the parametric coefficient a_1 in (7.4.1), while the right side shows results for a_2 in (7.4.2).

In Fig. 7.5, we observe that wider DNNs generally outperform narrower ones when $d = 4$ dimensions. Specifically, all ELU DNN architectures achieve relative errors closer to two digits of accuracy for the affine coefficient, and for the coefficient a_2 when $d = 4$. Also in all four plots the 10×100 ELU DNN shows a decay rate close to m^{-1} for $m < 100$ samples. Despite this, all architectures present a worse decay rate as the training sample size increases. We also note a relatively small variance across the 12 trials. Additionally, when comparing the ELU DNN, we notice a decrease in the performance as the problem dimension increases from $d = 4$ to $d = 8$ regardless of the parametric coefficient.

Figure 7.6 employs the same parameters as Figure 7.5, but using a ReLU DNN architecture. Here, we observe a better decay rate on average (solid line) after $m = 300$ points compared to that of the ELU activation function. In addition, all ReLU DNNs achieve an error below 10^{-1} at some point, but in average these rates do not go beyond 10^{-2}.

In contrast, Figure 7.7 reveals that deeper and wider 10×100 tanh DNNs perform worse than their ReLU or ELU counterparts. Additionally, a larger standard deviation in the error after 100 training points for wider DNNs. In average shallow tanh DNNs are similar to ELU newtworks in this case, but with a larger standard deviation in the error.

Parametric NSB equations

Here, we present the average relative $L^2_\varrho([-1,1]^d, \mathcal{V})$ approximation errors versus the number of samples m for various DNN architectures solving the Banach-valued NSB problem in (7.4.10). In particular, we focus on the approximation of $\boldsymbol{u}(\boldsymbol{y}) \in \mathbf{L}^4(\Omega)$ in Figures 7.8–7.10 and the approximation of $p \in \mathrm{L}^2_0(\Omega)$ in Figures 7.11–7.13. As in the previous case, testing involved computing the average performance over 12 trials. We used the same sparse grid rule, level 5 with 1105 points when $d = 4$ and level 4 with 3937 points when $d = 8$. The figures are organized in the same way as before.

Regarding the approximation of $\boldsymbol{u}$, when comparing Figure 7.8 to the rates in Figure 7.5 for the diffusion problem, we observe a similar behaviour between their corresponding architectures. In the case where $d = 4$, increasing the number of weight and biases achieves better performance when using ELU activation functions and more training points. In contrast, when $d = 8$ dimensions for the affine coefficient using larger DNNs does not improve the performance. It is interesting that the ELU 10×100 architecture tends to follow a decay rate similar to m^{-1} in some cases, e.g., the log-transformed diffusion with $d = 4$. In contrast, smaller DNNs show a faster rate up to 100 sample points.

The situation changes in Figure 7.9, where the performance of the DNNs deteriorates compared to Figure 7.8. The decay rate now is slower than using ELU DNNs.

In Figure 7.10, we notice that the error for a 10×100 DNN in general follows a straight line but at a worse decay rate than the other architectures. In general, the performance of

smaller DNNs seems to be better than the larger ones but with a larger standard deviation in the error.

The results for p are comparable to those of u for similar architectures. We note that larger DNNs in Figure 7.11 perform better. Note that the approximation tends to follow the decay rate m^{-1} in Figure 7.11. As for Figure 7.12, it shows better decay rate in comparison to Figure 7.6 near 500 points. In addition, as shown in Figure 7.13, smaller DNNs tend to have a smaller standard deviation in the error.

Parametric Boussineq equations

As in the previous example, we need to introduce additional details about the experiments and figures for the Boussinesq problem:

i) In Figures 7.14–7.22, we present the average relative $L^2_\varrho([-1,1]^d, \mathcal{V})$ approximation errors versus the number of samples m for various DNN architectures solving the Banach-valued Boussinesq problem in (7.4.13). In contrast to the Poisson and NSB problems, here we explore the capabilities of DNNs over 8 trials. Testing involved computing the average performance over all trials. We employ a smaller number of trials than before due to the increased complexity of this problem, which requires more computational resources. We also include the decay rate m^{-1} in all figures for comparison.

ii) To approximate the Bochner norms, we also used a sparse grid rule of level 5 with 1105 points when $d = 4$, but a level 3 with 849 points when $d = 8$ to compute the relative testing error. The figures are divided into the same sections as in the previous cases.

Upon examining Fig. 7.14, it becomes apparent that for the affine coefficient, the performance of the DNN improves as we increase the number of samples m until 100 training points for both $d = 4$ and $d = 8$. After this point the testing error appears to ceases to decrease. In contrast, when using the coefficient a_2, the DNN is still able to decrease the testing error, albeit at a slightly lower rate than before after reaching 100 training points.

In Fig. 7.15, we observe that the testing error decays at a slower rate, reaching approximately $2 \cdot 10^{-2}$ near 500 points.

Fig. 7.16 reveals that the shallow tanh DNNs exhibit a similar performance near 500 training points as the ELU DNNs, but worse performance for smaller numbers of training points. However, similar to the previous two problems, the 10×100 architecture does not perform well as the number of samples increases.

The pattern persists for the remaining approximations. Notably, in Fig. 7.20, the ELU DNN approximating the pressure p closely follows the approximation rate of m^{-1} for the affine coefficient.

7.6 Additional discussion

These experiments consistently highlight the reliability of ELU DNNs, which generally exhibit a smaller standard deviation from the mean error as we increase the number of training samples m. This makes them a favorable choice in many scenarios, offering a balance between performance and sample complexity. However, it is crucial to note that deeper and wider DNNs are not universally superior, as they can lead to diminished performance or larger standard deviations. This becomes particularly evident when using the tanh activation function. The results also underscore the tanh DNNs, despite their potential to achieve low generalization error as shown in Fig. 7.22. This suggests that while tanh DNNs can be effective in certain contexts, their performance may be less consistent compared to ELU DNNs.

Furthermore, the experiments do not provide significant evidence to support the notion that DNNs perform better when approximating functions in Hilbert spaces compared to Banach spaces. This suggests that the choice of function space may not have a substantial impact on the performance of DNNs in function approximation tasks. This highlights a clear gap between theory and practice and keeps the door open for research to try to achieve better approximation rates for the Banach-valued case, as the theory still claims a worse decay rate than the Hilbert case.

Regarding the practical side approximating complex parametric PDE problems with DNNs, the results suggest starting with ELU DNNs and then considering ReLU DNNs, with the caveat that ReLU DNNs tend to have a larger standard deviation than ELU DNNs. Additionally, the findings caution against using larger and wider tanh DNNs in approximating parametric PDEs, as their behaviour shows to be worst than smaller architectures as the number of samples increases, unlike for ReLUs or ELUs.

7.7 Conclusions

In this chapter, we investigated the practical approximation capabilities of different DNN architectures for three parametric PDEs. We tested 4×40, 5×50, and 10×100 DNNs architectures in combination with ReLU, tanh, and ELU activation functions. We applied these DNN archtiectures to learn the FE coefficients of solution maps for three problems: the Poisson problem in a 2D physical domain, a Navier-Stokes-Brinkman problem in a 2D physical domain, and a Boussinesq problem in a 3D physical domain.

Keeping this in mind and the discussion in §7.6, we answer Question 6 in the affirmative and Question 7 of §1.6 in the negative.

Answer to Question 6

It is possible to efficiently apply DL to learn smooth parametric Banach-valued functions using standard architectures and training procedures to approximate parametric PDEs from limited samples

Answer to Question 7

There is no preliminary empirical evidence indicating that learning the FEM coefficients of a Banach-valued solution of a parametric DE exhibits a decay rate in terms of the number of samples that is worse than in the Hilbert-valued case.

In summary, while DNNs demonstrate the capability to approximate complex parametric PDEs and exhibit a decay rate in the error as the number of training samples increases, there are limitations to this rate of improvement. Increasing the dimensionality of the problem tends to increase the error on average, but it does not significantly affect the rate of decay. Notably, simpler functions with fewer degrees of freedom, such as the pressure or temperature in the Boussinesq problem, are learned more effectively by DNNs compared to more complex functions like the velocity field $\boldsymbol{u}$. These observations provide valuable insights into the behaviour and performance of DNNs in function approximation tasks, highlighting the importance of thoughtful architecture selection and careful consideration of problem complexity and number of coefficients to be approximated.

7.8 Future work

There are several interesting directions for future research.

- First, it is important to investigate the impact of other activation functions such as Rectified Power Unit (RePU), Leaky Rectified Linear Unit (Leaky ReLU), and Scaled Exponential Linear Unit (SELU) on the performance of DNN approximating parametric PDEs. These activation functions may offer distinct advantages in terms of learning dynamics and model performance, and their comparative analysis could reveal which functions are most effective under different conditions, such as different boundary conditions. Moreover, additional experiments are needed to explore different DNN architectures by varying the widths and depths of the DNNs. For instance, by keeping the width constant and incrementally increasing the depth, we can study how the DNN approximation capabilities evolve. This would be important to optimize the choice of parameters in the DNN.

- In this work, we exclusively utilized the He uniform initializer for weights and biases. It is essential to investigate the effects of different initialization methods on model

performance and convergence. By comparing various initialization techniques, we can gain valuable insights into its effect on the approximation capabilities of the DNN.

- In this chapter we used a maximum number of training points of $m_{\max} = 500$. It is interesting to investigate wheter increasing the number of samples will make the approximation better or if after a certain amount of samples the DNN is not capable of reduce the testing error.

- Here we only considered steady-state PDEs. It would be important to extend this work to time-dependet problems. This could involve training the model on dynamic datasets to evaluate its effectiveness in handling temporal variations. Additionally, explore more complex situations where boundary conditions significantly impact the solution, providing insights into the ability of the model to handle real-world, dynamic scenarios.

- Finally, it would be interesting to study the theoretical aspects of the different parametric PDEs shown in this chapter. Unfortunately, it is not clear how to show that the Navier-Stokes-Boussinesq problem and Boussinesq equations are $(\boldsymbol{b}, \varepsilon)$-holomorphic in a specific complex region. Showing holomorphic extensions of these problems in a Banach-valued setting may help to bridge the gap between theory and practice between Chapter 5 and Chapter 7.

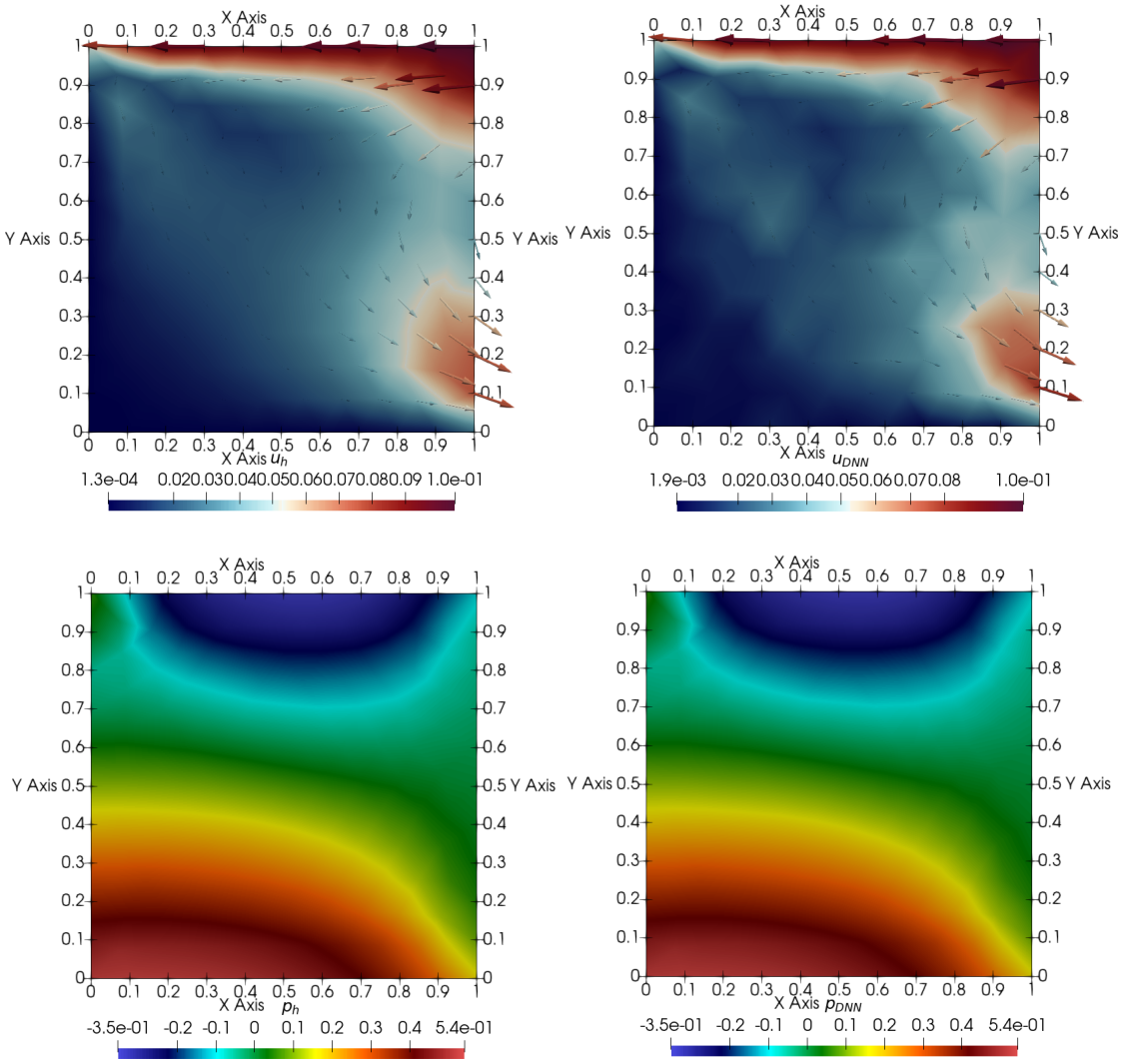

Figure 7.3: The figure shows the solution $(\boldsymbol{u}, p)(\boldsymbol{y})$ to the parametric NSB problem in (7.4.9) for a given parameter $\boldsymbol{y} = (1, 0, 0, 0)^\top$ with affine coefficient a_1, utilizing a total of 1464 degrees of freedom (DoF) for $\boldsymbol{u}$ and 244 DoF for p. The left column displays the solution given by the FEM solver, while the right column shows the 4×40 ELU DNN approximation after 60000 epochs of training with $m = 500$ sample points. The top displays the magnitude of the vector field $\boldsymbol{u}$ and its direction with white arrows. On the bottom side, we show the pressure p.

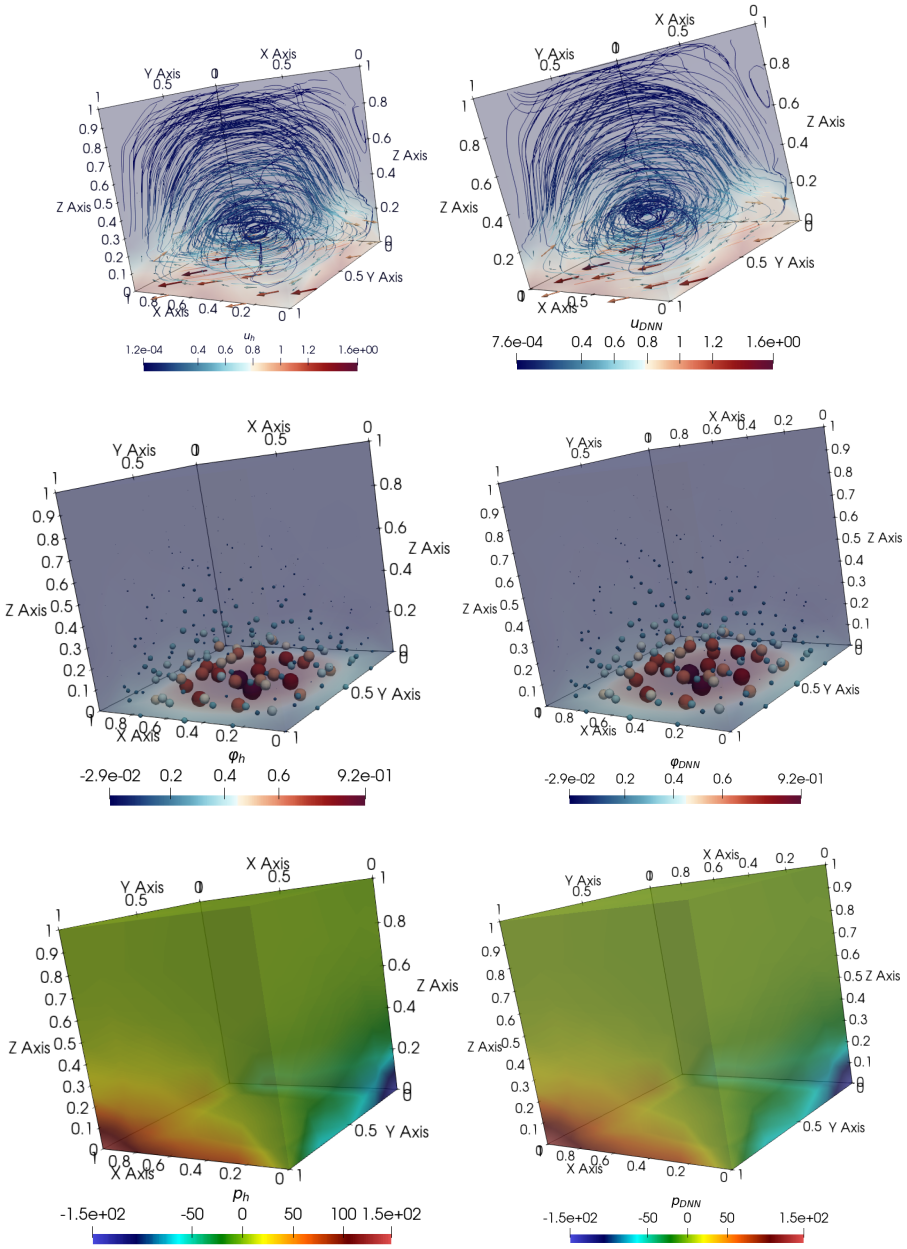

Figure 7.4: The figure shows the solution $(\boldsymbol{u}, \varphi, p)(\boldsymbol{y})$ to the parametric Boussinesq problem in (7.4.13) for a given parameter $\boldsymbol{y} = (1, 0, 0, 0)^\top$ with an affine coefficient a_1. The solution utilizes a total of 18480 degrees of freedom (DoF) for $\boldsymbol{u}$ and 528 DoF for both φ and p. The left column displays the solution given by the FEM solver, while the right column shows the 4×40 ELU DNN approximation after 60000 epochs of training with $m = 500$ sample points. The top row displays streamlines of the vector field $\boldsymbol{u}$ and their direction with colored arrows. In the middle row, we visualize the temperature distribution inside the cube using colored spheres, with the hottest region at the center of the cube. The bottom row illustrates the points of highest pressure p.

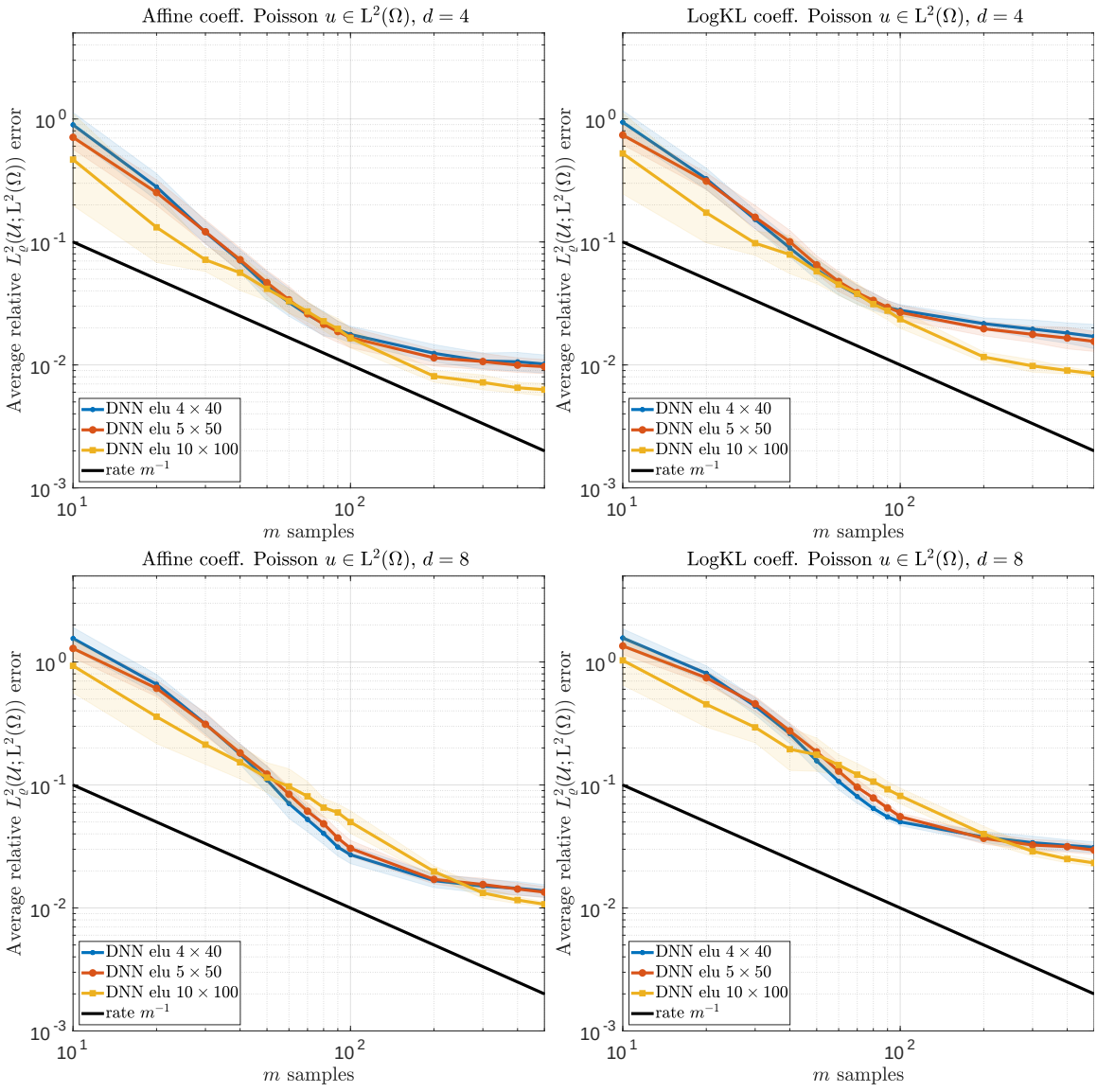

Figure 7.5: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ELU DNNs solving the Hilbert-valued diffusion equation in (7.4.4)–(7.4.5).

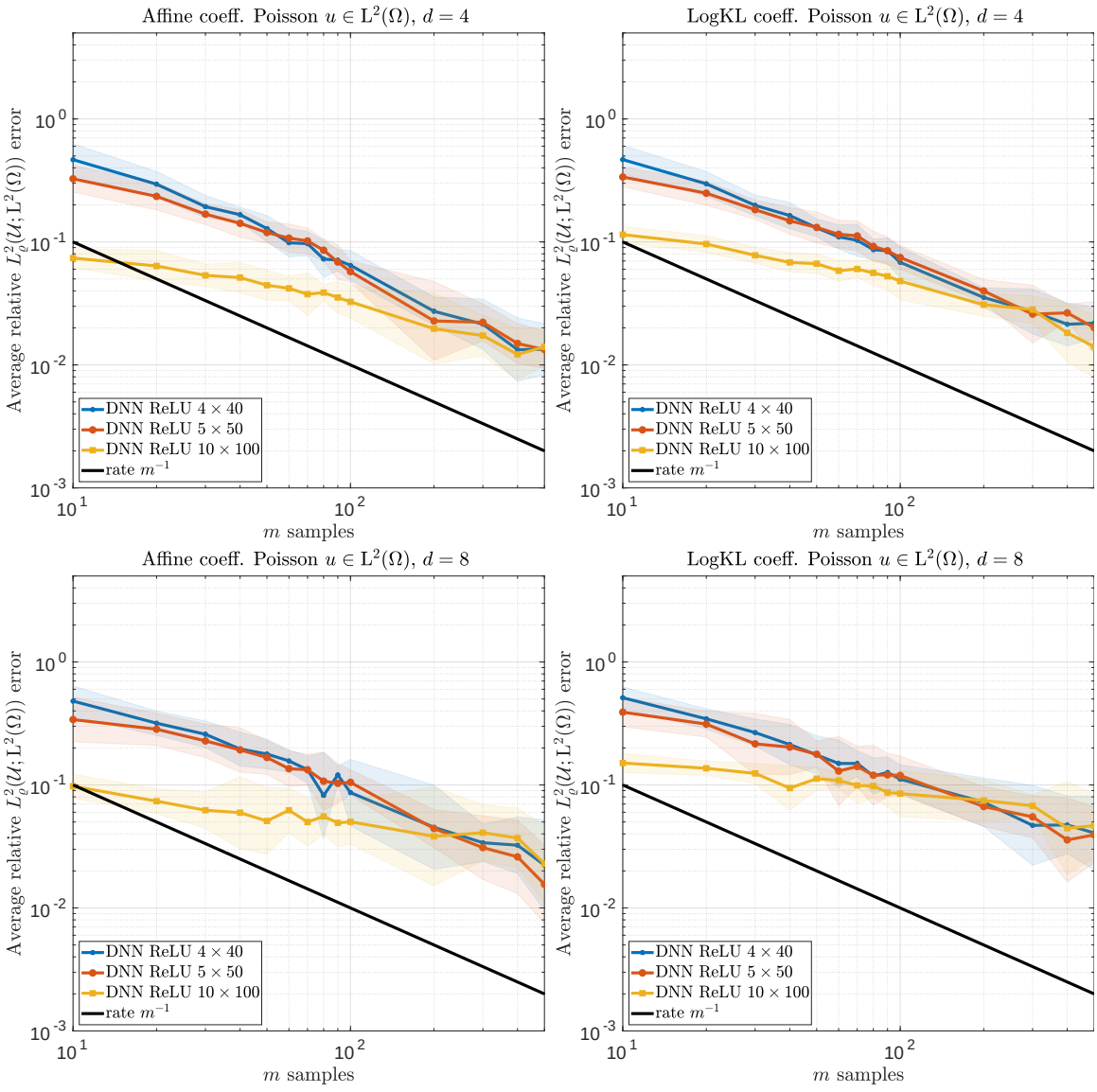

Figure 7.6: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ReLU DNNs solving the Hilbert-valued diffusion equation in (7.4.4)–(7.4.5).

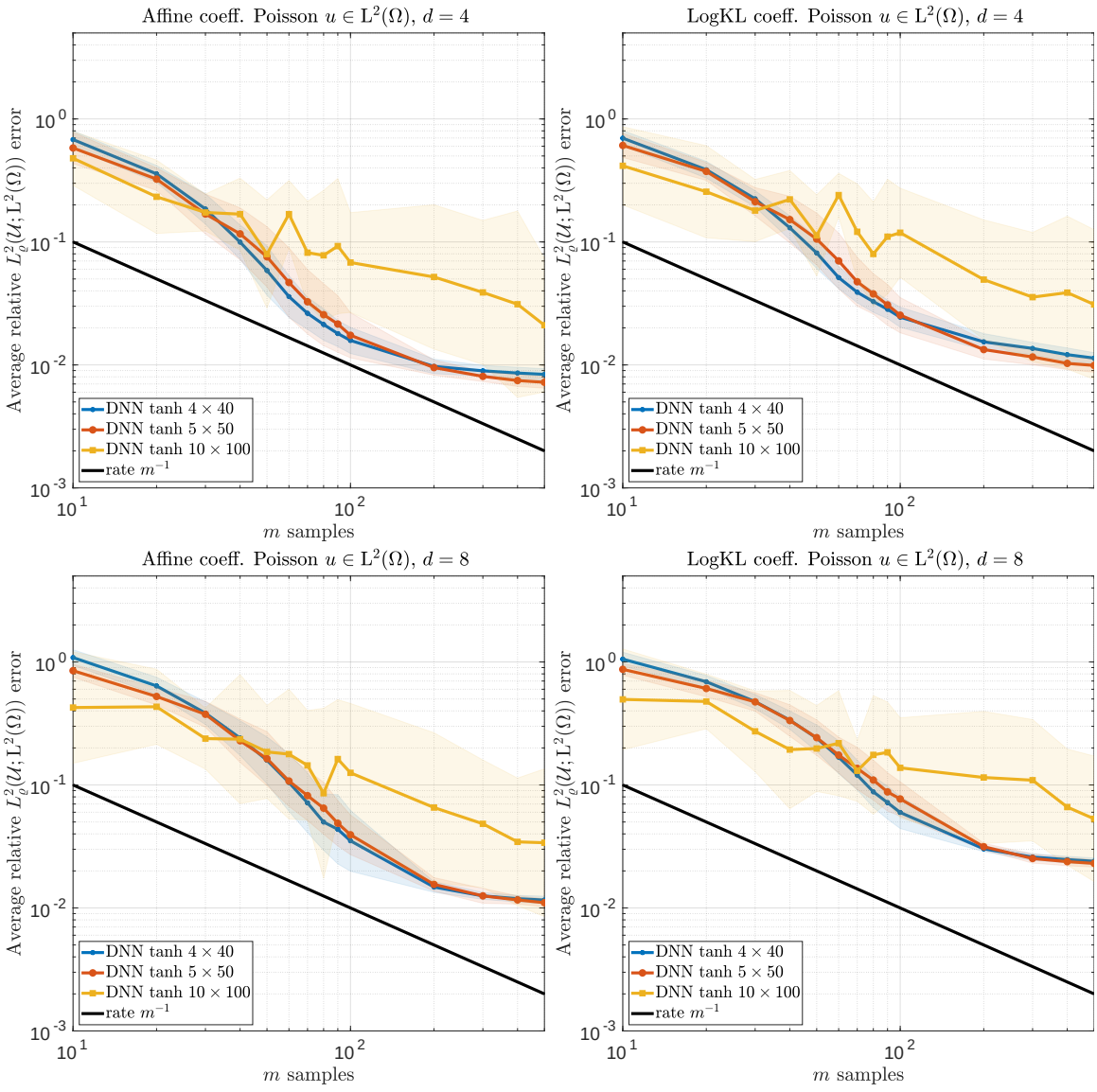

Figure 7.7: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for tanh DNNs solving the Hilbert-valued diffusion equation in (7.4.4)–(7.4.5).

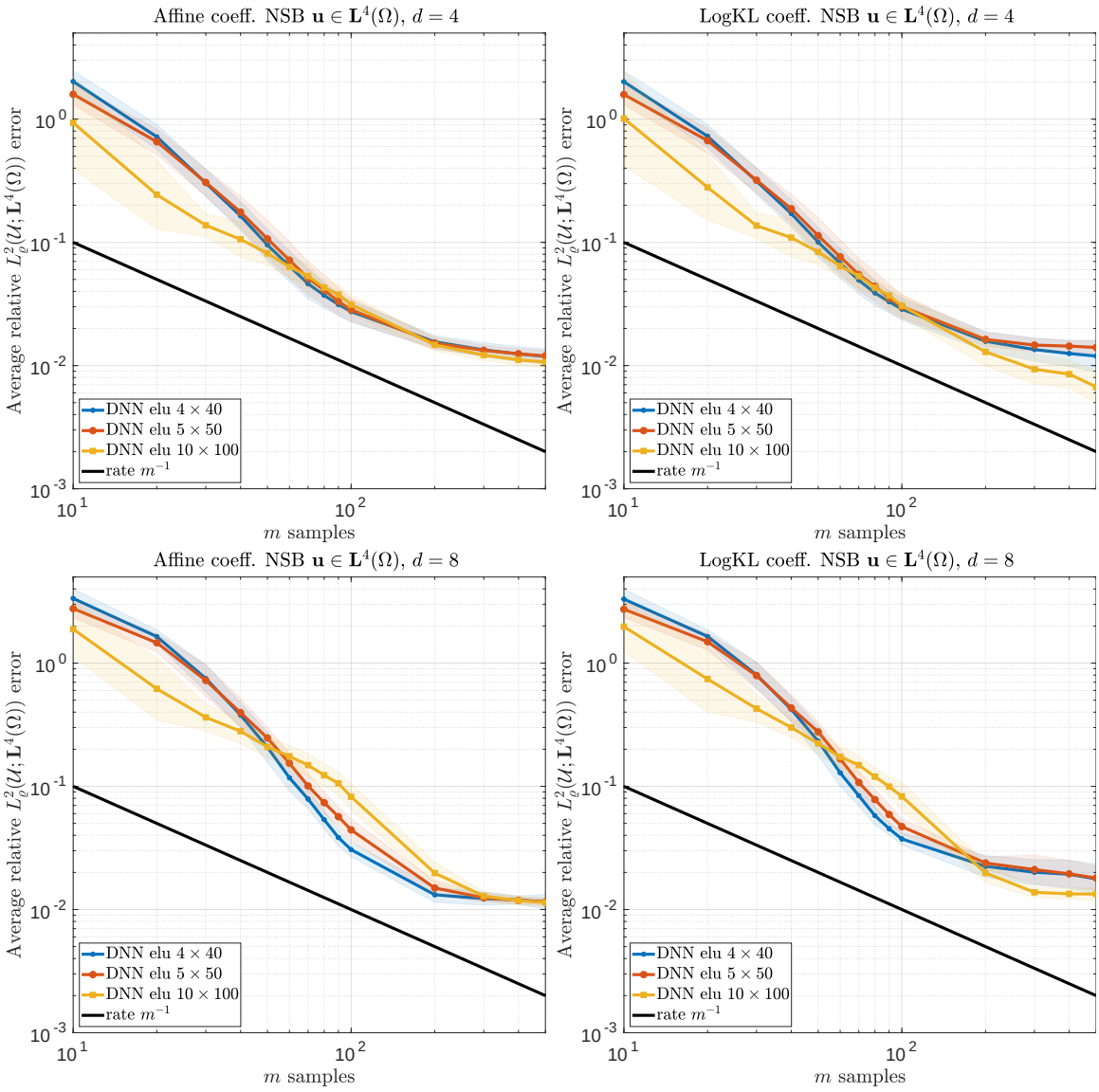

Figure 7.8: Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued NSB problem in (7.4.10).

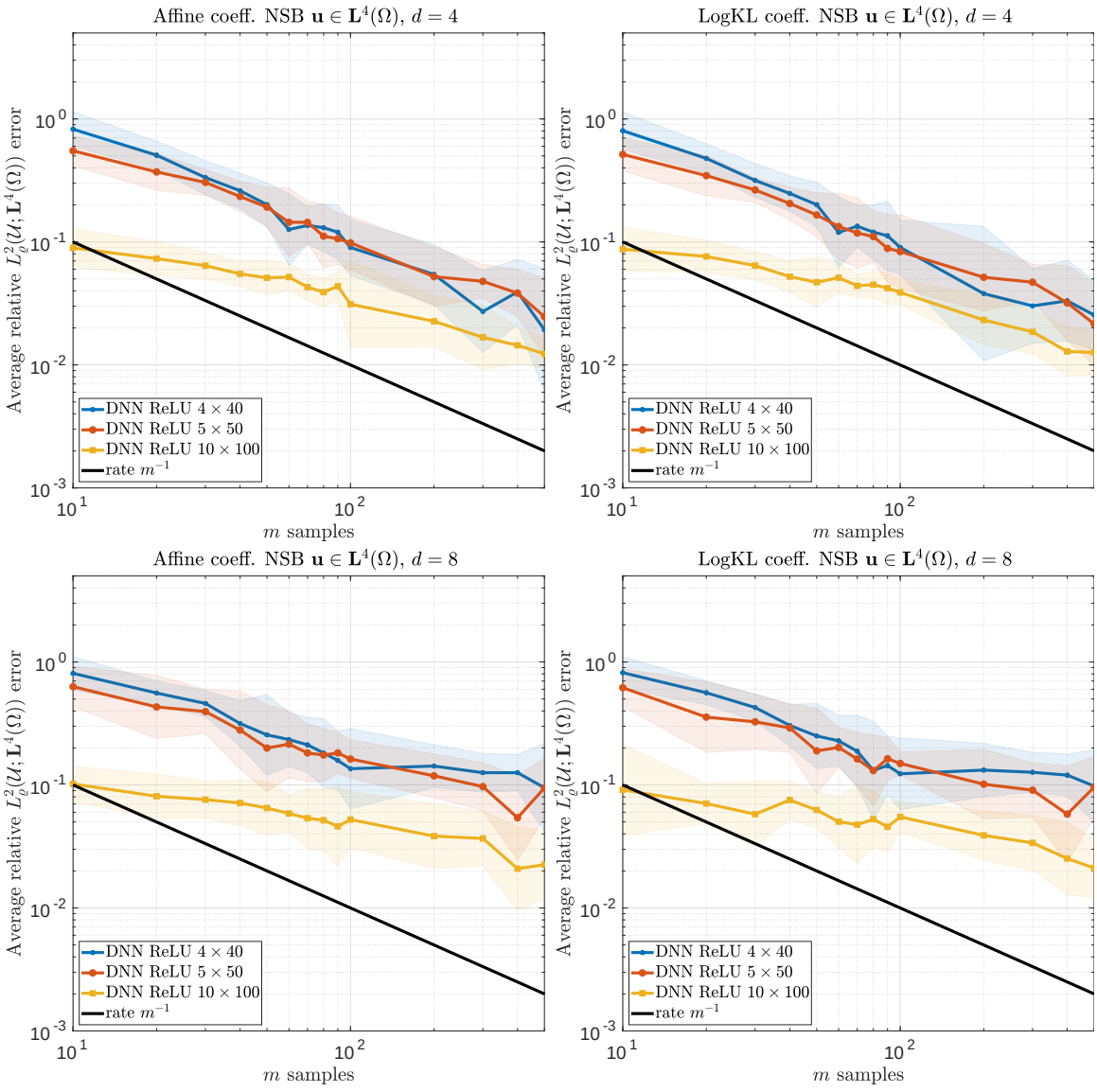

Figure 7.9: Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued NSB problem in (7.4.10).

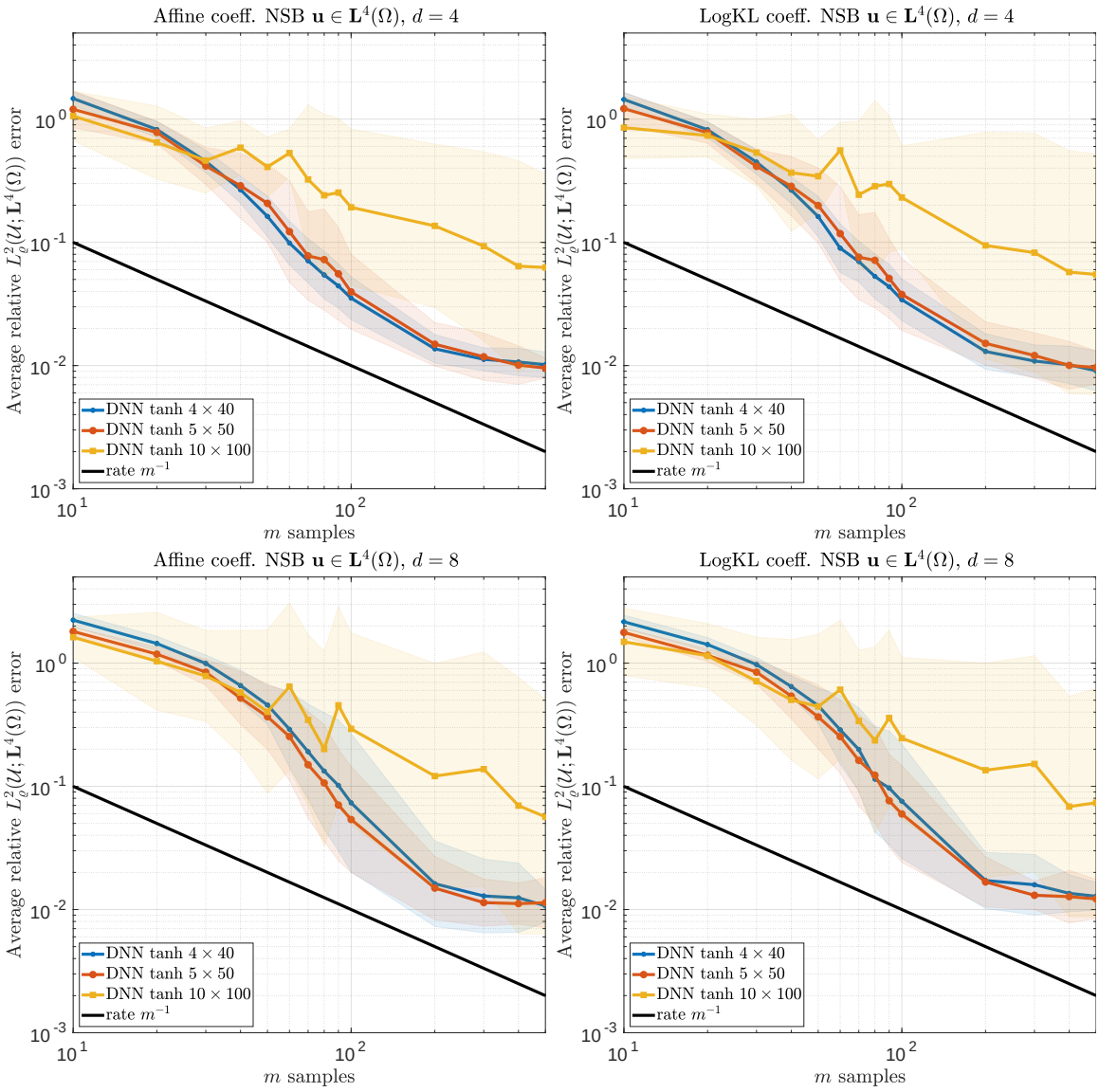

Figure 7.10: Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued NSB problem in (7.4.10).

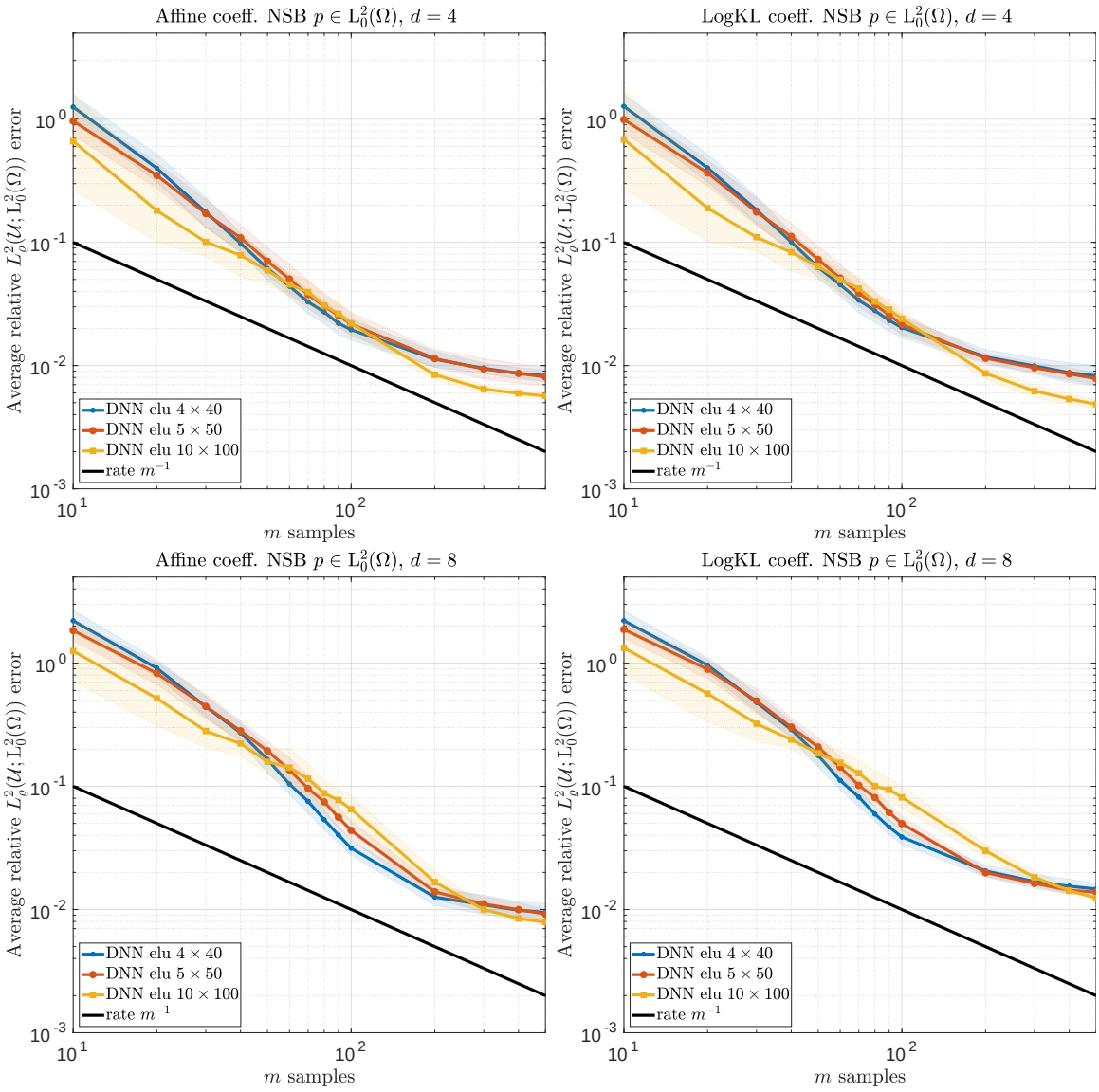

Figure 7.11: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued NSB problem in (7.4.10).

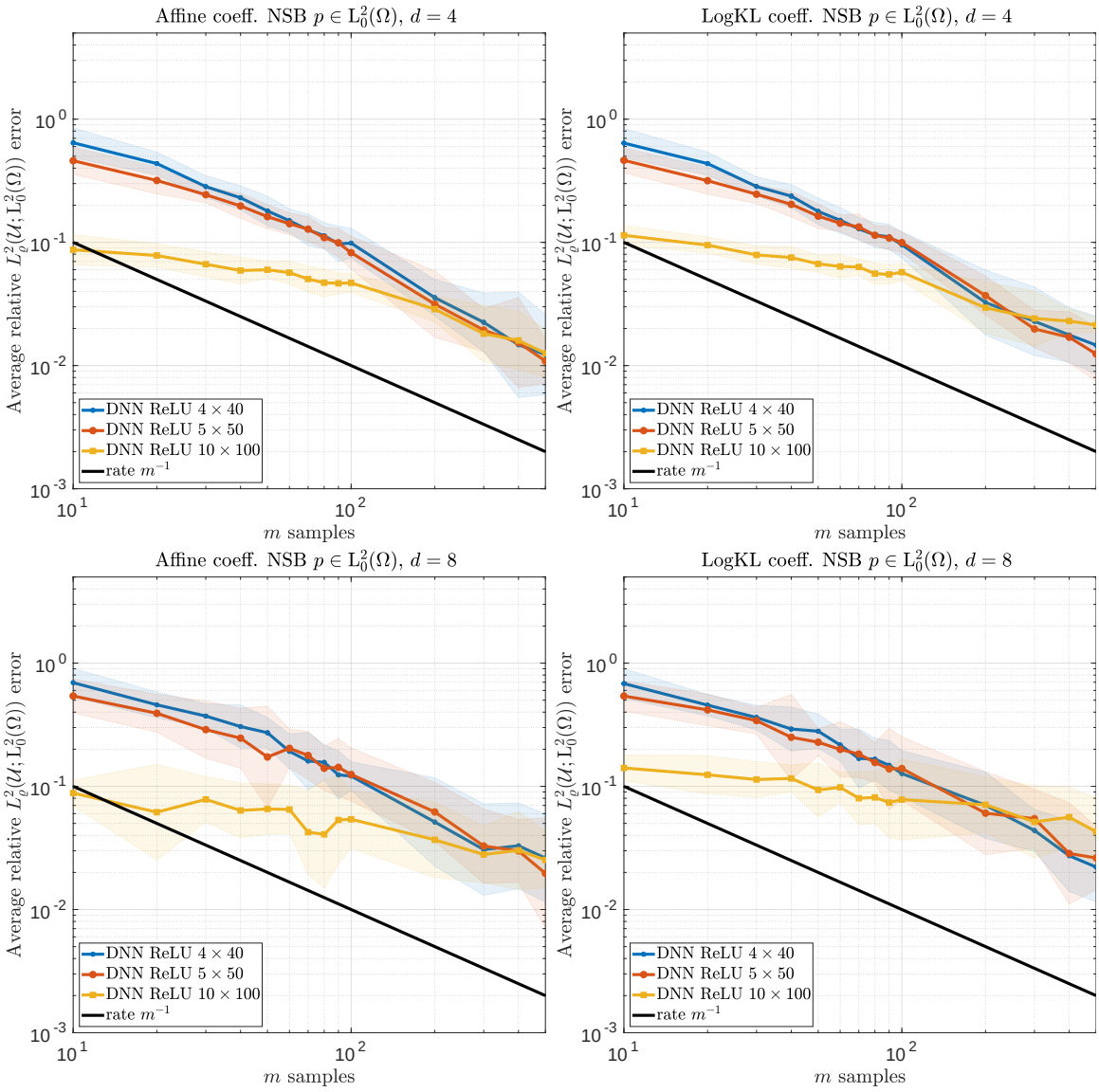

Figure 7.12: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued NSB problem in (7.4.10).

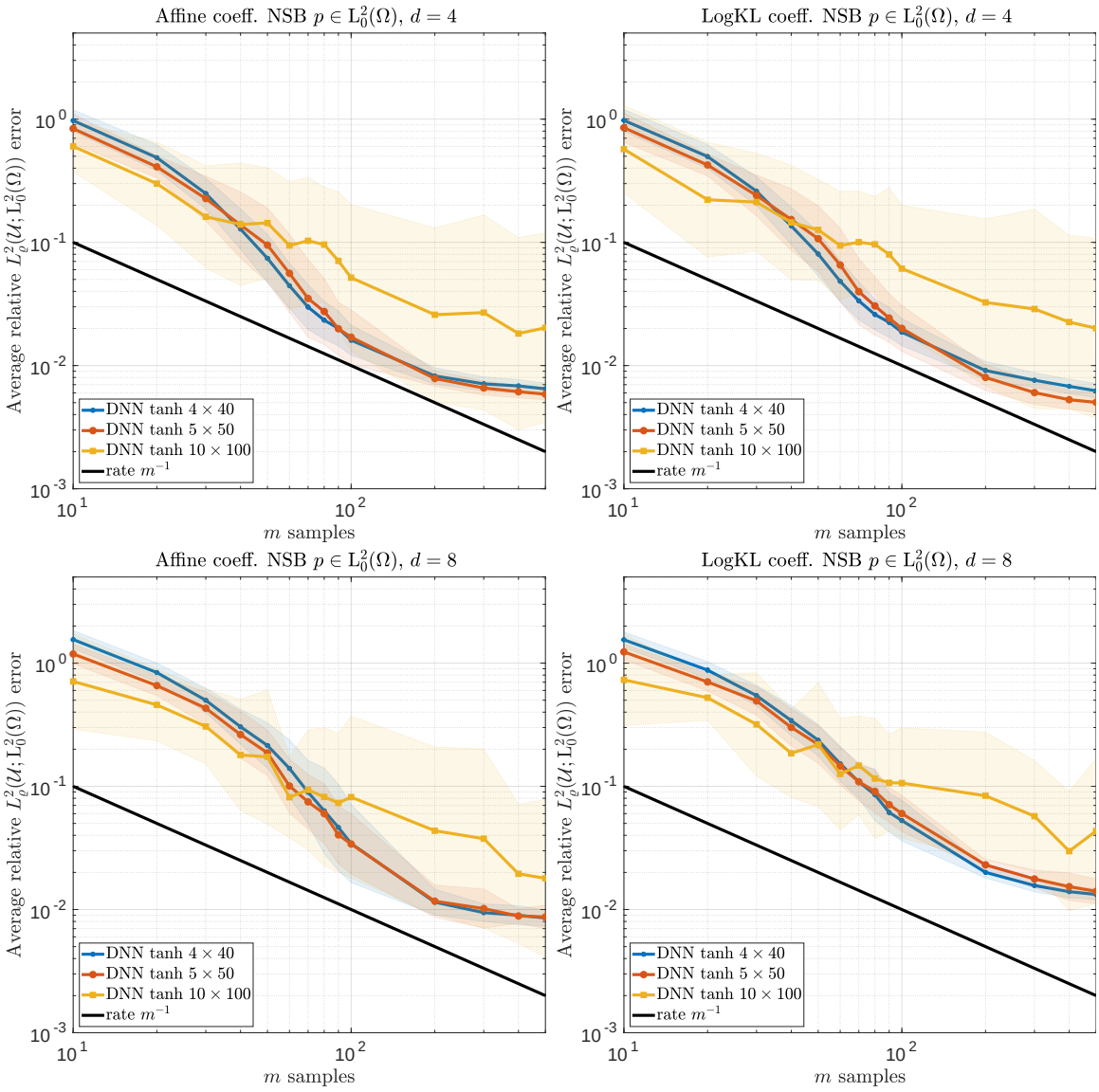

Figure 7.13: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued NSB problem in (7.4.10).

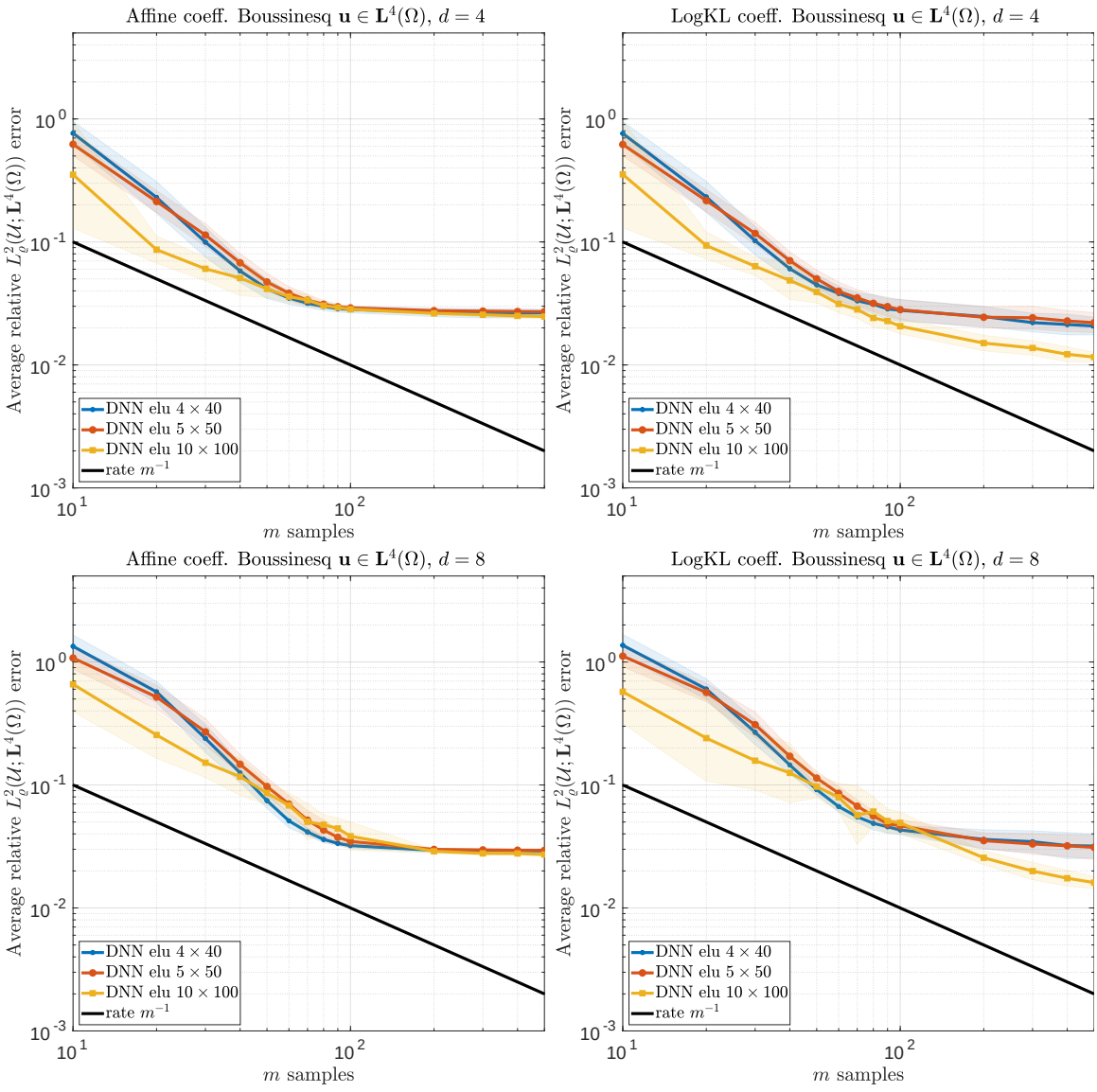

Figure 7.14: Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

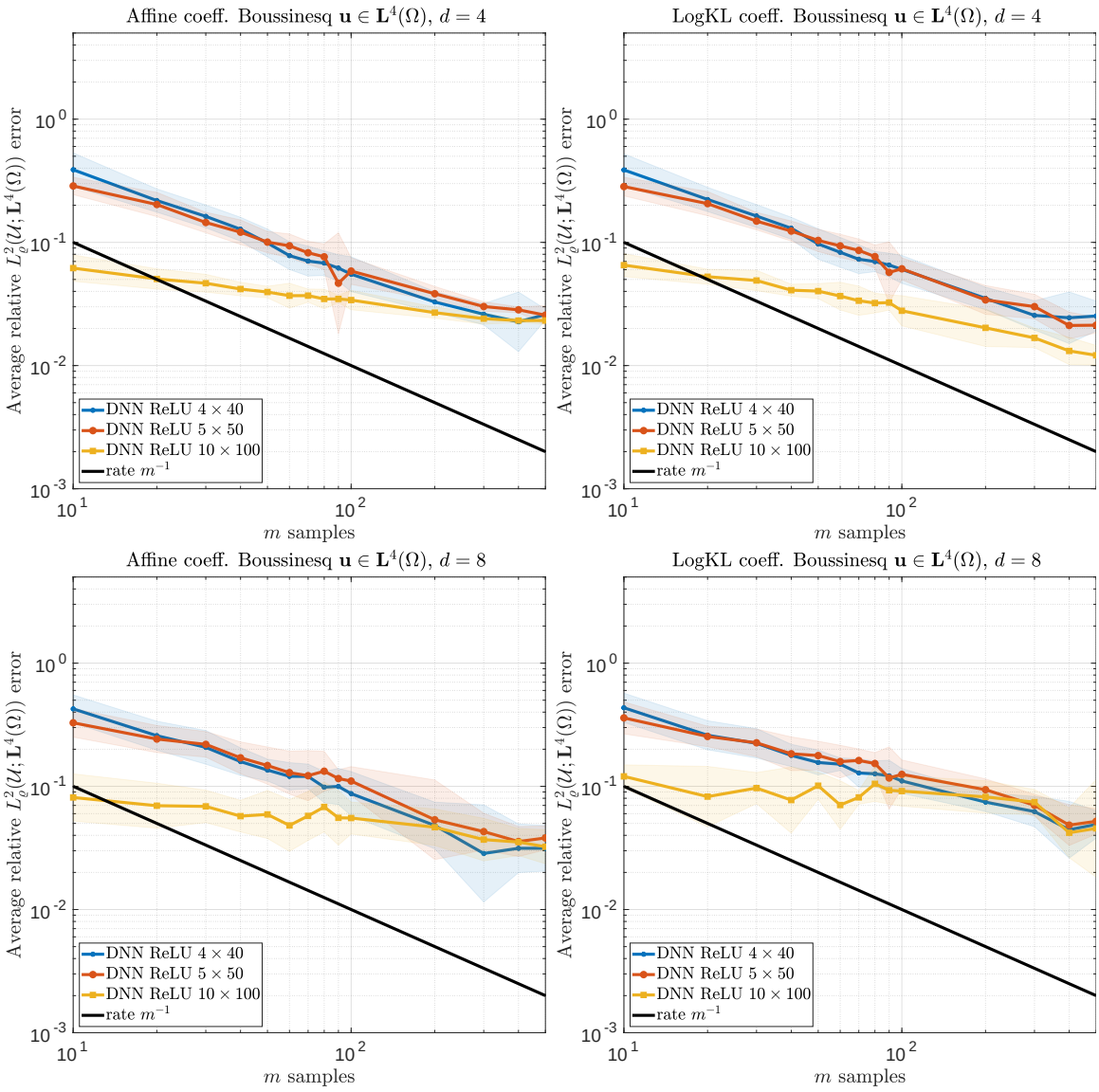

Figure 7.15: Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

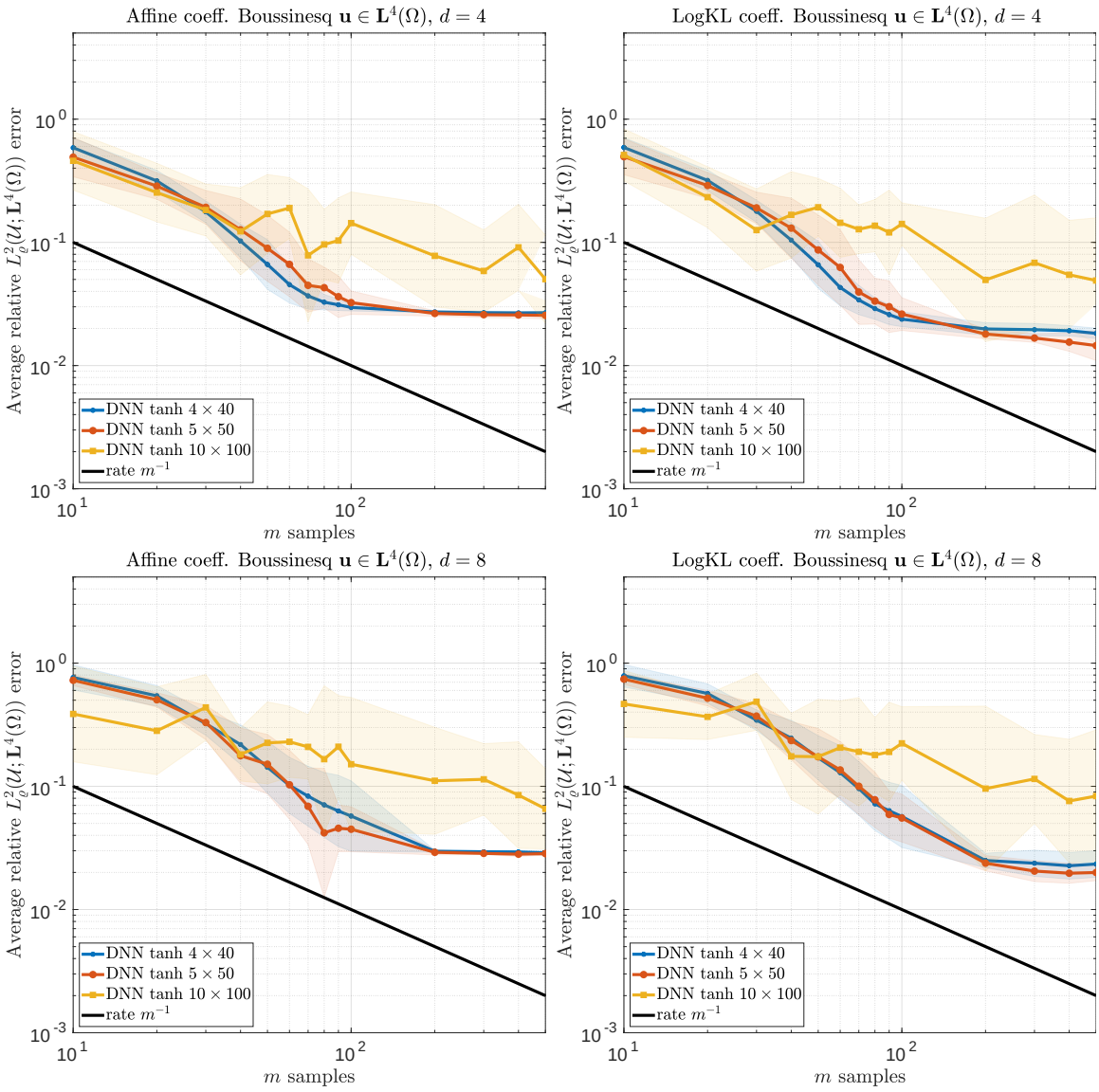

Figure 7.16: Average relative $L^2_\varrho([-1,1]^d, \mathbf{L}^4(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $\boldsymbol{u} \in \mathbf{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

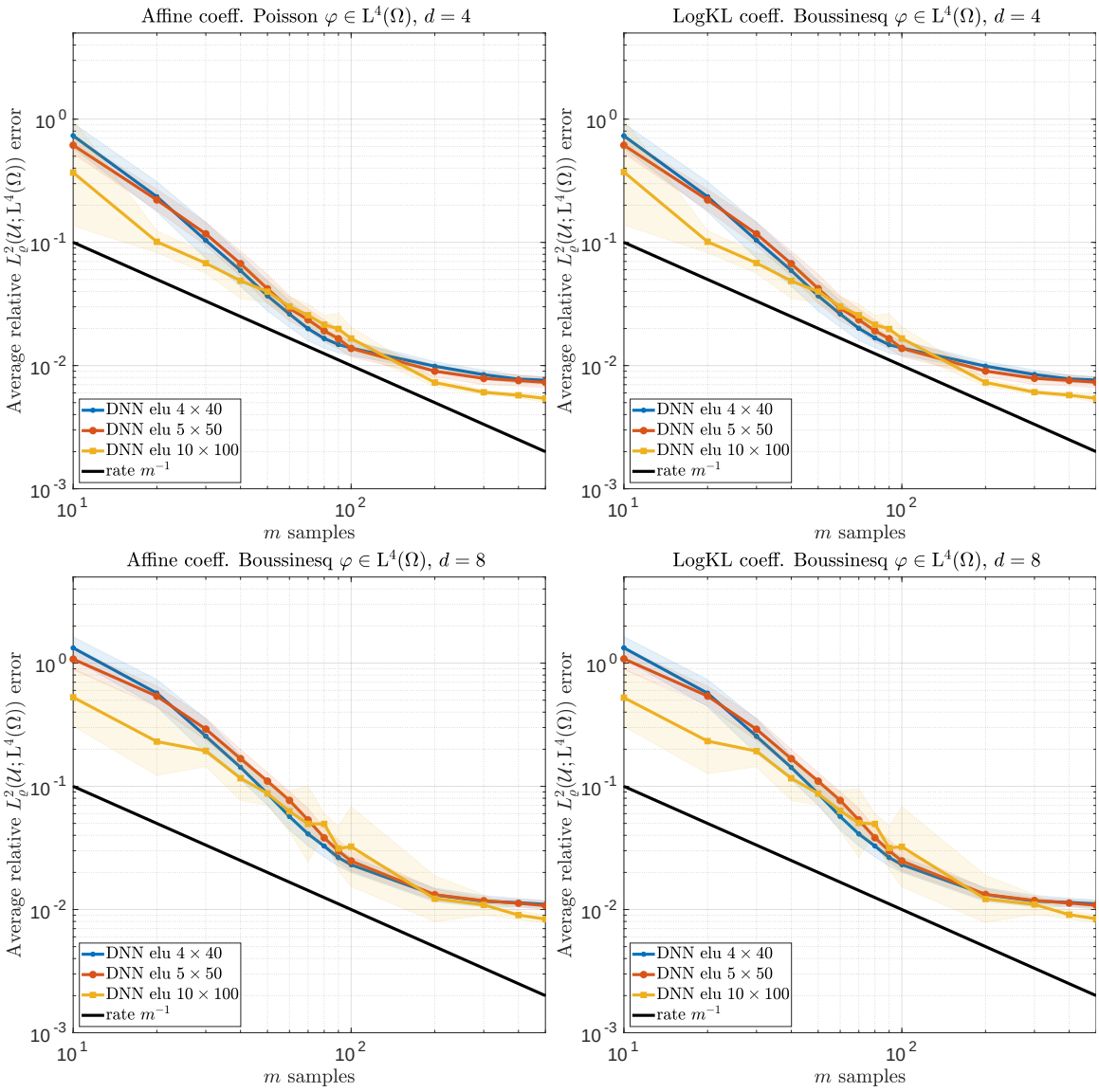

Figure 7.17: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^4(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $\varphi \in \mathrm{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

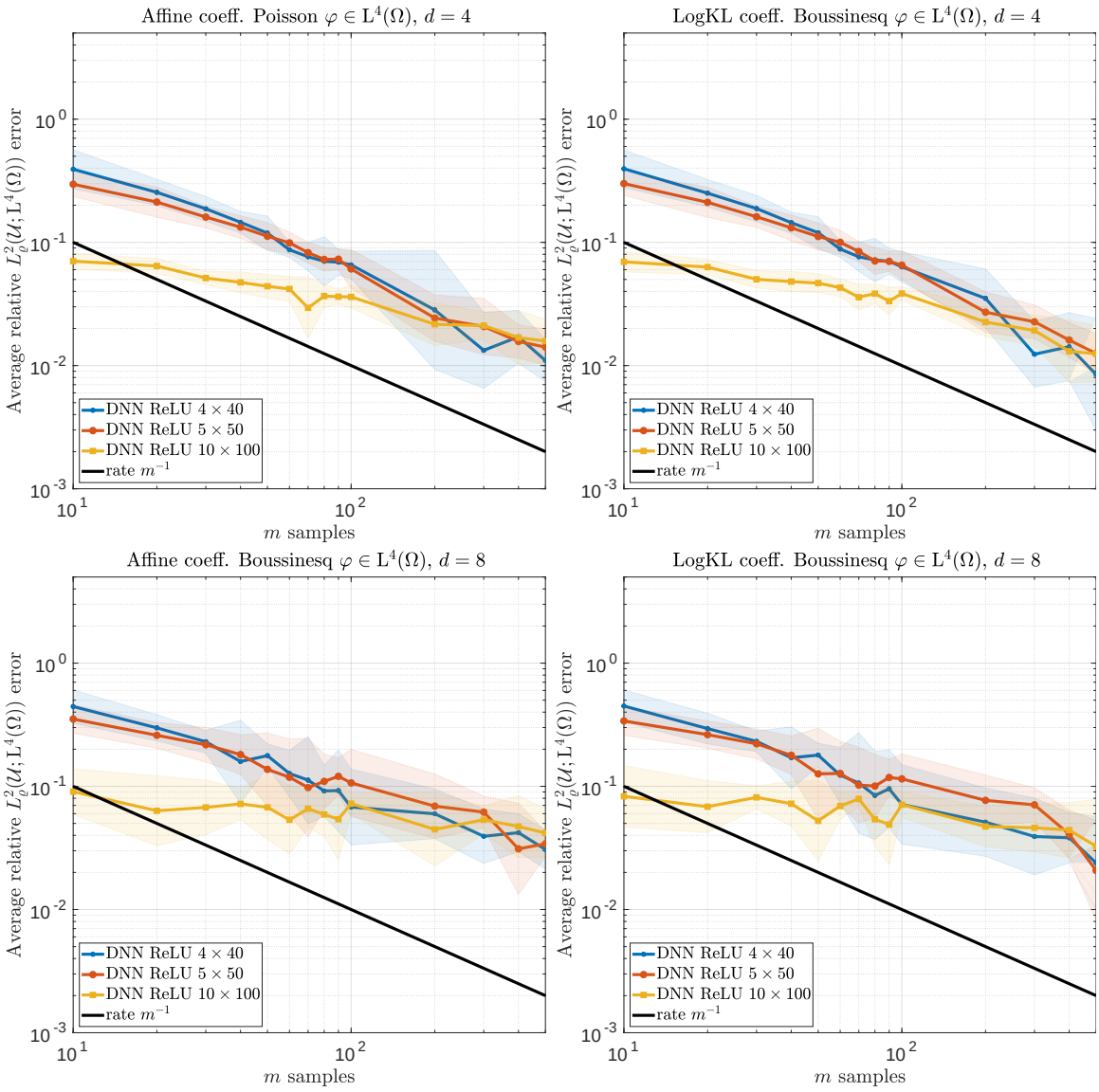

Figure 7.18: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^4(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $\varphi \in \mathrm{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

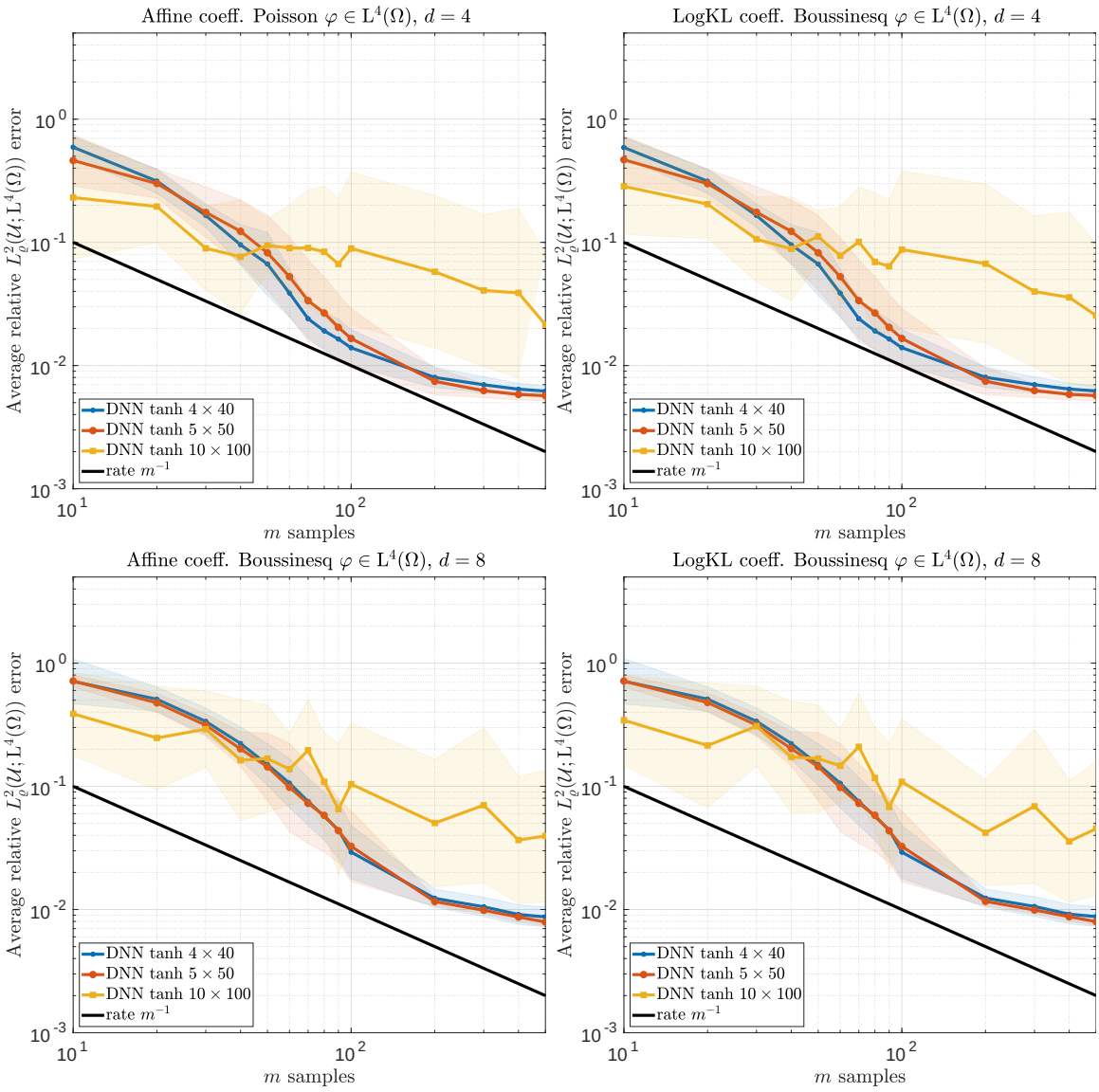

Figure 7.19: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^4(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $\varphi \in \mathrm{L}^4(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

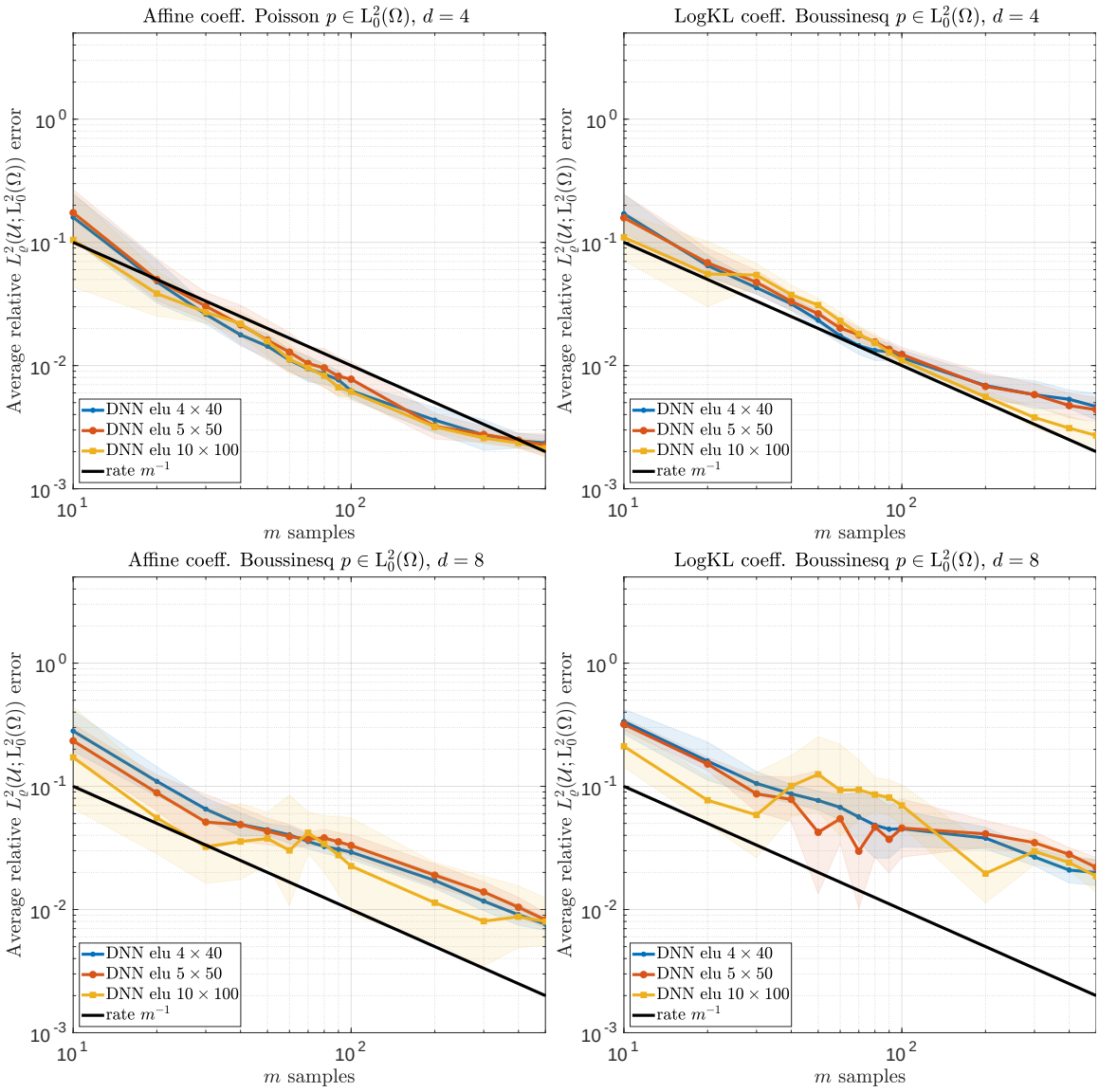

Figure 7.20: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ELU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

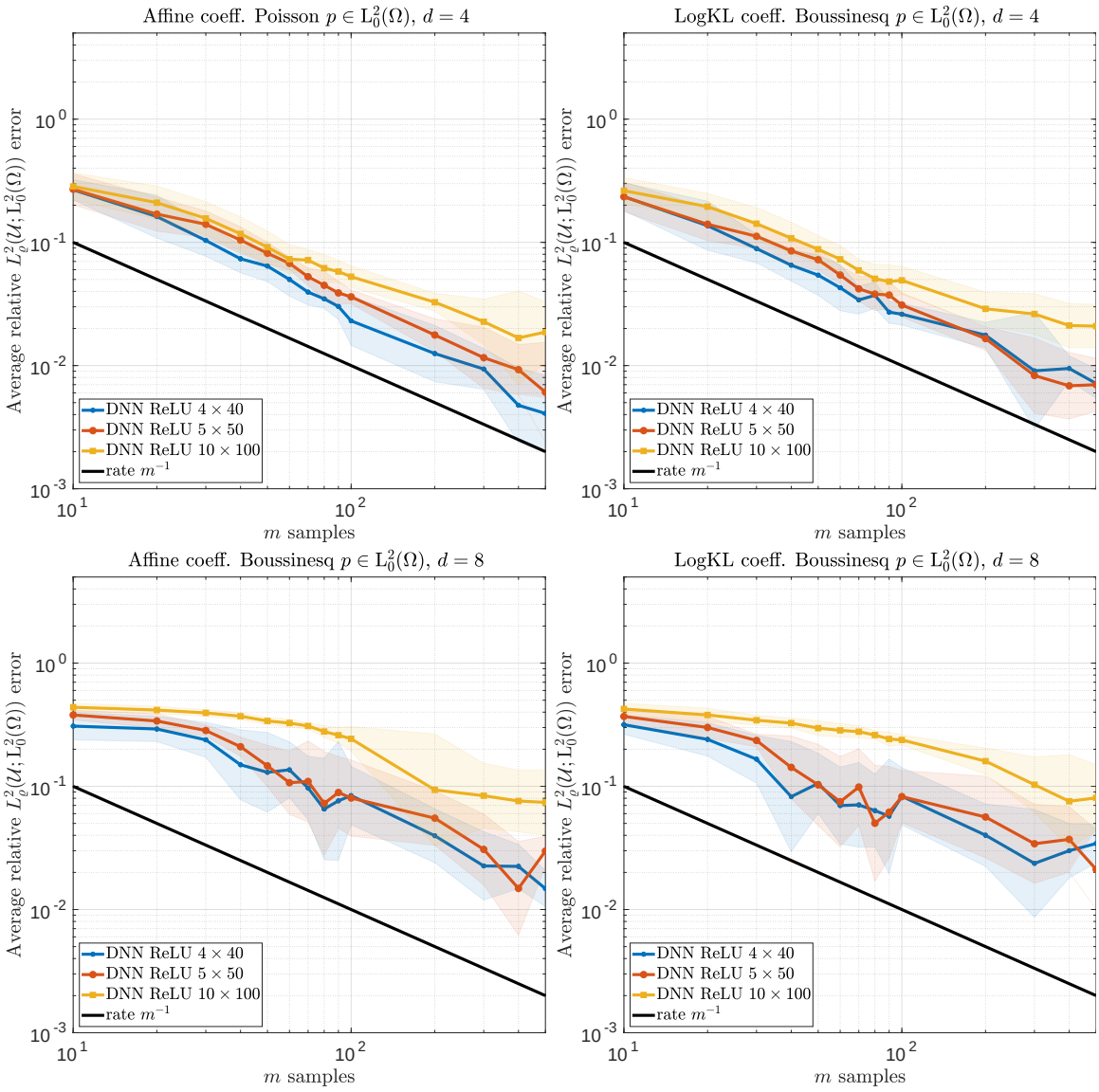

Figure 7.21: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for ReLU DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

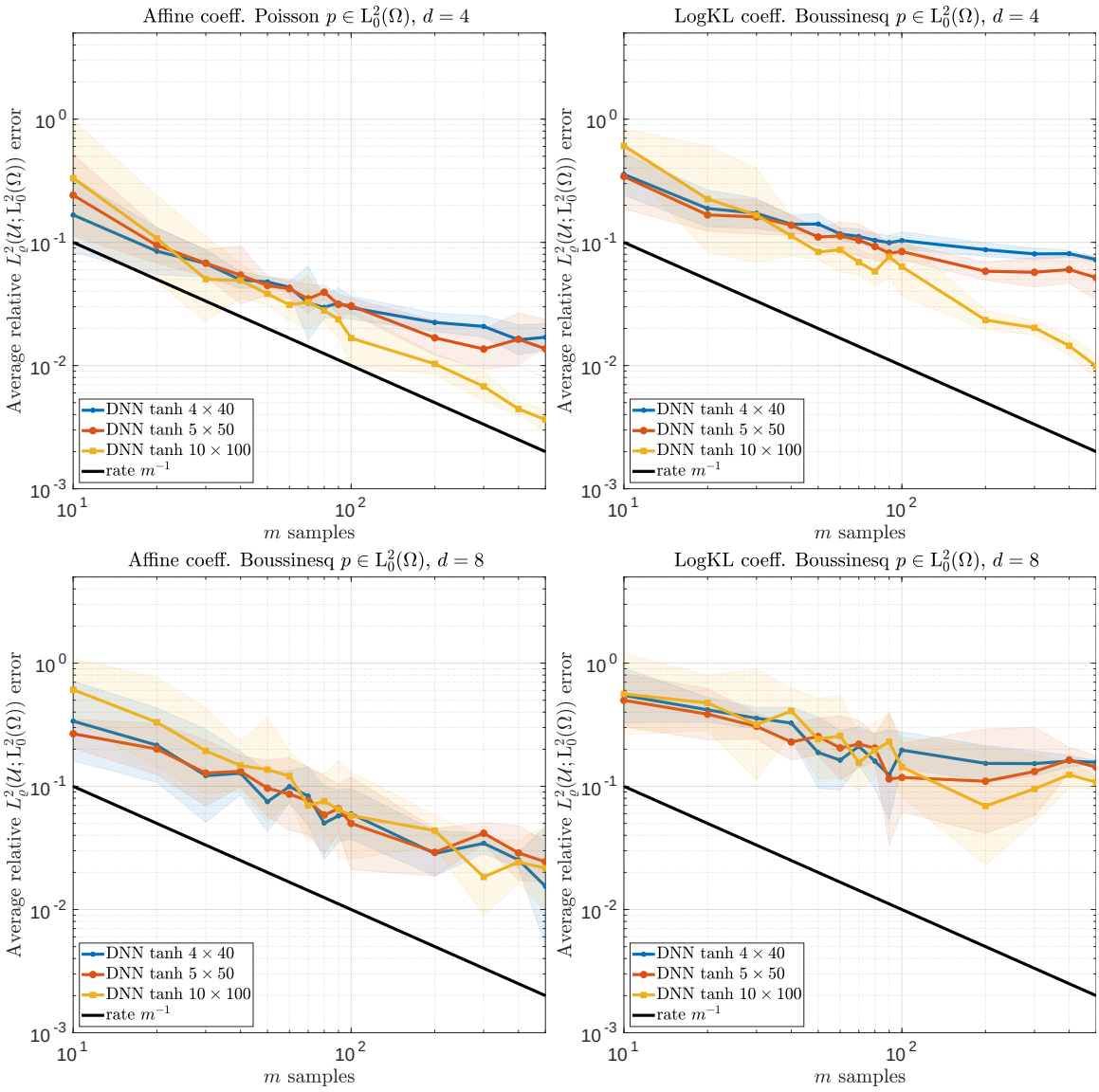

Figure 7.22: Average relative $L^2_\varrho([-1,1]^d, \mathrm{L}^2(\Omega))$ approximation error versus number of samples m for tanh DNNs approximating $p \in \mathrm{L}^2_0(\Omega)$ for the Banach-valued Boussinesq problem in (7.4.13).

Bibliography

[1] B. Adcock. Infinite-dimensional ℓ^1 minimization and function approximation from pointwise data. *Constr. Approx.*, 45(3):343–390, 2017.

[2] B. Adcock. Infinite-dimensional compressed sensing and function interpolation. *Found. Comput. Math.*, 18(3):661–701, 2018.

[3] B. Adcock, A. Bao, and S. Brugiapaglia. Correcting for unknown errors in sparse high-dimensional function approximation. *Numer. Math.*, 142(3):667–711, 2019.

[4] B. Adcock, A. Bao, J. D. Jakeman, and A. Narayan. Compressed sensing with sparse corruptions: fault-tolerant sparse collocation approximations. *SIAM/ASA J. Uncertain. Quantif.*, 6(4):1424–1453, 2018.

[5] B. Adcock and S. Brugiapaglia. Sparse approximation of multivariate functions from small datasets via weighted orthogonal matching pursuit. In S. Sherwin, D. Moxey, J. Peiró, P. Vincent, and C. Schwab, editors, *Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2018*, volume 134 of *Lect. Notes Comput. Sci. Eng.*, pages 611–621, Cham, Switzerland, 2020. Springer.

[6] B. Adcock and S. Brugiapaglia. Monte Carlo is a good sampling strategy for polynomial approximation in high dimensions. *arXiv:2208.09045*, 2023.

[7] B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. Deep neural networks are effective at learning high-dimensional Hilbert-valued functions from limited data. In J. Bruna, J. S. Hesthaven, and L. Zdeborová, editors, *Proceedings of The Second Annual Conference on Mathematical and Scientific Machine Learning*, volume 145 of *Proc. Mach. Learn. Res. (PMLR)*, pages 1–36. PMLR, 2021.

[8] B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. Near-optimal learning of Banach-valued, high-dimensional functions via deep neural networks. *arXiv:2211.12633*, 2023.

[9] B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. Learning smooth functions in high dimensions: from sparse polynomials to deep neural networks. *arXiv:2404.03761*, 2024.

[10] B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. On efficient algorithms for computing near-best polynomial approximations to high-dimensional, Hilbert-valued functions from limited samples. *Mem. Eur. Math. Soc. (In press)*, 2024.

[11] B. Adcock, S. Brugiapaglia, and M. King-Roskamp. Do log factors matter? On optimal wavelet approximation and the foundations of compressed sensing. *Found. Comput. Math.*, 22:99–159, 2022.

[12] B. Adcock, S. Brugiapaglia, and C. G. Webster. *Sparse Polynomial Approximation of High-Dimensional Functions*. Comput. Sci. Eng. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2022.

[13] B. Adcock and J. M. Cardenas. Near-optimal sampling strategies for multivariate function approximation on general domains. *SIAM J. Math. Data Sci.*, 2(3):607–630, 2020.

[14] B. Adcock, J. M. Cardenas, and N. Dexter. CAS4DL: Christoffel Adaptive Sampling for function approximation via Deep Learning. *Preprint*, 2022.

[15] B. Adcock, J. M. Cardenas, N. Dexter, and S. Moraga. *Towards optimal sampling for learning sparse approximation in high dimensions*, chapter 2, pages 9–77. Number 191 in Springer Optim. Appl. Springer, Cham, 2022.

[16] B. Adcock and N. Dexter. The gap between theory and practice in function approximation with deep neural networks. *SIAM J. Math. Data Sci.*, 3(2):624–655, 2021.

[17] B. Adcock, N. Dexter, and S. Moraga. Optimal approximation of infinite-dimensional holomorphic functions II: recovery from i.i.d. pointwise samples. *arXiv:2310.16940*, 2023.

[18] B. Adcock, N. Dexter, and S. Moraga. Optimal approximation of infinite-dimensional holomorphic functions. *Calcolo*, 61(1):12, 2024.

[19] B. Adcock and A. C. Hansen. *Compressive Imaging: Structure, Sampling, Learning*. Cambridge University Press, Cambridge, UK, 2021.

[20] B. Adcock and D. Huybrechs. Approximating smooth, multivariate functions on irregular domains. *Forum Math. Sigma*, 8:e26, 2020.

[21] B. Adcock and Y. Sui. Compressive Hermite interpolation: sparse, high-dimensional approximation from gradient-augmented measurements. *Constr. Approx.*, 50(1):167–207, 2019.

[22] N. Alemazkoor and H. Meidani. Divide and conquer: an incremental sparsity promoting compressive sampling approach for polynomial chaos expansions. *Comput. Methods Appl. Mech. Engrg.*, 318:937–956, 2017.

[23] N. Alemazkoor and H. Meidani. A near-optimal sampling strategy for sparse recovery of polynomial chaos expansions. *J. Comput. Phys.*, 371:137–151, 2018.

[24] K. Allali. A priori and a posteriori error estimates for Boussinesq equations. *Int. J. Numer. Anal. Model.*, 2:179–196, 2005.

[25] S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The FEniCS Project Version 1.5. *Archive of Numerical Software*, 3(100), 2015.

[26] T. Ando. Contractive projections in L_p spaces. *Pacific J. Math.*, 17:391–405, 1966.

[27] R. Aylwin, C. Jerez-Hanckes, C. Schwab, and J. Zech. Domain uncertainty quantification in computational electromagnetics. *SIAM/ASA J. Uncertain. Quantif.*, 8(1):301–341, 2020.

[28] M. Bachmayr and A. Cohen. Kolmogorov widths and low-rank approximations of parametric elliptic PDEs. *Math. Comput.*, 86(304):701–724, 2016.

[29] M. Bachmayr, A. Cohen, R. DeVore, and G. Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part II: lognormal coefficients. *ESAIM. Math. Model. Numer. Anal.*, 51(1):341–363, 2017.

[30] M. Bachmayr, A. Cohen, and G. Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients. *ESAIM. Math. Model. Numer. Anal.*, 51(1):321–339, 2017.

[31] G. Baird, R. Bürger, P. E. Méndez, and R. Ruiz-Baier. Second-order schemes for axisymmetric Navier-Stokes-Brinkman and transport equations modelling water filters. *Numer. Math.*, 147(2):431–479, 2021.

[32] F. Bartel, M. Schäfer, and T. Ullrich. Constructive subsampling of finite frames with applications in optimal function recovery. *Appl. Comput. Harmon. Anal.*, 65:209–248, 2023.

[33] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, 2000.

[34] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, New York, NY, 2011.

[35] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal stochastic Galerkin methods for a class of PDEs with random coefficients. *Comput. Math. Appl.*, 67(4):732–751, 2014.

[36] J. Beck, R. Tempone, F. Nobile, and L. Tamellini. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Models Methods Appl. Sci.*, 22(9):1250023, 2012.

[37] S. Becker, A. Jentzan, M. S. Mũller, and P. von Wurstemberger. Learning the random variables in Monte Carlo simulations with stochastic gradient descent: Machine learning for parametric PDEs and financial derivative pricing. *Math. Finance*, 34(1):90–150, 2023.

[38] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[39] R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, Princeton, NJ, 1961.

[40] A. Belloni, V. Chernozhukov, and L. Wang. Square-root LASSO: pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806, 2011.

[41] B. Bernardi, C. Métivet and B. Pernaud-Thomas. Couplage des équations de navier-stokes et dela chaleur : le modéle et son approximationpar éléments finis. *RAIRO Modél. Math. Anal. Numér.*, 29:8871–921, 1995.

[42] K. Bhattacharya, N. Hosseini, B. Kovachki, and A. Stuart. Model reduction and neural networks for parametric PDEs. *J. Comput. Math.*, 7:121–157, 2021.

[43] M. Bieri, R. Andreev, and C. Schwab. Sparse tensor discretization of elliptic SPDEs. *SIAM J. Sci. Comput.*, 31(6):4281–4304, 2010.

[44] P. Binev, A. Bonito, R. DeVore, and G. Petrova. Optimal learning. *Calcolo*, 61(1), 2024.

[45] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comput. Phys.*, 230:2345–2367, 2011.

[46] D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications*. Springer Berlin, Heidelberg, 1 edition, 2013.

[47] A. Bonito, R. DeVore, D. Guignard, P. Jantsch, and G. Petrova. Polynomial approximation of anisotropic analytic functions of several variables. *Constr. Approx.*, 53:319–348, 2021.

[48] J.-L. Bouchot, H. Rauhut, and C. Schwab. Multi-level compressed sensing Petrov-Galerkin discretization of high-dimensional parametric PDEs. *arXiv:1701.01671*, 2017.

[49] S. Brenner and R. L. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, 2nd edition, 2005.

[50] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[51] S. Brugiapaglia, S. Dirksen, H. C. Jung, and H. Rauhut. Sparse recovery in bounded Riesz systems with applications to numerical methods for PDEs. *Appl. Comput. Harmon. Anal.*, 53:231–269, 2021.

[52] C. Cai, C. Wang, and S. Zhang. Mixed finite element methods for incompressible flow: stationary Navier-Stokes equations. *SIAM J. Numer. Anal.*, 48:79–94, 2010.

[53] S. Cai, Z. Wang, L. Lu, T. A. Zaki, and G. E. Karniadakis. DeepM&Mnet: Inferring the electroconvection multiphysics fields based on operator approximation by neural networks. *J. Comput. Phys.*, 436:110296, 2021.

[54] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[55] C. Cao and J. Wu. Global regularity for the two-dimensional anisotropic Boussinesq equations with vertical dissipation. *Arch. Rational Mech. Anal.*, 208:985–1004, 2013.

[56] J. E. Castrillon-Candas, F. Nobile, and R. F. Tempone. Analytic regularity and collocation approximation for elliptic PDEs with random domain deformations. *Comput. Math. Appl.*, 71(6):1173–1197, 2016.

[57] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[58] A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numer.*, 25:161–319, 2016.

[59] A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal-dual algorithm. *Math. Program.*, 159(1-2):253–287, 2016.

[60] A. Chernov and D. Dũng. New explicit-in-dimension estimates for the cardinality of high-dimensional hyperbolic crosses and approximation of functions having mixed smoothness. *J. Complexity*, 32:92–121, 2016.

[61] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone. Discrete least squares polynomial approximation with random evaluations - application to parametric and stochastic elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 49(3):815–837, 2015.

[62] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Found. Comput. Math.*, 14(4):601–633, 2014.

[63] A. Chkifa, A. Cohen, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. *J. Math. Pures Appl.*, 103(2):400–428, 2015.

[64] A. Chkifa, N. Dexter, H. Tran, and C. G. Webster. Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. *Math. Comp.*, 87(311):1415–1450, 2018.

[65] B. Choi, M. A. Iwen, and F. Krahmer. Sparse harmonic transforms: a new class of sublinear-time algorithms for learning functions of many variables. *Found. Comput. Math.*, 21(2):275–329, 2021.

[66] B. Choi, M. A. Iwen, and T. Volkmer. Sparse harmonic transforms II: best s-term approximation guarantees for bounded orthonormal product bases in sublinear-time. *Numer. Math.*, 148(2):293–362, 2021.

[67] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. SIAM, 2002.

[68] L. Cicci, S. Fresca, and A. Manzoni. Deep-HyROMnet: A deep learning-based operator approximation for hyper-reduction of nonlinear parametrized PDEs. *J. Sci. Comput.*, 93:57, 2022.

[69] A. Cohen, W. Dahmen, and R. A. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.

[70] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13:819–834, 2013.

[71] A. Cohen and R. A. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numer.*, 24:1–159, 2015.

[72] A. Cohen, R. A. DeVore, and C. Schwab. Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10:615–646, 2010.

[73] A. Cohen, R. A. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl. (Singap.)*, 9(1):11–47, 2011.

[74] A. Cohen and G. Migliorati. Optimal weighted least-squares methods. *SMAI J. Comput. Math.*, 3:181–203, 2017.

[75] A. Cohen and G. Migliorati. Multivariate approximation in downward closed polynomial spaces. In Josef Dick, Frances Y. Kuo, and Henryk Woźniakowski, editors, *Contemporary Computational Mathematics – A Celebration of the 80th Birthday of Ian Sloan*, pages 233–282. Springer, Cham, Switzerland, 2018.

[76] A. Cohen, G. Migliorati, and F. Nobile. Discrete least-squares approximations over optimized downward closed polynomial spaces in arbitrary dimension. *Constr. Approx.*, 45:497–519, 2017.

[77] A. Cohen, C. Schwab, and J. Zech. Shape holomorphy of the stationary Navier-Stokes Equations. *SIAM J. Math. Anal.*, 50(2):1720–1752, 2018.

[78] M. J. Colbrook. WARPd: A linearly convergent first-order primal-dual algorithm for inverse problems with approximate sharpness conditions. *SIAM J. Imaging Sci.*, 15(3):1539–1575, 2022.

[79] M. J. Colbrook, V. Antun, and A. C. Hansen. The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and smale's 18th problem. *Proc. Natl. Acad. Sci. USA*, 119(12):e2107151119, 2022.

[80] E. Colmenares, G. N. Gatica, and S. Moraga. A Banach spaces-based analysis of a new fully-mixed finite element method for the Boussinesq problem. *ESAIM Math. Model. Numer. Anal.*, 54(5):1525–1568, 2020.

[81] E. Colmenares and M. Neilan. Dual-mixed finite element methods for the stationary Boussinesq problem. *Comp. Math. Appl.*, 72:1828–1850, 2016.

[82] D. Dũng and V. K. Nguyen. Deep ReLU neural networks in high-dimensional approximation. *Neural Netw.*, 142:619–635, 2021.

[83] D. Dũng, V. K. Nguyen, and D. T. Pham. Deep ReLU neural network approximation in Bochner spaces and applications to parametric PDEs. *J. Complexity*, 79:101779, 2023.

[84] D. Dũng, V. K. Nguyen, and M. X. Thao. Computation complexity of deep ReLU neural networks in high-dimensional approximation. *JCC*, 37(3):291–320, 2021.

[85] N. Dal Santo, S. Deparis, and L. Pegolotti. Data driven approximation of parametrized PDEs by reduced basis and neural betworks. *J. Comput. Phys.*, 416:109550, 2020.

[86] I. Danaila, R. Moglan, F. Hecht, and S. Le Masson. A newton method with adaptive finite elements for solving phase-change problems with natural convection. *J. Comput. Phys.*, 274:826–840, 2014.

[87] J. Daws and C. Webster. Analysis of deep neural networks with quasi-optimal polynomial approximation rates. *arXiv:1912.02302*, 2019.

[88] J. Daws and C. G. Webster. A Polynomial-Based Approach for Architectural Design and Learning with Deep Neural Networks. *arXiv:1905.10457*, 2019.

[89] C. de Boor and A. Ron. Computational aspects of polynomial interpolation in several variables. *Math. Comp.*, 58:705–727, 1992.

[90] M. De Hoop, D. Z. Huang, E. Qian, and A. Stuart. The cost-accuracy trade-off in operator learning with neural networks. *J. Mach. Learn.*, 1:299–341, 2022.

[91] T. De Ryck, S. Lanthaler, and S. Mishra. On the approximation of functions by tanh neural networks. *Neural Networks*, 143:732–750, 2021.

[92] F. Deutsch. Linear selections for the metric projection. *J. Funct. Anal.*, 49:269–292, 1982.

[93] R. A. DeVore, R. Howard, and C. Micchelli. Optimal nonlinear approximation. *Manuscripta Mathematica*, 63(4):469–478, 1989.

[94] N. Dexter. *Sparse reconstruction techniques for solutions of high-dimensional parametric PDEs*. PhD thesis, University of Tennessee, 2018.

[95] N. Dexter, H. Tran, and C. Webster. A mixed ℓ_1 regularization approach for sparse simultaneous approximation of parameterized PDEs. *ESAIM Math. Model. Numer. Anal.*, 53:2025–2045, 2019.

[96] N. Dexter, C. Webster, and G. Zhang. Explicit cost bounds of stochastic Galerkin approximations for parameterized PDEs with random coefficients. *Comput. Math. Appl.*, 71(11):2231–2256, 2016.

[97] P. Diaz, A. Doostan, and J. Hampton. Sparse polynomial chaos expansions via compressed sensing and D-optimal design. *Comput. Methods Appl. Mech. Engrg.*, 336:640–666, 2018.

[98] J. Dick, F. Y. Kuo, Q. T. Le Gia, and D. Nuyens. Higher order QMC Petrov-Galerkin discretization for affine parametric operator equations with random field inputs. *SIAM J. Numer. Anal.*, 52(6):2676–2702, 2014.

[99] J. Dick, F. Y. Kuo, Q. T. Le Gia, and C. Schwab. Higher order Quasi-Monte Carlo integration for holomorphic, parametric operator equations. *SIAM/ASA J. Uncertain. Quantif.*, 4(1):48–79, 2016.

[100] M. Dolbeault and A. Cohen. Optimal sampling and Christoffel functions on general domains. *Constr. Approx.*, 56:121–163, 2021.

[101] M. Dolbeault and A. Cohen. Optimal pointwise sampling for L^2 approximation. *J. Complexity*, 68:101602, 2022.

[102] M. Dolbeault, D. Krieg, and M. Ullrich. A sharp upper bound for sampling numbers in L_2. *Appl Comput Harmon Anal.*, 63:113–134, 2023.

[103] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[104] A. Doostan and H. Owhadi. A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.*, 230(8):3015–3034, 2011.

[105] Y. Dutil, D. R. Rousse, N. B. Salah, S. Lassue, and L. Zalewski. A review on phase-change materials: Mathematical modeling and simulations. *Renew. Sustain. Energy Rev.*, 15(1):112–130, 2011.

[106] M. Eigel, S. Farchmin, N. Heidenreich, and P. Trunschke. Adaptive non-intrusive reconstruction of solutions to high-dimensional parametric PDEs. *SIAM J. Comput.*, 45(2):A457–A479, 2021.

[107] M. Eigel, C. J. Gittelson, C. Schwab, and E. Zander. A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes. *M2AN Math. Model. Numer. Anal.*, 49(5):1367–1398, 2015.

[108] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Appl. Math. Sci.* Springer–Verlag, New York, NY, 2004.

[109] L. C. Evans. *Partial Differential Equations*. AMS, 1998.

[110] M. Farhloul, S. Nicaise, and L. Paquet. A priori and a posteriori error estimations for the dual mixed finite element method of the Navier-Stokes problem. *Numer. Methods Partial Differ. Equ.*, 25:843–869, 2009.

[111] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich. The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. *J. Complex.*, 26(6):629–640, 2010.

[112] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Appl. Numer. Harmon. Anal. Birkhäuser, New York, NY, 2013.

[113] P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5):205–228, 2005.

[114] G. N. Gatica. *A Simple Introduction to the Mixed Finite Element Method*. Springer Cham, 2014.

[115] G. N. Gatica, N. Nuñez, and R. Ruiz-Baier. New non-augmented mixed finite element methods for the navier–stokes–brinkman equations using banach spaces. *In review*, 2022.

[116] G. N. Gatica, R. Oyarzúa, R. Ruiz-Baier, and Y. D. Sobral. Banach spaces-based analysis of a fully-mixed finite element method for the steady-state model of fluidized beds. *Comput. Math. Appl.*, 84:244–276, 2021.

[117] L. F. Gatica, R. Oyarzúa, and N. Sánchez. A priori and a posteriori error analysis of an augmented mixed-FEM for the Navier-Stokes-Brinkman problem. *Comput. Math. Appl.*, 75(7):2420–2444, 2018.

[118] M. Geist, P. Petersen, M. Raslan, R. Schneider, and G. Kutyniok. Numerical solution of the parametric diffusion equation by deep neural networks. *J. Sci. Comput.*, 88:22, 2021.

[119] R. Ghanem, D. Higdon, and H. Owhadi. *Handbook of Uncertainty Quantification.* Springer, Switzerland, 2017.

[120] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations: Theory and algorithms* . Springer Berlin, Heidelberg, 1986.

[121] T. J. Grady II, R. Khan, M. Louboutin, Z. Yin, P. A. Witte, R. Chandra, R. J. Hewett, and F. J. Herrmann. Model-parallel Fourier neural operators as learned surrogates for large-scale parametric PDEs. Technical Report TR-CSE-2022-1, 04 2022.

[122] M. Gunzburger, C. G. Webster, and G. Zhang. Stochastic finite element methods for partial differential equations with random input data. *Acta Numer.*, 23:521–650, 2014.

[123] L. Guo, Y. Liu, and L. Yan. Sparse recovery via ℓ_q-minimization for polynomial chaos expansions. *Numer. Math. Theor. Meth. Appl.*, 10(4):775–797, 2017.

[124] L. Guo, A. Narayan, L. Yan, and T. Zhou. Weighted approximate Fekete points: sampling for least-squares polynomial approximation. *SIAM J. Sci. Comput.*, 40(1):A366–A387, 2018.

[125] L. Guo, A. Narayan, and T. Zhou. A gradient enhanced ℓ_1-minimization for sparse approximation of polynomial chaos expansions. *J. Comput. Phys.*, 367:49–64, 2018.

[126] L. Guo, A. Narayan, and T. Zhou. Constructing least-squares polynomial approximations. *SIAM Rev.*, 62(2):483–508, 2020.

[127] L. Guo, A. Narayan, T. Zhou, and Y. Chen. Stochastic collocation methods via ℓ_1 minimization using randomized quadratures. *SIAM J. Sci. Comput.*, 39(1):A333–A359, 2017.

[128] Z. Guo, B. Shi, and C. Zheng. A coupled lattice BGK model for the Boussinesq equations. *Int. J. Numer. Meth. Fluids*, 39:325–342, 2002.

[129] L. Guta and S. Sundar. Navier-stokes-brinkman system for interaction of viscous wves with a submerged porous structure. *Tamkang J. Math.*, 41:2017–243, 2010.

[130] M. Hadigol and A. Doostan. Least squares polynomial chaos expansion: a review of sampling strategies. *Comput. Methods Appl. Mech. Engrg.*, 332:382–407, 2018.

[131] A.-L. Haji-Ali, F. Nobile, R. Tempone, and S. Wolfers. Multilevel weighted least squares polynomial approximation. *ESAIM. Math. Model. Numer. Anal.*, 54(2):649–677, 2020.

[132] J. Hampton and A. Doostan. Coherence motivated sampling and convergence analysis of least squares polynomial chaos regression. *Comput. Methods Appl. Mech. Engrg.*, 290:73–97, 2015.

[133] J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: convergence analysis and sampling strategies. *J. Comput. Phys.*, 280:363–386, 2015.

[134] J. Hampton and A. Doostan. Compressive sampling methods for sparse polynomial chaos expansions. In Roger Ghanem, David Higdon, and Houman Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 827–855. Springer, Cham, Switzerland, 2017.

[135] J. Hampton and A. Doostan. Basis adaptive sample efficient polynomial chaos (BASE-PC). *J. Comput. Phys.*, 371:20–49, 2018.

[136] M. Hansen and C. Schwab. Analytic regularity and nonlinear approximation of a class of parametric semilinear elliptic PDEs. *Math. Nachr.*, 286(8-9):832–860, 2013.

[137] M. Hansen and C. Schwab. Sparse adaptive approximation of high dimensional parametric initial value problems. *Vietnam J. Math.*, 41(2):181–215, 2013.

[138] C. Heiß, I. Gühring, and M. Eigel. A neural multilevel method for high-dimensional parametric PDEs. In *Advances in Neural Information Processing Systems*, 2021.

[139] C. Heiß, I. Gühring, and M. Eigel. Multilevel CNNs for parametric PDEs. *J. Mach. Learn. Res.*, 24:1–42, 2023.

[140] F. Henriquez and C. Schwab. Shape Holomorphy of the Caldéron Projector for the Laplacian in $\mathbb{R}^2$. *Integral Equations Operator Theory*, 93(4):43, 2021.

[141] L. Herrmann, J. A. A. Opschoor, and C. Schwab. Constructive deep ReLU neural network approximation. *J. Sci. Comput.*, 90:75, 2022.

[142] M. Hervé. *Analyticity in Infinite Dimensional Spaces*, volume 10 of *De Gruyter Stud. Math.* Walter de Gruyter, Berlin, Germany, 1989.

[143] J. Hesthaven, G. Rozza, and B. Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. SpringerBriefs Math. Springer, Cham, 2016.

[144] R. Hiptmair, L. Scarabosio, C. Schillings, and C. Schwab. Large deformation shape uncertainty quantification in acoustic scattering. *Adv. Comput. Math.*, 44(5):1475–1518, 2018.

[145] L. S. T. Ho, H. Schaeffer, G. Tran, and R. Ward. Recovery guarantees for polynomial coefficients from weakly dependent data with outliers. *J. Approx. Theory*, 259:105472, 2020.

[146] V. H. Hoang and C. Schwab. Regularity and generalized polynomial chaos approximation of parametric and random second-order hyperbolic partial differential equations. *Anal. Appl. (Singap.)*, 10(3):295–326, 2012.

[147] V. H. Hoang and C. Schwab. Sparse tensor Galerkin discretization of parametric and random parabolic PDEs—analytic regularity and generalized polynomial chaos approximation. *SIAM J. Math. Anal.*, 45(5):3050–3083, 2013.

[148] R.B. Holmes and B.R. Kripke. Smoothness of approximation. *Michigan Math. J.*, 15:225–248, 1968.

[149] J. Howell and N. Walkington. Dual-mixed finite element methods for the Navier-Stokes equations. *ESAIM Math. Model. Numer. Anal.*, 47(3):789–805, 2016.

[150] W. R. Hwang and S. G. Advani. Numerical simulations of Stokes–Brinkman equations for permeability prediction of dual scale fibrous porous media. *Phys. Fluids.*, 22:113101, 2010.

[151] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. Bochner spaces. In: Analysis in Banach Spaces. *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics, vol 63. Springer, Cham*, pages 23:521–650, 2016.

[152] R. Ingram. Finite element approximation of nonselenoidal, viscous flows around porous and solid obstacles. *SIAM J. Numer. Anal.*, 49(2):491–520, 2011.

[153] J. D. Jackson. *Classical Electrodynamics*. Wiley, New York, 3rd edition, 1999.

[154] J. D. Jakeman, M. S. Eldred, and K. Sargsyan. Enhancing ℓ_1-minimization estimates of polynomial chaos expansions using basis selection. *J. Comput. Phys.*, 289:18–34, 2015.

[155] J. D. Jakeman, A. Narayan, and T. Zhou. A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions. *SIAM J. Sci. Comput.*, 39(3):A1114–A1144, 2017.

[156] P. A. Jantsch. *Efficient methods for multidimensional global polynomial approximation with applications to random PDEs*. PhD thesis, University of Tennessee, 2017.

[157] C. Jerez-Hanckes, C. Schwab, and J. Zech. Electromagnetic wave scattering by random surfaces: shape holomorphy. *Math. Models Methods Appl. Sci.*, 27(12):2229–2259, 2017.

[158] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, K. Tunyasuvunakool, O. Ronneberger, R. Bates, A. Zidek, A. Bridgland, C. Meyer, S. A. A. Kohl, A. Potapenko, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, M. Steinegger, M. Pacholska, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. High Accuracy Protein Structure Prediction Using Deep Learning. *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstract Book)*, 2020.

[159] J.H. Jung, S. Gottlieb, S.O. Kim, C.L. Bresten, and D. Higgs. Recovery of high order accuracy in radial basis function approximations of discontinuous problems. *J Sci Comput*, 45:359–381, 2010.

[160] L. Kämmerer, T. Ullrich, and T. Volkmer. Worst case recovery guarantees for least squares approximation using random samples. *Constr. Approx.*, 54(2):295–352, 2021.

[161] K. Khadra, P. Angot, S. Parneix, and J. P. Caltagirone. Fictitious domain approach for numerical modelling of Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, 34:651–684, 2000.

[162] B. Khara, A. Balu, A. Joshi, S. Sarkar, C. Hegde, A. Krishnamurthy, and B. Ganapathysubramanian. NeuFENet: Neural finite element solutions with theoretical bounds for parametric PDEs. *Eng. Comput.*, 2024.

[163] Y. Khoo, J. Lu, and L. Ying. Solving parametric PDE problems with artificial neural networks. *European J. Appl. Math.*, 32(3):421–435, 2021.

[164] D. P. Kingma and J. Ba. Adam: a method for stochastic optimization. *arXiv:1412.6980*, 2017.

[165] A. N. Kolmogorov. Über die bester Annäherung von Funktionen einer gegebenen Funktionenklasse. *Ann. Maths*, 37:107–110, 1936.

[166] I. A. Kougioumtzoglou, I. Petromichelakis, and A. F. Psaros. Sparse representations and compressive sampling approaches in engineering mechanics: a review of theoretical concepts and diverse applications. *Probabilistic Eng. Mech.*, 61:103082, 2020.

[167] N. Kovachki, Z. Li, B. Liu, K. Azizzadnesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: learning maps between function spaces with applications to PDEs. *J. Mach. Learn. Res.*, 24:1–97, 2023.

[168] F. Kröpfl, R. Maier, and D. Peterseim. Operator compression with deep neural networks. *Adv. Cont. Discr. Mod.*, 29, 2022.

[169] T. Kühn, W. Sickel, and T. Ullrich. Approximation of mixed order Sobolev functions on the d-torus: asymptotics, preasymptotics, and d-dependence. *Constr. Approx.*, 42:353–398, 2015.

[170] A. Kunoth and C. Schwab. Analytic regularity and GPC approximation for control problems constrained by linear parametric elliptic and parabolic PDEs. *SIAM J. Control Optim.*, 51(3):2442–2471, 2013.

[171] J. Kuntzman. *Méthodes Numériques – Interpolation, Dérivées*. Dunod, Paris, France, 1959.

[172] G. Kutyniok, P. Petersen, M. Raslan, and R. Schneider. A theoretical analysis of deep neural networks and parametric PDEs. *Constr. Approx.*, 55:73–125, 2022.

[173] B. Lamichhane. Mixed finite element methods for the Poisson equation using biorthogonal and quasi-biorthogonal systems. *Adv. Numer. Anal.*, 2013:1–9, 2013.

[174] S. Lanthaler, S. Mishra, and G. E. Karniadakis. Error estimates for DeepOnets: A deep learning framework in infinite dimensions. *Trans. math. appl.*, 6(1):tnac001, 2022.

[175] A. Larios, E. Lunasin, and E. S. Titi. Global well-posedness for the 2D Boussinesq system with anisotropic viscosity and without heat diffusion. *J. Differ. Equ.*, 255:2636–2654, 2013.

[176] O. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Sci. Comput. Springer, Dordrecht, Netherlands, 2010.

[177] K. Lee, H. C. Elman, and B. Sousedík. A low-rank solver for the Navier-Stokes equations with uncertain viscosity. *SIAM-ASA J. Uncertain. Quantif.*, 7(4):1275–1300, 2019.

[178] Z. Lei, L. Shi, and C. Zeng. Solving parametric partial differential equations with deep rectified quadratic unit neural networks. *J. Sci. Comput.*, 93:80, 2022.

[179] B. Li, S. Tang, and H. Yu. Better approximations of high dimensional smooth functions by deep neural networks with rectified power units. *1903.05858*, 2019.

[180] B. Li, S. Tang, and H. Yu. Better approximations of high dimensional smooth functions by deep neural networks with rectified power units. *Commun. Comput. Phys.*, 27:379–411, 2020.

[181] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *ICLR*, 2021.

[182] I. Limonova and V. Temlyakov. On sampling discretization in L_2. *J. Math. Anal. Appl.*, 515(2):126457, 2022.

[183] Y. Liu and L. Guo. Stochastic collocation via l_1-minimisation on low discrepancy point sets with application to uncertainty quantification. *East Asian J. Appl. Math.*, 6(2):171–191, 2016.

[184] S. A. Lorca and J. L. Boldrini. Stationary solutions for generalized Boussinesq models. *J. Differential equations*, 134:389–406, 1996.

[185] S. A. Lorca and J. L. Boldrini. The initial value problem for a generalized Boussinesq model. *Nonlinear Anal. Theory Methods Appl.*, 36(4):457–480, 1999.

[186] G. G. Lorentz and R. A. Lorentz. Solvability problems of bivariate interpolation I. *Constr. Approx.*, 2:153–169, 1986.

[187] G. G. Lorentz, M. v. Golitschek, and Y. Makovoz. *Constructive approximation: advanced problems*, volume 304. Springer Berlin, 1996.

[188] J. G. Lu, S. B. Lee, T. S. Lundström, and W. R. Hwang. Numerical simulation on void formation and migration using Stokes-Brinkman coupling with effective dual-scale fibrous porous media. *Composites Part A: Applied Science and Manufacturing*, 152:106683, 2022.

[189] L. Lu, P. Jin, Z. Pang, G. Zhang, and G. E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.*, 3:218–229, 2021.

[190] N. Lüthen, S. Marelli, and B. Sudret. Sparse polynomial chaos expansions: literature survey and benchmark. *SIAM/ASA J. Uncertain. Quantif.*, 9(2):593–649, 2021.

[191] L. Mathelin and K. A. Gallivan. A compressed sensing approach for partial differential equations with random input data. *Commun. Comput. Phys.*, 12(4):919–954, 2012.

[192] B. J. Matkowsky and G. I. Sivashinsky. An asymptotic derivation of two models in flame theory associated with the constant density approximation. *SIAM J. Appl. Math.*, 37(3):686–699, 1979.

[193] W. McCulloch and W. Pitts. A logical calculus of ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943.

[194] C. A. Micchelli and T. J. Rivlin. *A survey of optimal recovery*, volume Optimal Estimation in Approximation Theory of *The IBM Research Symposia Series*. Springer, Boston, MA, 1977.

[195] G. Migliorati. *Polynomial approximation by means of the random discrete L^2 projection and application to inverse problems for PDEs with stochastic data*. PhD thesis, Politecnico di Milano, 2013.

[196] G. Migliorati. Adaptive polynomial approximation by means of random discrete least squares. In Assyr Abdulle, Simone Deparis, Daniel Kressner, Fabio Nobile, and Marco Picasso, editors, *Numerical Mathematics and Advanced Applications – ENUMATH 2013*, pages 547–554, Cham, Switzerland, 2015. Springer.

[197] G. Migliorati. Adaptive approximation by optimal weighted least squares methods. *SIAM J. Numer. Anal*, 57(5):2217–2245, 2019.

[198] G. Migliorati. Multivariate approximation of functions on irregular domains by weighted least-squares methods. *IMA J. Numer. Anal.*, 41(2):1293–1317, 2021.

[199] G. Migliorati and F. Nobile. Analysis of discrete least squares on multivariate polynomial spaces with evaluations in low-discrepancy point sets. *J. Complexity*, 31:517–542, 2015.

[200] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Analysis of the discrete L^2 projection on polynomial spaces with random evaluations. *Found. Comput. Math.*, 14:419–456, 2014.

[201] R. Mondal, S. Mondal, K. V. Kurada, S. Bhattacharjee, S. Sengupta, M. Mondal, S. Karmakar, S. De, and I. M. Griffiths. Modelling the transport and adsorption dynamics of arsenic in a soil bedfilter. *Chem. Eng. Sci.*, 210:115205, 2019.

[202] H. Montanelli, H. Yang, and Q. Du. Deep ReLU networks overcome the curse of dimensionality for bandlimited functions. *J. Comput. Math.*, 39(6):801–815, 2021.

[203] A. Narayan. Computation of induced orthogonal polynomial distributions. *Electron. Trans. Numer. Anal.*, 50:71–97, 2018.

[204] A. Narayan, J. D. Jakeman, and T. Zhou. A Christoffel function weighted least squares algorithm for collocation approximations. *Math. Comp.*, 86:1913–1947, 2017.

[205] A. Narayan and T. Zhou. Stochastic collocation on unstructured multivariate meshes. *Commun. Comput. Phys.*, 18(1):1–36, 2015.

[206] N. H. Nelsen and A. M. Stuart. The random feature model for input-output maps between Banach spaces. *SIAM J. Sci. Comput.*, 43(5):A3212–A3243, 2021.

[207] A. S. Nemirovski. Prox-method with rate of convergence $\mathcal{O}(1/t)$ for vari-ational inequalities with Lipschitz continuous monotone operators and smoothconvex-concave saddle point problems. *SIAM J. Optim.*, 15:229–251, 2004.

[208] L. W. T. Ng and M. S. Eldred. Multifidelity uncertainty quantification using nonintrusive polynomial chaos and stochastic collocation. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, volume 45. AIAA, 2012.

[209] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[210] E. Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*. Number 1. Springer Berlin, Heidelberg, 1988.

[211] E. Novak and H. Woźniakowski. *Tractability of Multivariate Problems, Volume I: Linear Information*, volume 6. European Math. Soc. Publ. House, Zürich, 2008.

[212] E. Novak and H. Woźniakowski. *Tractability of Multivariate Problems, Volume II: Standard Information for functionals*, volume 12. European Math. Soc., Zürich, 2010.

[213] J. A. A. Opschoor, P. C. Petersen, and C. Schwab. Deep ReLU networks and high-order finite element methods. *Anal. Appl. (Singap.)*, 18(5):715–770, 2020.

[214] J. A. A. Opschoor, C. Schwab, and J. Zech. *Deep learning in high dimension: ReLU neural network expression for Bayesian PDE inversion*, pages 419–462. De Gruyter, Berlin, Boston, 2022.

[215] J. A. A. Opschoor, Ch. Schwab, and J. Zech. Exponential ReLU DNN expression of holomorphic maps in high dimension. *Constr. Approx.*, 55(1):537–582, 2022.

[216] R. Oyarzúa, T. Qin, and D. Schötzau. An exactly divergence-free finite element method for a generalized Boussinesq problem. *IMA J. Numer. Anal.*, 34:1104–1135, 2014.

[217] J. Peng, J. Hampton, and A. Doostan. A weighted ℓ_1-minimization approach for sparse polynomial chaos expansions. *J. Comput. Phys.*, 267:92–111, 2014.

[218] J. Peng, J. Hampton, and A. Doostan. On polynomial chaos expansion via gradient-enhanced l_1-minimization. *J. Comput. Phys.*, 310:440–458, 2016.

[219] I. Perugia and D. Schötzau. An hp-analysis of the local discontinuous galerkin method for diffusion problems. *J. Sci. Comput.*, 17:561–571, 2002.

[220] M. Pinkus. *N-widths in Approximation Theory*. Springer–Verlag, Berlin, 1968.

[221] D. R. Poirier and G. H. Geiger. *Conduction of Heat in Solids*, pages 281–327. Springer International Publishing, Cham, 2016.

[222] C. E. Powel and D. J. Solvester. Preconditioning steady-state Navier-Stokes equations with random data. *SIAM J. Sci. Comput.*, 34(5):A2482–A2506, 2012.

[223] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction.* vol. 92 of UNITEXT, Springer, Cham, 2015.

[224] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*, volume 23 of *Springer Ser. Comput. Math.* Springer–Verlag, Berlin Heidelberg, Germany, 1994.

[225] H. Rauhut and C. Schwab. Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations. *Math. Comp.*, 86:661–700, 2017.

[226] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1-minimization. *J. Approx. Theory*, 164(5):517–533, 2012.

[227] H. Rauhut and R. Ward. Interpolation via weighted ℓ^1 minimization. *Appl. Comput. Harmon. Anal.*, 40(2):321–351, 2016.

[228] J. Renegar and B. Grimmer. A simple nearly optimal restart scheme for speeding up first-order methods. *Found. Comput. Math.*, 22(1):211–256, 2022.

[229] V. Roulet and A. Boumal, N. d'Aspremont. Computational complexity versus statistical performance on sparse recovery problems. *Inf. Inference*, 9(1):1–32, 2020.

[230] V. Roulet and A. d'Aspremont. Sharpness, restart, and acceleration. *SIAM J. Optim.*, 30(1):262–289, 2020.

[231] W. Rudin. *Principles of Mathematical Analysis*, volume 3. McGraw-Hill New York, 1964.

[232] W. Rudin. *Functional Analysis.* McGraw–Hill , Inc., New York, NY, 2nd edition, 1991.

[233] F. Sabetghadam, S. Sharafatmandjoor, and F. Norouzi. Fourier spectral embedded boundary solution of the Poisson's and Laplace equations with Dirichlet boundary conditions. *J. Comput. Phys.*, 228:55–74, 2009.

[234] C. Schwab and C. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.

[235] C. Schwab and J. Zech. Deep learning in high dimension: neural network expression rates for generalized polynomial chaos expansions in UQ. *Anal. Appl. (Singap.)*, 17(1):19–55, 2019.

[236] C. Schwab and J. Zech. Deep learning in high dimension: neural network approximation of analytic functions in $L^2(\mathbb{R}^d, \gamma_d)$. *arXiv:2111.07080*, 2021.

[237] P. Seshadri, A. Narayan, and S. Mahadevan. Effectively subsampled quadratures for least squares polynomial approximations. *SIAM/ASA J. Uncertain. Quantif.*, 5:1003–1023, 2017.

[238] Y. Shin and D. Xiu. Correcting data corruption errors for multivariate function approximation. *SIAM J. Sci. Comput.*, 38(4):A2492–A2511, 2016.

[239] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Comput. Sci. Eng. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[240] B. Sousedík and H. C. Elman. Stochastic Galerkin methods for the steady-state Navier–Stokes equations. *J. Comput. Phys.*, 316:435–452, 2016.

[241] M. I. Stesin. Aleksandrov diameters of finite-dimensional sets and of classes of smooth functions. *Dokl. Akad. Nauk SSSR*, 220(6):1278–1281, 1975.

[242] M. Stoyanov. User manual: Tasmanian sparse grids. Technical Report ORNL/TM-2015/596, Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN, 2015.

[243] M. Stoyanov, D. Lebrun-Grandie, J. Burkardt, and D. Munster. Tasmanian, 9 2013.

[244] M. K. Stoyanov and C. G. Webster. A dynamically adaptive sparse grid method for quasi-optimal interpolation of multidimensional functions. *Comput. Math. Appl.*, 71(11):2449–2465, 2016.

[245] Y. Sui. *Weighted ℓ^1 minimization techniques for compressed sensing and their applications*. PhD thesis, Simon Fraser University, 2020.

[246] T. J. Sullivan. *Introduction to Uncertainty Quantification*, volume 63 of *Texts Appl. Math.* Springer, Cham, Switzerland, 2015.

[247] T. Sun and C.-H. Zhang. Scaled sparse linear regression. *Biometrika*, 99(4):879–898, 2012.

[248] G. Szegö. *Orthogonal Polynomials*, volume 23 of *Amer. Math. Soc. Colloq. Publ.* American Mathematical Society, Providence, RI, 4th edition, 1975.

[249] G. Tang. *Methods for high dimensional uncertainty quantification: regularization, sensitivity analysis, and derivative enhancement*. PhD thesis, Stanford University, 2013.

[250] G. Tang and G. Iaccarino. Subsampled Gauss quadrature nodes for estimating polynomial chaos expansions. *SIAM/ASA J. Uncertain. Quantif.*, 2(1):423–443, 2014.

[251] T. Tang and T. Zhou. On discrete least-squares projection in unbounded domain with random evaluations and its application to parametric uncertainty quantification. *SIAM J. Sci. Comput.*, 36(5):A2272–A2295, 2014.

[252] T. Tao. *An Introduction to Measure Theory*, volume 126 of *Grad. Stud. Math.* American Mathematical Society, Providence, RI, 2011.

[253] V. Temlyakov. On optimal recovery in L_2. *J. Complexity*, 65:101545, 2021.

[254] V. N. Temlyakov. Approximation of periodic functions of several variables with bounded mixed derivative. *Trudy Mat. Inst. Steklov*, 156:233–260; English translation in Proc. Steklov Inst. Math., 2 (1983), 1980.

[255] R. A. Todor and C. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.*, 27(2):232–261, 2007.

[256] M. Torrilhon and N. Sarna. Hierarchical Boltzmann simulations and model error estimation. *J. Comput. Phys.*, 342:66–84, 2017.

[257] H. Tran and C. Webster. A class of null space conditions for sparse recovery via nonconvex, non-separable minimizations. *Results Appl. Math.*, 3:100011, 2019.

[258] H. Tran, C. G. Webster, and G. Zhang. Analysis of quasi-optimal polynomial approximations for parameterized PDEs with deterministic and stochastic coefficients. *Numer. Math.*, 137(2):451–493, 2017.

[259] Y. Traonmilin and R. Gribonval. Stable recovery of low-dimensional cones in Hilbert spaces: one RIP to rule them all. *Appl. Comput. Harmon. Anal.*, 45(1):170–205, 2018.

[260] J. F. Traub, H. Woźniakowski, and G. W. Wasilkowski. *Information-Based Complexity.* Elsevier Science and Technology Books, 1988.

[261] L. N. Trefethen. *Approximation Theory and Approximation Practice.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[262] P. Tsilifis, X. Huan, C. Safta, K. Sargsyan, G. Lacaze, J. C. Oefelein, H. N. Najm, and R. G. Ghanem. Compressive sensing adaptation for polynomial chaos expansions. *J. Comput. Phys.*, 380:29–47, 2019.

[263] K. Urban. *Wavelet Methods for Elliptic Partial Differential Equations.* Oxford University Press, 2009.

[264] G. G. Walter and X. Shen. *Wavelets and other orthogonal systems.* CRC Press, 2001.

[265] W. Walter. *Ordinary Differential Equations*, volume 182 of *Grad. Texts in Math.* Springer–Verlag, New York, NY, 1998.

[266] S. Wang, A. Faghri, and T. L. Bergman. A comprehensive numerical model for melting with natural convection. *Int. J. Heat Mass Transfer.*, 53(9-10):1986–200, 2010.

[267] S. Wang, H. Wang, and P. Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed DeepOnets. *Sci. Adv.*, 7(40):eabi8605, 2021.

[268] C. G. Webster. *Sparse grid stochastic collocation techniques for the numerical solution of partial differential equations with random input data.* PhD thesis, Florida State University, 2007.

[269] C. L. Winter and Daniel M. Tartakovsky. Groundwater flow in heterogeneous composite aquifers. *Water Resources Research*, 38(8), August 2002.

[270] Y. Xu, A. Narayan, H. Tran, and C. Webster. Analysis of the ratio of ℓ_1 and ℓ_2 norms in compressed sensing. *Appl. Comput. Harmon. Anal.*, 55:486–511, 2020.

[271] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using ℓ_1-minimization. *Int. J. Uncertain. Quantif.*, 2(3):279–293, 2012.

[272] L. Yan, Y. Shin, and D. Xiu. Sparse approximation using $\ell_1 - \ell_2$ minimization and its application to stochastic collocation. *SIAM J. Sci. Comput.*, 39(1):A229–A254, 2017.

[273] X. Yang and G. E. Karniadakis. Reweighted ℓ_1 minimization method for stochastic elliptic differential equations. *J. Comput. Phys.*, 248:87–108, 2013.

[274] X. Yang, H. Lei, N. A. Baker, and G. Lin. Enhancing sparsity of Hermite polynomial expansions by iterative rotations. *J. Comput. Phys.*, 307:94–109, 2016.

[275] X. Yang, W. Li, and A. Tartakovsky. Sliced-inverse-regression–aided rotated compressive sensing method for uncertainty quantification. *SIAM/ASA J. Uncertain. Quantif.*, 6(4):1532–1554, 2018.

[276] X. Yang, X. Wan, L. Lin, and H. Lei. A general framework for enhancing sparsity of generalized polynomial chaos expansions. *Int. J. Uncertain. Quantif.*, 9(3):221–243, 2019.

[277] T. Zhou, A. Narayan, and D. Xiu. Weighted discrete least-squares polynomial approximation using randomized quadratures. *J. Comput. Phys.*, 298:787–800, 2015.

[278] T. Zhou, A. Narayan, and Z. Xu. Multivariate discrete least-squares approximations with a new type of collocation grid. *SIAM J. Sci. Comput.*, 36(5):A2401–A2422, 2014.

[279] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. 2008.

Appendix A

Holomorphic maps for mixed formulations

In Chapter 7, §7.4, we introduced a parametric DE whose weak solution was obtained via a mixed formulation. The following appendix section includes a variety of results and extensions proving that the solution map $\boldsymbol{y} \mapsto (\boldsymbol{\sigma}, u)(\boldsymbol{y})$ of (7.4.4)–(7.4.5) admits a holomorphic extension to an open neighbourhood $\mathcal{O}$ such that $\mathcal{E}_{\boldsymbol{\rho}} \subseteq \mathcal{O}$ for some $\boldsymbol{\rho} \geq \mathbf{1}$. This shows an example of a non-trivial parametric PDE whose solution is Hilbert-valued and satisfies Assumption 2.3.4. The analysis of the holomorphic extension of the solution map in this appendix is based on [12, §4], and the analysis of the mixed formulation is based on [114] (see also [108, 120]).

A.1 Babuska-Brezzi theory for real and complex variables

First we introduce the general framework and notation to study mixed formulations. Let $(H, \langle\cdot,\cdot\rangle_H)$ and $(Q, \langle\cdot,\cdot\rangle_Q)$ be Hilbert spaces over $\mathbb{C}$. We say that $a : H \times H \to \mathbb{C}$ and $b : H \times Q \to \mathbb{C}$ are bounded sesquilinear operators if they are linear with respect to their first argument, antilinear in their second argument, i.e.,

$$\begin{aligned} a(\sigma, c_1\tau_1 + c_2\tau_2) &= \overline{c_1}a(\sigma,\tau_1) + \overline{c_2}a(\sigma,\tau_2) \quad \forall \sigma,\tau_1,\tau_2 \in H, \quad \forall c_1, c_2 \in \mathbb{C} \\ b(\sigma, c_1u_1 + c_2u_2) &= \overline{c_1}b(\sigma,u_1) + \overline{c_2}b(\sigma,u_2) \quad \forall \sigma \in H, \quad \forall u_1,u_2 \in Q, \quad \forall c_1, c_2 \in \mathbb{C} \end{aligned}$$

and if there exists $M_a, M_b > 0$ such that

$$|a(\sigma,\tau)| \leq M_a\|\sigma\|_H\|\tau\|_H \quad \forall \sigma,\tau \in H \quad \text{and} \quad |b(\sigma,u)| \leq M_b\|\sigma\|_H\|u\|_Q, \quad \forall \sigma \in H,\, u \in Q.$$

Let $A : H \to H$ and $B : H \to Q$ be the linear operators induced by a, b. That is,

$$a(\sigma,\tau) = \langle A(\sigma),\tau\rangle_H \quad \forall \sigma,\tau \in H \quad \text{and} \quad b(\sigma,u) = \langle B(\sigma),u\rangle_Q \quad \forall \sigma \in H,\, u \in Q.$$

Note that these operators can be defined through the Riesz operator [108, Thm. A.16] (see also [50, Thm. 5.5]) and they satisfy the following identity:

$$a(\sigma,\tau) = \langle \sigma, A'(\tau)\rangle_H \quad \forall \sigma,\tau \in H \quad \text{and} \quad b(\sigma,u) = \langle \sigma, B'(u)\rangle_H, \quad \forall \sigma \in H,\, u \in Q,$$

where $A' : H \to H$ and $B' : Q \to H$ are the adjoint operators of A and B, respectively. Now consider the problem: find $(\sigma, u) \in H \times Q$ such that

$$\begin{aligned} a(\sigma,\tau) + b(\tau,u) &= G(\tau) \qquad \forall \tau \in H, \\ b(\sigma,v) &= F(v) \qquad \forall v \in Q, \end{aligned}$$

where $F \in Q^*$ and $G \in H^*$ are the continuous duals of Q and H, respectively. Then, using the Riesz operator once more, we can rewrite this problem as: find $(\sigma, u) \in H \times Q$ such that

$$\begin{aligned} A(\sigma) + B'(u) &= \mathcal{R}_H(G), \\ B(\sigma) &= \mathcal{R}_Q(F), \end{aligned}$$

where $\mathcal{R}_H : H^* \to H$ and $\mathcal{R}_Q : Q^* \to Q$ are the respective Riesz operators.

A.1.1 The inf-sup condition for b

In this section we consider the inf-sup condition for b. For more details on other frameworks we refer the reader to [50, 108]. Let $b : H \times Q \to \mathbb{C}$ be the sesquilinear form defined above. Then, we say that b satisfies the *continuous inf-sup condition* if there exists a $\beta > 0$ such that

$$\inf_{\substack{v \in Q \\ v \neq 0}} \sup_{\substack{\tau \in H \\ \tau \neq 0}} \frac{|b(\tau,v)|}{\|\tau\|_H \|v\|_Q} \geq \beta.$$

Notice that using the adjoint operator we obtain

$$\sup_{\substack{\tau \in H \\ \tau \neq 0}} \frac{|b(\tau,v)|}{\|\tau\|_H} = \sup_{\substack{\tau \in H \\ \tau \neq 0}} \frac{|\langle B'(v), \tau\rangle|}{\|\tau\|_H} = \|B'(v)\|_H, \quad \forall v \in Q$$

and therefore the inf-sup condition is equivalent to

$$\|B'(v)\|_H \geq \beta \|v\|_Q, \quad \forall v \in Q.$$

The following result establishes an equivalent condition to the previous inequalities. Its proof can be found in [114, Lem 2.1] for the real-valued case. A simple inspection of the proof reveals that it is also valid in the complex-valued case.

Lemma A.1.1. *Let $b : H \times Q \to \mathbb{C}$ be a bounded sesquilinear form with induced operator $B : H \to Q$. Then the following statements are equivalent.*

i) There exists $\beta > 0$ such that

$$\sup_{\substack{\tau\in H\\ \tau\neq 0}} \frac{|b(\tau,v)|}{\|\tau\|_H} \geq \beta\|v\|_Q, \quad \forall v \in Q.$$

ii) The operator $B' : Q \to N(B)^\perp$ is an isomorphism (linear bijection), and

$$\|B'(v)\|_H \geq \beta\|v\|_Q, \quad \forall v \in Q.$$

iii) The operator $B : N(B)^\perp \to Q$ is an isomorphism (linear bijection), and

$$\|B(\tau)\|_H \geq \beta\|\tau\|_H, \quad \forall \tau \in N(B)^\perp.$$

iv) The operator $B : H \to Q$ is surjective.

A.1.2 The BNB theorem for complex-valued Hilbert spaces

The following saddle-point structure is taken from [108, §2.4]. In particular, the real-valued case is implied by [108, Thm. 2.34]. However, a simple inspection to the proof of [114, Thm. 2.2–2.3] reveals that its real-valued version can be extended to the complex-valued case following similar arguments. We present its proof for completeness.

Theorem A.1.2 (Complex-valued Babuška-Nečas-Brezzi (BNB) theorem)**.** *Let $(H, \langle\cdot,\cdot\rangle_H)$ and $(Q, \langle\cdot,\cdot\rangle_Q)$ be Hilbert spaces over $\mathbb{C}$, $a : H \times H \to \mathbb{C}$ and $b : H \times Q \to \mathbb{C}$ be bounded sesquilinear operators with induced linear forms $A : H \to H$ and $B : H \to Q$, such that*

$$\begin{aligned} a(\sigma,\tau) = \langle A(\sigma),\tau\rangle_H, \quad &\text{and} \quad b(\sigma,u) = \langle B(\sigma),u\rangle_Q, \\ |a(\sigma,\tau)| \leq \|A\|\|\sigma\|_H\|\tau\|_H \quad &\text{and} \quad |b(\sigma,u)| \leq \|B\|\|\sigma\|_H\|u\|_Q \end{aligned}$$

for all $\sigma, \tau \in H$, and $u \in Q$. Let $V := N(B)$ and assume that

i) there exists $\alpha > 0$ such that

$$|a(\tau,\tau)| \geq \alpha\|\tau\|_H^2, \quad \forall \tau \in V,$$

(this is known as the V-ellipticity condition)

ii) there exists $\beta > 0$ such that

$$\sup_{\substack{\tau\in H\\ \tau\neq 0}} \frac{|b(\tau,v)|}{\|\tau\|_H} \geq \beta\|v\|_Q, \quad \forall v \in Q. \tag{A.1.1}$$

Then for each pair $(G, H) \in (H^* \times Q^*)$ there exists a unique solution $(\sigma, u) \in H \times Q$ to the problem

$$\begin{aligned} a(\sigma,\tau) + b(\tau,u) &= G(\tau) \qquad \forall \tau \in H, \\ b(\sigma,v) &= F(v) \qquad \forall v \in Q, \end{aligned}$$

that satisfyies

$$\|u\|_Q \leq \frac{\alpha + \|A\|}{\alpha\beta}\left(\|G\|_{H^*} + \frac{\|A\|}{\beta}\|F\|_{Q^*}\right)$$
$$\|\sigma\|_H \leq \frac{1}{\alpha}\left(\|G\|_{H^*} + \frac{(\alpha + \|A\|)}{\alpha\beta}\|F\|_{Q^*}\right)$$

Proof. We first reformulate the problem as follows: find $(\sigma, u) \in H \times Q$ such that

$$A(\sigma) + B'(u) = \mathcal{R}_H(G)$$
$$B(\sigma) = \mathcal{R}_Q(F),$$

where $\mathcal{R}_H : H^* \to H$ and $\mathcal{R}_Q : Q^* \to Q$ are the respective Riesz operators. Notice that $R_Q(F) \in Q$ and from (A.1.1) we know that B is an isomorphism from $V^\perp$ into Q. Therefore exists $\sigma_g \in N(B)^\perp = V^\perp \subseteq H$ such that

$$B(\sigma_g) = \mathcal{R}_Q(F),$$

and

$$\|\sigma_g\|_H \leq \frac{1}{\beta}\|B(\sigma_g)\|_Q = \frac{1}{\beta}\|F\|_{Q^*}.$$

Next, let $\Pi : H \to V$ be the orthogonal projector from H to V, and $\sigma_g \in H$ be as defined above. Then, $\Pi(\mathcal{R}_H(G) - A(\sigma_g)) \in V$. Since a is a sesquilinear V-elliptic operator, the complex-valued Lax-Milgram lemma in [224, Rmk. 5.1.2] implies that there exists a unique $\sigma_0 \in V$ such that

$$A(\sigma_0) = \Pi(\mathcal{R}_H(G) - A(\sigma_g)) \tag{A.1.2}$$

and

$$\|\sigma_0\|_H \leq \frac{1}{\alpha}\|\Pi(\mathcal{R}_H(G) - A(\sigma_g))\|_H \leq \frac{1}{\alpha}\left(\|G\|_{H^*} + \frac{\|A\|}{\beta}\|F\|_{Q^*}\right).$$

Observe that, by definition, $A(\sigma_0) \in V$ and therefore $\Pi(\mathcal{R}_H(G) - A(\sigma_g + \sigma_0)) = 0$. This implies that $\mathcal{R}_H(G) - A(\sigma_g + \sigma_0) \in V^\perp$. Since B is an isomorphism from $V^\perp$ to Q, there exists $u \in Q$ such that $B'(u) = \mathcal{R}_H(G) - A(\sigma_g + \sigma_0)$, that is

$$A(\sigma_g + \sigma_0) + B'(u) = \mathcal{R}_H(G).$$

This u satisfies

$$\|u\|_Q \leq \frac{1}{\beta}\|B'(u)\|_H = \frac{1}{\beta}\|\mathcal{R}_H(G) - A(\sigma_g + \sigma_0)\|_H \leq \frac{\alpha + \|A\|}{\alpha\beta}\left(\|G\|_{H^*} + \frac{\|A\|}{\beta}\|F\|_{Q^*}\right),$$

as required. Defining $\sigma = \sigma_g + \sigma_0 \in H$ and noticing that $B(\sigma_0) = 0$ conclude the existence of a desired pair (σ, u). It remains to show uniqueness. This is implied by the solution of the homogeneous problem, since it follows the same arguments as [114, Theorem 2.1] we omit its proof. □

A.2 The parametric diffusion equation

This section is mainly based on [12, §4.2]. Specifically, we use the same arguments to prove that the mixed formulation of the non-homogeneous Poisson problem is well-defined and that it has a holomorphic extension to an open neighborhood of a suitable filled-in Bernstein polyellipse. First, consider a bounded Lipschitz domain $\Omega \subset \mathbb{R}^2$, let $\partial\Omega$ be the boundary of Ω, $f \in \mathrm{L}^2(\Omega;\mathbb{R})$ and $g \in \mathrm{H}^{1/2}(\partial\Omega)$. Here $\mathrm{H}^{1/2}(\partial\Omega)$ is the trace space on the boundary $\partial\Omega$, that is,

$$\mathrm{H}^{1/2}(\partial\Omega) := \gamma_0(\mathrm{H}^1(\Omega)),$$

were $\gamma_0 : \mathrm{H}^1(\Omega) \to \mathrm{L}^2(\partial\Omega)$ is the bounded linear operator such that $\gamma_0(\varphi) = \varphi|_\gamma$ for all φ in the space of restrictions to Ω of functions that are of class C_0^∞ in an open set containing $\overline{\Omega}$. Note that the space $(\mathrm{H}^{1/2}(\partial\Omega), \|\cdot\|_{1/2,\partial\Omega})$ is complete, where

$$\|g\|_{1/2,\partial\Omega} := \inf\{\|v\|_{\mathrm{H}^1(\Omega)} : v \in \mathrm{H}^1(\Omega) \text{ such that } \gamma_0(v) = g\}, \quad \forall g \in \mathrm{H}^{1/2}(\partial\Omega).$$

For a more complete review on trace spaces we refer to [114, §1.3.2].

Now, consider the linear elliptic equation with Dirichlet boundary conditions

$$\begin{aligned} -\mathrm{div}(a(\boldsymbol{x},\boldsymbol{y})\nabla u(\boldsymbol{x},\boldsymbol{y})) &= f(\boldsymbol{x}), \quad \text{in } \Omega \\ u(\boldsymbol{x},\boldsymbol{y}) &= g(\boldsymbol{x}), \quad \text{on } \partial\Omega. \end{aligned} \tag{A.2.1}$$

The variable $\boldsymbol{y} \in \mathcal{U} = [-1,1]^{\mathbb{N}}$, the coefficient $a(\boldsymbol{x},\boldsymbol{y})$ is parametric, the term $f(\boldsymbol{x})$ is nonparametric and $g(\boldsymbol{x})$ is nonparametric as well. Our main goal is to study the regularity of the parametric map $\boldsymbol{y} \mapsto u(\cdot,\boldsymbol{y})$. With a slight abuse of notation, we sometimes switch between the notation $u(\cdot,\boldsymbol{y})$ and $u(\boldsymbol{y})$ when referring to the parametric solution map.

Our first step is to identify sufficient considitions on $\boldsymbol{y} \mapsto a(\boldsymbol{y})$ for the map $\boldsymbol{y} \mapsto u(\boldsymbol{y})$ to be well defined. Now we turn our attention to the weak mixed variational formulation of the elliptic problem in (A.2.1).

First, assume that there exists $r, M > 0$, independent of $\boldsymbol{y}$ such that

$$0 < r \le \operatorname*{ess\,inf}_{\boldsymbol{x}\in\Omega} a(\boldsymbol{x},\boldsymbol{y}) = a_{\min}(\boldsymbol{y}), \text{ and } a_{\max}(\boldsymbol{y}) = \operatorname*{ess\,sup}_{\boldsymbol{x}\in\Omega} a(\boldsymbol{x},\boldsymbol{y}) \le M, \quad \forall \boldsymbol{y} \in \mathcal{U}. \tag{A.2.2}$$

Next, for any value of the parameter $\boldsymbol{y} \in \mathcal{U}$, define the additional unknown $\boldsymbol{\sigma}(\boldsymbol{y}) = a(\boldsymbol{y})\nabla u(\boldsymbol{y})$ in Ω. Then the problem can be stated as a first-order system: given $\boldsymbol{y} \in \mathcal{U}$, find $(\boldsymbol{\sigma}, u)(\boldsymbol{y})$ such that

$$\begin{aligned} a^{-1}(\boldsymbol{y})\boldsymbol{\sigma}(\boldsymbol{y}) &= \nabla u(\boldsymbol{y}), \quad \text{in } \Omega, \\ -\mathrm{div}(\boldsymbol{\sigma}(\boldsymbol{y})) &= f, \quad \text{in } \Omega, \\ u(\boldsymbol{y}) &= g, \quad \text{on } \partial\Omega. \end{aligned}$$

Multiplying the first equation by $\boldsymbol{\tau} \in \mathbf{H}(\mathrm{div};\Omega)$, and applying Green's identity [114, Lem. 1.4], we get

$$\langle a^{-1}\boldsymbol{\sigma}, \boldsymbol{\tau}\rangle_{\mathrm{L}^2(\Omega)} + \langle u, \mathrm{div}(\boldsymbol{\tau})\rangle_{\mathrm{L}^2(\Omega)} = \langle \gamma_0(\boldsymbol{\tau})\cdot\boldsymbol{n}, g\rangle_{\mathrm{H}^{-1/2}(\partial\Omega)\times\mathrm{H}^{1/2}(\partial\Omega)},$$

where $\boldsymbol{n}$ is the outward normal vector to $\partial\Omega$, $\langle\cdot,\cdot\rangle_{\mathrm{H}^{-1/2}(\partial\Omega)\times\mathrm{H}^{1/2}(\partial\Omega)}$ denotes the duality pair between $\mathrm{H}^{1/2}(\partial\Omega)$ and its dual $\mathrm{H}^{-1/2}(\partial\Omega)$ with respect to the inner product of $\mathrm{L}^2(\partial\Omega)$. For simplicity, we write $H = \mathbf{H}(\mathrm{div};\Omega)$, $Q = \mathrm{L}^2(\Omega)$, and $\langle\cdot,\cdot\rangle_{\mathrm{H}^{-1/2}(\partial\Omega)\times\mathrm{H}^{1/2}(\partial\Omega)}$ as $\langle\cdot,\cdot\rangle_{1/2,\partial\Omega}$. Next, multiplying the second equation by $v \in Q$ we get the mixed variational formulation: find $(\boldsymbol{\sigma}, u) \in H \times Q$ such that

$$\begin{aligned} d_a(\boldsymbol{\sigma},\boldsymbol{\tau}) + b(\boldsymbol{\tau},u) &= G(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in H, \\ b(\boldsymbol{\sigma},v) &= F(v), \quad \forall v \in Q, \end{aligned} \tag{A.2.3}$$

where d and b are the bilinear forms defined by

$$\begin{aligned} d_a(\boldsymbol{\sigma},\boldsymbol{\tau}) &= \langle a^{-1}\boldsymbol{\sigma},\boldsymbol{\tau}\rangle_{\mathrm{L}^2(\Omega)}, \quad \forall(\boldsymbol{\tau},\boldsymbol{\sigma}) \in H \times H \\ b(\boldsymbol{\tau},v) &= \langle \mathrm{div}(\boldsymbol{\tau}), v\rangle_{\mathrm{L}^2(\Omega)}, \quad \forall(\boldsymbol{\tau},v) \in H \times Q \end{aligned}$$

and the functionals $G \in H^*$ and $F \in Q^*$ are defined by

$$F(v) = \langle -f, v\rangle_{\mathrm{L}^2(\Omega)}, \quad \forall v \in Q, \text{ and } G(\boldsymbol{\tau}) = \langle \gamma_0(\boldsymbol{\tau})\cdot\boldsymbol{n}, g\rangle_{1/2,\partial\Omega}, \quad \forall \boldsymbol{\tau} \in H.$$

Now, we show the well-posedness of this variational formulation. Observe that this requires a real-valued version of the BNB theorem. However, a simple inspection reveals that this is implied by the complex version above. For the general version of this theorem in real-valued Banach spaces see [120, Theorem 4.1] (see also [114, Theorem 2.1] for the real-valued case in Hilbert spaces).

Theorem A.2.1 (Poisson equation; real-valued case)**.** *Suppose that $a \in L^\infty(\mathcal{U};\mathrm{L}^\infty(\Omega;\mathbb{R}))$ is bounded above and below by positive constants r, M, i.e.,*

$$0 < r \le \operatorname*{ess\,inf}_{\boldsymbol{x}\in\Omega} a(\boldsymbol{x},\boldsymbol{y}) = a_{\min}(\boldsymbol{y}), \text{ and } a_{\max}(\boldsymbol{y}) = \operatorname*{ess\,sup}_{\boldsymbol{x}\in\Omega} a(\boldsymbol{x},\boldsymbol{y}) \le M, \quad \forall \boldsymbol{y} \in \mathcal{U}. \tag{A.2.4}$$

Given $\boldsymbol{y} \in \mathcal{U}$, consider the mixed formulation in (A.2.3). Then, for any $f \in \mathrm{L}^2(\Omega)$, $g \in \mathrm{H}^{1/2}(\partial\Omega)$, there exists a unique solution $(\boldsymbol{\sigma}(\boldsymbol{y}), u(\boldsymbol{y})) \in H \times Q$ to the problem

$$\begin{aligned} d_a(\boldsymbol{\sigma},\boldsymbol{\tau}) + b(\boldsymbol{\tau},u) &= G(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in H, \\ b(\boldsymbol{\sigma},v) &= F(v), \quad \forall v \in Q, \end{aligned} \tag{A.2.5}$$

and this solution satisfies

$$\begin{aligned} \|u\|_Q &\le \frac{r+M}{r\beta}\left(\|g\|_{1/2,\partial\Omega} + \frac{1}{r\beta}\|f\|_{\mathrm{L}^2(\Omega)}\right), \\ \|\boldsymbol{\sigma}\|_H &\le M\left(\|g\|_{1/2,\partial\Omega} + \frac{(r+M)}{r\beta}\|f\|_{\mathrm{L}^2(\Omega)}\right). \end{aligned}$$

Proof. This proof is based on [114, §2.4.1]. First, observe that (A.2.4) implies

$$0 < M^{-1} \le \operatorname*{ess\,inf}_{\boldsymbol{x}\in\Omega} a^{-1}(\boldsymbol{x},\boldsymbol{y}) = a^{-1}_{\min}(\boldsymbol{y}), \text{ and } a^{-1}_{\max}(\boldsymbol{y}) = \operatorname*{ess\,sup}_{\boldsymbol{x}\in\Omega} a^{-1}(\boldsymbol{x},\boldsymbol{y}) \le r^{-1}, \quad \forall \boldsymbol{y} \in \mathcal{U}.$$

Naturally d, b are bounded bilinear forms and F, G are bounded linear functionals, with bounds

$$|d_a(\boldsymbol{\tau}, \boldsymbol{\sigma})| \leq r^{-1}\|\boldsymbol{\tau}\|_H\|\boldsymbol{\sigma}\|_H, \quad |b(\boldsymbol{\tau}, v)| \leq \|\boldsymbol{\tau}\|_H\|v\|_Q,$$
$$|F(v)| \leq \|f\|_{\mathrm{L}^2(\Omega)}\|v\|_Q, \quad |G(\boldsymbol{\tau})| \leq \|g\|_{1/2,\gamma}\|\boldsymbol{\tau}\|_H,$$

for all $\boldsymbol{\tau}, \boldsymbol{\sigma} \in H$ and $v \in Q$, respectively. Now let $B : H \to Q$ be the induced operator of b given by $B(\boldsymbol{\tau}) = \mathrm{div}(\boldsymbol{\tau})$ for all $\boldsymbol{\tau} \in H$. Hence

$$V = N(B) = \{\boldsymbol{\tau} \in H : \mathrm{div}(\boldsymbol{\tau}) = 0 \text{ in } \Omega\}.$$

This implies that the operator d_a is V-elliptic. That is

$$d_a(\boldsymbol{\tau}, \boldsymbol{\tau}) = \langle a^{-1}\boldsymbol{\tau}, \boldsymbol{\tau}\rangle_{\mathrm{L}^2(\Omega)} \geq M^{-1}\langle \boldsymbol{\tau}, \boldsymbol{\tau}\rangle_{\mathrm{L}^2(\Omega)} = M^{-1}\|\boldsymbol{\tau}\|^2_{\mathrm{L}^2(\Omega)} = M^{-1}\|\boldsymbol{\tau}\|^2_H, \quad \forall \boldsymbol{\tau} \in V.$$

Now we need to prove the surjectivity of the operator B, which is equivalent to the inf-sup condition (A.1.1). Let $v \in Q$ and consider the auxiliary variational problem

$$\begin{aligned} -\Delta z &= v, \quad \text{in } \Omega \\ z &= 0, \quad \text{on } \Omega. \end{aligned}$$

Due to Lax-Milgram lemma there exists a unique $z \in \mathrm{H}^1_0(\Omega)$ such that

$$\|\nabla z\|_{\mathbf{L}^2(\Omega)} \leq C\|v\|_Q,$$

where $C > 0$ is a constant depending on the Friedrich-Poincaré inequality [114, Lem. 1.1]. Then, defining $\widetilde{\boldsymbol{\tau}} = -\nabla z$ in Ω we deduce $\widetilde{\boldsymbol{\tau}} \in H$ and $B(\widetilde{\boldsymbol{\tau}}) = v$ in Ω. Therefore B is surjective. Using the real-valued version of the BNB theorem we get the result. □

In summary, under the sufficient conditions of Theorem A.2.1, the parametric solution map

$$(\boldsymbol{\sigma}, u) : \mathcal{U} \to H \times Q,$$

is well-defined, Hilbert-valued mapping belonging to the space $L^\infty(\mathcal{U}, H \times Q)$.

The complex-valued version of this result follows the similar arguments to those of the real-valued version.

Corollary A.2.2 (Poisson equation; complex-valued case)**.** *Suppose that* $a \in L^\infty(\mathcal{U}; \mathrm{L}^\infty(\Omega; \mathbb{C}))$ *is bounded below and above by positive constants* r, M*, i.e.,*

$$0 < r \leq \operatorname*{ess\,inf}_{\boldsymbol{x}\in\Omega} \mathrm{Re}\,(a(\boldsymbol{x}, \boldsymbol{y})) = a_{\min}(\boldsymbol{y}), \text{ and } a_{\max}(\boldsymbol{y}) = \operatorname*{ess\,sup}_{\boldsymbol{x}\in\Omega} |a(\boldsymbol{x}, \boldsymbol{y})| \leq M, \quad \forall \boldsymbol{y} \in \mathcal{U}. \tag{A.2.6}$$

Given $\boldsymbol{y} \in \mathcal{U}$*, consider the mixed formulation in* (A.2.3)*. Then, for any* $f \in \mathrm{L}^2(\Omega; \mathbb{C})$*,* $g \in \mathrm{H}^{1/2}(\partial\Omega; \mathbb{C})$*, there exists a unique solution* $(\boldsymbol{\sigma}(\boldsymbol{y}), u(\boldsymbol{y})) \in (H \times Q) = (\mathbf{H}(\mathrm{div}; \Omega; \mathbb{C}) \times \mathrm{L}^2(\Omega; \mathbb{C}))$ *to the problem*

$$\begin{aligned} d_a(\boldsymbol{\sigma}, \boldsymbol{\tau}) + b(\boldsymbol{\tau}, u) &= G(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in H, \\ b(\boldsymbol{\sigma}, v) &= F(v), \quad \forall v \in Q, \end{aligned} \tag{A.2.7}$$

and this solution satisfies

$$\|u\|_Q \leq \frac{1+(r/M)^2}{\beta}\left(\|g\|_{1/2,\partial\Omega} + \frac{1}{r\beta}\|f\|_{\mathrm{L}^2(\Omega)}\right),$$
$$\|\boldsymbol{\sigma}\|_H \leq \frac{M^2}{r}\left(\|g\|_{1/2,\partial\Omega} + \frac{(1+(r/M)^2)}{\beta}\|f\|_{\mathrm{L}^2(\Omega)}\right).$$

Proof. Under the assumption (A.2.6), we get

$$\begin{aligned}|d_a(\boldsymbol{\tau},\boldsymbol{\tau})| \geq \mathrm{Re}\,(d_a(\boldsymbol{\tau},\boldsymbol{\tau})) &= \int_\Omega \mathrm{Re}\,(a^{-1})|\boldsymbol{\tau}|^2 \\ &= \int_\Omega \frac{\mathrm{Re}\,(a)}{\mathrm{Re}^{\,2}(a)+\mathrm{Im}^{\,2}(a)}|\boldsymbol{\tau}|^2 \\ &\geq \frac{r}{M^2}\|\boldsymbol{\tau}\|_{\mathrm{L}^2(\Omega)} = \frac{r}{M^2}\|\boldsymbol{\tau}\|_H, \quad \forall \boldsymbol{\tau} \in N(B).\end{aligned}$$

Moreover, the first inequality implies that

$$|d_a(\boldsymbol{\sigma},\boldsymbol{\tau})| = |\langle a^{-1}(\boldsymbol{y})\boldsymbol{\sigma},\boldsymbol{\tau}\rangle_{\mathrm{L}^2(\Omega)}| \leq \frac{1}{r}\|\boldsymbol{\sigma}\|_H\|\boldsymbol{\tau}\|_H \quad \forall \boldsymbol{\tau},\boldsymbol{\sigma} \in H.$$

Replacing $\frac{r}{M^2}$ and $1/r$ instead of α and $\|A\|$ in Theorem A.1.2, and following the same steps as in Theorem A.2.1 we get the result. □

A.2.1 Affine parametric dependence

As in [12, §4.2.1], we now assume that there exists functions $a_0 \in \mathrm{L}^\infty(\Omega)$ and $\{\psi_j\}_{j\in\mathbb{N}} \subset \mathrm{L}^\infty(\Omega)$ such that

$$a(\boldsymbol{x},\boldsymbol{y}) = a_0(\boldsymbol{x}) + \sum_{j\in\mathbb{N}} y_j\psi_j(\boldsymbol{x}), \quad \forall x \in \Omega, \forall \boldsymbol{y} \in \mathcal{U}. \tag{A.2.8}$$

Under the parametric dependence assumption, the uniform ellipticity condition (A.2.6) is equivalent to

$$\sum_{j\in\mathbb{N}} |\psi_j(\boldsymbol{x})| \leq a_0(\boldsymbol{x}) - r, \quad \forall \boldsymbol{x} \in \Omega,$$

for some $r > 0$. This implies the absolute convergence of the series $\sum_{j\in\mathbb{N}} y_j\psi_j(\boldsymbol{x})$ for every $\boldsymbol{x} \in \Omega$ and $\boldsymbol{y} \in \mathcal{U}$ and therefore (A.2.6) holds wit $M = 2\|a_0\|_{\mathrm{L}^\infty(\Omega)}$.

A.3 Holomorphic extension of the solution map

We now follow the same arguments as those in [12, §4.2.2] to prove the holomorphic extension of the solution map. First, we generalize the definition in §2.3 to holomorphic maps between Banach spaces [12, Def. 4.7].

Definition A.3.1 (holomorphic mapping between Banach spaces)**.** Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces and $\mathcal{O} \subseteq X$ be an open set. A mapping $\mathfrak{F} : X \to Y$ *is*

holomorphic in $\mathcal{O}$ if, for every $x \in \mathcal{O}$, $\mathfrak{F}$ has a Fréchet derivative $\mathrm{d}\mathfrak{F}(x)$, i.e., there exists a bounded linear operator $\mathrm{d}\mathfrak{F}(x) : X \to Y$ that satisfies

$$\lim_{\|h\|_X \to 0} \frac{\|\mathfrak{F}(x+h) - \mathfrak{F}(x) - \mathrm{d}\mathfrak{F}(x)h\|_Y}{\|h\|_X} = 0.$$

We now show that the solution map $\boldsymbol{y} \mapsto (\boldsymbol{\sigma}(\boldsymbol{y}), u(\boldsymbol{y}))$ of (A.2.7) with diffusion coefficient $a = a(\cdot, \boldsymbol{y})$ given by (A.2.8) admits a holomorphic extension to an open neighborhood of a suitable filled-in Bernstein polyellipse and therefore satisfies the holomorphic extension assumption, Assumption 2.3.2.

Step 1. The holomorphic extension of $a \to (\boldsymbol{\sigma}, u)(a)$. We first consider the solution $(\boldsymbol{\sigma}, u)$ as a function of the diffusion coefficient $a \in \mathrm{L}^\infty(\Omega; \mathbb{R})$ and study the solution map $a \mapsto (\boldsymbol{\sigma}, u)(a)$. This map is an operator between the Banach spaces $\mathrm{L}^\infty(\Omega; \mathbb{R})$ and $H \times Q = \mathbf{H}(\mathrm{div}; \Omega; \mathbb{R}) \times \mathrm{L}^2(\Omega; \mathbb{R})$, which is well defined for those a satisfying (A.2.4). The main result below shows that this mapping admits a holomorphic extension to the open region

$$\mathcal{O}^{\mathrm{UE}} = \bigcup_{\substack{r>0 \\ M>0}} \{a \in \mathrm{L}^\infty(\Omega; \mathbb{C}) : \operatorname{ess\,inf}_{\boldsymbol{x} \in \Omega} \mathrm{Re}(a(\boldsymbol{x})) \geq r, \|a\|_{\mathrm{L}^\infty(\Omega, \mathbb{C})} < M\} \subset \mathrm{L}^\infty(\Omega, \mathbb{C}).$$

Proof of step 1. Let $H \times Q = \mathbf{H}(\mathrm{div}; \Omega; \mathbb{R}) \times \mathrm{L}^2(\Omega; \mathbb{R})$ and $f \in Q$ and $g \in H$. Consider the parametric operator $\mathscr{L} : \mathrm{L}^\infty(\mathcal{U}) \to \mathcal{L}(H \times Q, H \times Q)$ given by

$$\mathscr{L}_{\mathcal{T}(a)} = \mathscr{L}(\mathcal{T}(a)) = \begin{bmatrix} D_{\mathcal{T}(a)} & B^* \\ B & 0 \end{bmatrix}$$

where $D_{\mathcal{T}(a)} : H \to H$ and $B : H \to Q$ are such that

$$\langle D_{\mathcal{T}(a)}(\boldsymbol{\sigma}), \boldsymbol{\tau} \rangle = \langle \mathcal{T}(a)\boldsymbol{\sigma}, \boldsymbol{\tau} \rangle_{\mathrm{L}^2(\Omega)}, \quad \langle B(\boldsymbol{\tau}), v \rangle = \langle \mathrm{div}(\boldsymbol{\tau}), v \rangle_{\mathrm{L}^2(\Omega)}. \tag{A.3.1}$$

for all $\mathcal{T}(a) = 1/a \in \mathrm{L}^\infty(\Omega)$ satisfying (A.2.2). Assuming that there exists $r, M > 0$ such that

$$0 < r \leq \operatorname*{ess\,inf}_{\boldsymbol{x} \in \Omega} a(\boldsymbol{x}, \boldsymbol{y}) = \widetilde{a}_{\min}(\boldsymbol{y}), \text{ and } a_{\max}(\boldsymbol{y}) = \operatorname*{ess\,sup}_{\boldsymbol{x} \in \Omega} a(\boldsymbol{x}, \boldsymbol{y}) \leq M, \quad \forall \boldsymbol{y} \in \mathcal{U},$$

the inverse $\mathscr{L}^{-1}_{\mathcal{T}(a)}$ is well defined and we can write

$$\mathscr{L}^{-1}_{\mathcal{T}(a)} \begin{bmatrix} f \\ g \end{bmatrix} =: \begin{bmatrix} \widetilde{\boldsymbol{\sigma}}(\mathcal{T}(a)) \\ \widetilde{u}(\mathcal{T}(a)) \end{bmatrix}.$$

Then, we can write the mapping $a \mapsto (\boldsymbol{\sigma}, u)(a)$ as

$$a \mapsto \mathcal{T}(a) \mapsto (\boldsymbol{\sigma}, u)(a) = (\tilde{\boldsymbol{\sigma}}, \tilde{u})(\mathcal{T}(a)),$$

where once more $\mathcal{T}(a) = 1/a$ for any $a \in \mathrm{L}^\infty(\Omega)$ satisfying (A.2.2). This composition will be useful later. Our first goal is to find conditions such that the mapping $a \mapsto (\boldsymbol{\sigma}, u)(a)$ can be extended to complex-valued diffusion coefficients $a \in \mathrm{L}^\infty(\Omega; \mathbb{C})$.

The complex version of the BNB theorem gives the following result. Let $a \in \mathrm{L}^\infty(\Omega; \mathbb{C})$, and assume that (A.2.6) holds for some $r > 0$ and $M > 0$. Then the operator $\mathscr{L}_{\mathcal{T}(a)}$ is invertible and

$$\|\mathscr{L}^{-1}_{\mathcal{T}(a)}(f, g)\|_{H \times Q} \leq C_1(M, r, \beta)\|f\|_Q + C_2(M, r, \beta)\|g\|_H,$$

for all $f \in Q$ and $g \in H$, where

$$C_1(M, r, \beta) = (1 + \beta M)\left(\frac{1 + (r/M)^2}{r\beta^2}\right), \quad C_2(M, r, \beta) = \frac{1 + (r/M)^2}{\beta} + \frac{M^2}{r}. \tag{A.3.2}$$

Hence, a natural way to extend $(\boldsymbol{\sigma}, u)$ is to consider $(\bar{\boldsymbol{\sigma}}, \bar{u})(\mathcal{T}(a)) = \mathscr{L}^{-1}_{\mathcal{T}(a)}(f, g)$. We are now ready to prove that $(\bar{\boldsymbol{\sigma}}, \bar{u})$ is holomorphic in a suitable open set.

Proposition A.3.2 (holomorphic extension of $a \mapsto (\boldsymbol{\sigma}, u)(a)$)**.** *The solution map*

$$(\boldsymbol{\sigma}, u) : \mathrm{L}^\infty(\mathcal{U}) \to \mathbf{H}(\mathrm{div}; \Omega) \times \mathrm{L}^2(\Omega)$$

associated with prolem (A.2.5) *admits a well-defined and holomorphic extension*

$$(\bar{\boldsymbol{\sigma}}, \bar{u}) : \mathcal{O}^{UE} \to \mathbf{H}(\mathrm{div}; \Omega; \mathbb{C}) \times \mathrm{L}^2(\Omega; \mathbb{C}),$$

where $\mathcal{O}^{UE} \subset \mathrm{L}^\infty(\Omega; \mathbb{C})$ *is the open region defined as*

$$\mathcal{O}^{UE} = \bigcup_{\substack{r>0 \\ M>0}} \mathcal{R}^{UE}_{r,M}, \text{ with } \mathcal{R}^{UE}_{r,M} = \{a \in \mathrm{L}^\infty(\Omega; \mathbb{C}) : \operatorname{ess\,inf}_{\boldsymbol{x} \in \Omega} \mathrm{Re}\,(a(\boldsymbol{x})) \geq r, \|a\|_{\mathrm{L}^\infty(\Omega;\mathbb{C})} \leq M\}.$$

Moreover, the following upper bound holds in each region $\mathcal{R}^{UE}_{r,M}$ *of uniform ellipticity:*

$$\|(\bar{\boldsymbol{\sigma}}, \bar{u})\|_{L^\infty(\mathcal{R}^{UE}_{r,M}; \mathbf{H}(\mathrm{div};\Omega;\mathbb{C}) \times \mathrm{L}^2(\Omega;\mathbb{C}))} \leq C_1(M, r, \beta)\|f\|_{\mathrm{L}^2(\Omega)} + C_2(M, r, \beta)\|g\|_{\mathrm{H}^{1/2}(\partial\Omega)}$$

where C_1, C_2 *are the constants in* (A.3.2).

Proof. We follow [12, Prop. 4.8] to prove that $\mathcal{O}^{\mathrm{UE}}$ is open. Let $a \in \mathcal{O}^{\mathrm{UE}}$. Then there exists $r, M > 0$ such that $\operatorname{ess\,inf}_{\boldsymbol{x} \in \Omega} \mathrm{Re}\,(a(\boldsymbol{x})) \geq r$ and $\|a\|_{\mathrm{L}^\infty(\Omega;\mathbb{C})} \leq M$. Now, let $0 < \epsilon < r$. For any $b \in \mathrm{L}^\infty(\Omega; \mathbb{C})$ such that $\|b - a\|_{\mathrm{L}^\infty(\Omega;\mathbb{C})} < \epsilon$, we have

$$\operatorname{ess\,inf}_{\boldsymbol{x} \in \Omega} \mathrm{Re}\,(b(\boldsymbol{x})) > \operatorname{ess\,inf}_{\boldsymbol{x} \in \Omega} \mathrm{Re}\,(a(\boldsymbol{x})) - \|b - a\|_{\mathrm{L}^\infty(\Omega;\mathbb{C})} > r - \epsilon > 0,$$

and $\|b\|_{\mathrm{L}^\infty(\Omega;\mathbb{C})} \leq \|a\|_{\mathrm{L}^\infty(\Omega;\mathbb{C})} + \|b - a\|_{\mathrm{L}^\infty(\Omega;\mathbb{C})} \leq M + \epsilon$. Hence $b \in \mathcal{O}^{\mathrm{UE}}$. This proves the claim.

We now show that $(\boldsymbol{\sigma}, u)$ admits a holomorphic extension over $\mathcal{O}^{\mathrm{UE}}$. First we decompose the map $a \mapsto (\bar{\boldsymbol{\sigma}}, \bar{u})(a)$ into the following concatenation of four maps:

$$\begin{array}{ccccccccc} \mathcal{O}^{\mathrm{UE}} & \to & \mathrm{L}^\infty(\Omega;\mathbb{C}) & \to & \mathcal{L}(H\times Q, H\times Q) & \to & \mathcal{L}(H\times Q, H\times Q) & \to & H\times Q \\ a & \mapsto & \mathcal{T}(a)=1/a & \mapsto & \mathscr{L}_{\mathcal{T}(a)} & \mapsto & \mathscr{L}^{-1}_{\mathcal{T}(a)} & \mapsto & \mathscr{L}^{-1}_{\mathcal{T}(a)}(f,g), \end{array}$$

where $\mathscr{L}^{-1}_{\mathcal{T}(a)}(f,g) = (\bar{\boldsymbol{\sigma}}, \bar{u})(a)$.

As in [12, Prop. 4.8], the rest of the proof is devoted to showing that $(\bar{\boldsymbol{\sigma}}, \bar{u})$ is holomorphic in $\mathcal{O}^{\mathrm{UE}}$ by veryfying that $\mathcal{T}$, $\mathscr{L}_{\mathcal{T}(a)}$, the inversion mapping and the evaluation mapping are holomorphic in suitable open domains.

The map $\mathcal{T}(a) = 1/a$

The mapping $\mathcal{T} : \mathcal{O}^{\mathrm{UE}} \to \mathrm{L}^\infty(\Omega;\mathbb{C})$ defined as $\mathcal{T}(a) = 1/a$ has a Frechet derivative $\mathrm{d}(\frac{1}{a})h = -h/a^2$ in the open domain $\mathcal{O}^{\mathrm{UE}}$. Therefore, by Definition A.3.1 it is holomorphic in $\mathcal{O}^{\mathrm{UE}}$.

The parametric diffusion operator $\mathscr{L}$

We now prove that the operator $\mathscr{L} : \mathrm{L}^\infty(\mathcal{U}) \to \mathcal{L}(H\times Q, H\times Q)$ is holomorphic, where

$$\mathscr{L}_a = \mathscr{L}(a) = \begin{bmatrix} D_a & B^* \\ B & 0 \end{bmatrix},$$

and $D_a : H \to H$ and $B : H \to Q$ are such that (A.3.1) holds for all $a \in \mathrm{L}^\infty(\Omega)$ satisfying (A.2.2).

First, notice that the Frechet derivative of $\mathscr{L}$ is given by

$$\mathrm{d}\mathscr{L}(a)(h) = \begin{bmatrix} D_h & 0 \\ 0 & 0 \end{bmatrix}.$$

In particular we have

$$\begin{aligned} &\|\mathscr{L}(x+h) - \mathscr{L}(x) - \mathrm{d}\mathscr{L}(x)h\|_{\mathcal{L}(H\times Q, H\times Q)} \\ &= \left\| \begin{bmatrix} D_{(x+h)} & B^* \\ B & 0 \end{bmatrix} - \begin{bmatrix} D_x & B^* \\ B & 0 \end{bmatrix} - \begin{bmatrix} D_h & 0 \\ 0 & 0 \end{bmatrix} \right\|_{\mathcal{L}(H\times Q, H\times Q)} = \left\| \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right\|_{\mathcal{L}(H\times Q, H\times Q)} = 0. \end{aligned}$$

Therefore $\|\mathscr{L}(x+h) - \mathscr{L}(x) - \mathrm{d}\mathscr{L}(x)h\|_{\mathcal{L}(H\times Q, H\times Q)} = 0$ and by Definition A.3.1, the operator $\mathscr{L}$ is holomorphic in $\mathrm{L}^\infty(\Omega, \mathbb{C})$.

We refer to [12, Prop. 4.8], which proves that the inverse map is holomorphic. In particular, the operators in $\mathscr{L}(\mathcal{T}(\mathcal{R}^{\mathrm{UE}}_{r,M}))$ are invertible due to Theorem A.1.2. Moreover, the inverse mapping is well defined and holomorphic in the open set $\mathscr{L}(\mathcal{O}^{\mathrm{UE}}) = \bigcup_{\substack{r>0 \\ M>0}} \mathscr{L}(\mathcal{T}(\mathcal{R}^{\mathrm{UE}}_{r,M}))$.

As in [12, Prop. 4.8], the evaluation mapping is linear and hence holomorphic, and that the composition of holomorphic maps is holomorphic. We conclude that $a \mapsto (\bar{\boldsymbol{\sigma}}, \bar{u})(a)$ is

holomorphic in $\mathcal{O}^{\mathrm{UE}}$. Furthermore, the upper bound follows from applying the uniform bound in Theorem A.1.2. □

We now show that the map $\boldsymbol{y} \mapsto (\boldsymbol{\sigma}, u)(\boldsymbol{y})$ admits a complex holomorphic extension to an open neighborhood of the filled-in Bernstein polyellipse $\mathcal{E}_{\boldsymbol{\rho}}$ provided $\boldsymbol{\rho}$ satisfies a suitable summability condition.

Proposition A.3.3 (holomorphic extension of $\boldsymbol{y} \mapsto (\sigma, u)(\boldsymbol{y})$)**.** *Let $\beta > 0$ be a constant depending on the n-dimensional Friedrich-Poincaré inequality. Suppose that there exists functions $a_0 \in \mathrm{L}^\infty(\Omega)$ and $\{\psi_j\}_{j\in\mathbb{N}} \subset \mathrm{L}^\infty(\Omega)$ such that*

$$a(\boldsymbol{x}, \boldsymbol{y}) = a_0(\boldsymbol{x}) + \sum_{j\in\mathbb{N}} y_j \psi_j(\boldsymbol{x}), \quad \forall x \in \Omega, \forall \boldsymbol{y} \in \mathcal{U}.$$

Let $r, M > 0$ be such that

$$\|a_0\|_{\mathrm{L}^\infty(\Omega)} \leq M, \quad \text{and} \quad \sum_{j\in\mathbb{N}} |\psi_j(\boldsymbol{x})| \leq a_0(\boldsymbol{x}) - r, \tag{A.3.3}$$

holds for all $\boldsymbol{x} \in \Omega$. Moreover, let $\boldsymbol{\rho} \in \mathbb{R}^{\mathbb{N}}$ with $\boldsymbol{\rho} \geq \mathbf{1}$ and $0 < \varepsilon < r$ be such that

$$\sum_{j\in\mathbb{N}} \left(\frac{\rho_j + \rho_j^{-1}}{2} - 1 \right) \|\psi_j\|_{\mathrm{L}^\infty(\Omega)} \leq \varepsilon. \tag{A.3.4}$$

Then, the solution map

$$(\boldsymbol{\sigma}, u) : \mathcal{U} \to \mathrm{H}(\mathrm{div}; \Omega; \mathbb{R}) \times \mathrm{L}^2(\Omega; \mathbb{R})$$

associated with prolem (A.2.5) admits a well-defined and holomorphic extension

$$(\bar{\boldsymbol{\sigma}}, \bar{u}) : \mathcal{O} \to \mathbf{H}(\mathrm{div}; \Omega; \mathbb{C}) \times \mathrm{L}^2(\Omega; \mathbb{C})$$

to an open set $\mathcal{O} \subseteq \mathbb{C}^{\mathbb{N}}$ such that $\mathcal{E}_{\boldsymbol{\rho}} \subset \mathcal{O}$, where $\mathcal{E}_{\boldsymbol{\rho}}$ is the filled-in Bernstein polyellipse. In particular, $(\bar{\boldsymbol{\sigma}}, \bar{u}) : \mathcal{O} \to \mathbf{H}(\mathrm{div}; \Omega; \mathbb{C}) \times \mathrm{L}^2(\Omega; \mathbb{C})$ is such that $(\bar{\boldsymbol{\sigma}}, \bar{u})|_{\mathcal{U}} = (\boldsymbol{\sigma}, u)$. Moreover, the following upper bound holds:

$$\|(\bar{\boldsymbol{\sigma}}, \bar{u})\|_{L^\infty(\mathcal{E}_{\boldsymbol{\rho}}; \mathbf{H}(\mathrm{div};\Omega;\mathbb{C}) \times \mathrm{L}^2(\Omega;\mathbb{C}))} \leq C_1(M, r, \beta, \varepsilon) \|f\|_{\mathrm{L}^2(\Omega)} + C_2(M, r, \beta, \varepsilon) \|g\|_{\mathrm{H}^{1/2}(\partial\Omega)}$$

where the constants C_1 and C_2 are given by

$$C_1(M, r, \beta, \varepsilon) = (1 + 2\beta M) \left(\frac{1 + ((r-\varepsilon)/M)^2}{4(r-\varepsilon)\beta^2} \right), C_2(M, r, \beta, \varepsilon) = \frac{1 + ((r-\varepsilon)/M)^2}{4\beta} + \frac{4M^2}{(r-\varepsilon)}.$$

Proof. As in [12, Proposition 4.9], we define a complex extension of the parametrization $\boldsymbol{y} \mapsto a(\boldsymbol{y})$ as

$$\boldsymbol{z} \in \mathbb{C}^{\mathbb{N}} \mapsto a(\boldsymbol{x}, \boldsymbol{z}) = a_0(\boldsymbol{x}) + \sum_{j\in\mathbb{N}} z_j \psi_j(\boldsymbol{x}) \in \mathbb{C}, \quad \forall \boldsymbol{x} \in \Omega. \tag{A.3.5}$$

This is a well defined mapping for every $\boldsymbol{z} \in \mathcal{E}_{\boldsymbol{\rho}}$. In fact, using (A.3.4) and the uniform ellipticity condition (A.3.3) we get

$$\begin{aligned}\sum_{j\in\mathbb{N}} |z_j||\psi_j(\boldsymbol{x})| \leq \sum_{j\in\mathbb{N}} \frac{\rho_j+\rho_j^{-1}}{2}|\psi_j(\boldsymbol{x})| &\leq \sum_{j\in\mathbb{N}} |\psi_j(\boldsymbol{x})| + \sum_{j\in\mathbb{N}} \left(\frac{\rho_j+\rho_j^{-1}}{2}-1\right)\|\psi_j\|_{\mathrm{L}^\infty(\Omega)} \\ &\leq a_0(\boldsymbol{x}) - (r-\epsilon).\end{aligned}$$

Since $|z_j| \leq (\rho_j+\rho_j^{-1})/2$ for every $j \in \mathbb{N}$ by definition of the Bernstein polyellipse $\mathcal{E}_{\boldsymbol{\rho}}$, we deduce that the infinite sum in (A.3.5) converges absolutely and uniformly for $\boldsymbol{z} \in \mathcal{E}_{\boldsymbol{\rho}}$. Furthermore, for every $\boldsymbol{z} \in \mathcal{E}(\boldsymbol{\rho})$ we get

$$|a(\boldsymbol{x},\boldsymbol{z})| \leq |a_0(\boldsymbol{x})| + \sum_{j\in\mathbb{N}} |z_j||\psi_j(\boldsymbol{x})| \leq 2|a_0(\boldsymbol{x})| - (r-\varepsilon) \leq 2M, \quad \forall \boldsymbol{x} \in \Omega,$$

and

$$\begin{aligned}\operatorname*{ess\,inf}_{\boldsymbol{x}\in\Omega} \operatorname{Re}(a(\boldsymbol{x},\boldsymbol{z})) &= \operatorname*{ess\,inf}_{\boldsymbol{x}\in\Omega} \left(a_0(\boldsymbol{x}) + \sum_{j\in\mathbb{N}} \operatorname{Re}(z_j)\psi_j(\boldsymbol{x})\right) \\ &= \operatorname*{ess\,inf}_{\boldsymbol{x}\in\Omega} \left(a_0(\boldsymbol{x}) - \sum_{j\in\mathbb{N}} \frac{\rho_j+\rho_j^{-1}}{2}|\psi_j(\boldsymbol{x})|\right) \\ &\geq r-\varepsilon.\end{aligned}$$

Hence, $a(\mathcal{E}_{\boldsymbol{\rho}}) \subseteq \mathcal{R}^{\mathrm{UE}}_{r-\varepsilon,2M}$ is the region in which the complex uniform ellipticity condition holds with parameter $r-\varepsilon$ and bound $2M$.

Note that the mapping $\boldsymbol{z} \mapsto a(\boldsymbol{z})$ is affine and therefore holomorphic. Recall from Proposition A.3.2 that this parametrization maps $\mathcal{E}(\boldsymbol{\rho})$ into the open region $\mathcal{O}^{\mathrm{UE}} \subset \mathrm{L}^\infty(\Omega;\mathbb{C})$, where the mapping $a \mapsto (\sigma,u)(a)$ admits a holomorphic extension $a \mapsto (\bar{\boldsymbol{\sigma}},\bar{u})(a)$. Combining these two results we deduce that we can extend $\boldsymbol{y} \mapsto (\boldsymbol{\sigma},u)(\boldsymbol{y}) = (\boldsymbol{\sigma}(a(\boldsymbol{y})),u(a(\boldsymbol{y})))$ in a holomorphic way to $\mathcal{E}_{\boldsymbol{\rho}}$ as

$$\boldsymbol{z} \mapsto (\bar{\boldsymbol{\sigma}},\bar{u})(\boldsymbol{z}) = (\bar{\boldsymbol{\sigma}}(a(\boldsymbol{z})),\bar{u}(a(\boldsymbol{z}))),$$

since the composition of holomorphic maps is holomorphic. Now, from the upper bound $2M$ and lower bound $r-\varepsilon$ above, and using (A.3.2) we obtain the uniform upper bound in the result.

Finally, we provide an open set $\mathcal{O}$ containing $\mathcal{E}_{\boldsymbol{\rho}}$ in which $\boldsymbol{z} \mapsto (\bar{\sigma},\bar{u})(\boldsymbol{z})$ is holomorphic. The argument follows the same arguments as in [12, Prop. 4.9]. Let $0 < \delta < r-\varepsilon$, and let $\mathcal{O}$ be the open interior of the polyellipse $\mathcal{E}_{\widetilde{\boldsymbol{\rho}}}$ with parameter $\widetilde{\boldsymbol{\rho}} \geq \mathbf{1}$ defined implicitly by

$$\frac{\widetilde{\rho}_j+\widetilde{\rho}_j^{-1}}{2} = \frac{\rho_j+\rho_j^{-1}}{2} + \frac{\delta}{\sum_{j\in\mathbb{N}}\|\psi_j\|_{\mathrm{L}^\infty(\Omega)}}, \forall j \in \mathbb{N}.$$

Then for every $\boldsymbol{x} \in \Omega$, we have

$$\sum_{j\in\mathbb{N}} \frac{\widetilde{\rho}_j + \widetilde{\rho}_j^{-1}}{2}|\psi_j(\boldsymbol{x})| \leq \sum_{j\in\mathbb{N}} \frac{\rho_j + \rho_j^{-1}}{2}|\psi_j(\boldsymbol{x})| + \delta \leq a_0(\boldsymbol{x}) - (r - \varepsilon - \delta).$$

Hence, arguing as in the first part of the proof, we deduce that $a(\mathcal{O}) \subseteq \mathcal{R}^{\mathrm{UE}}_{r-\varepsilon-\delta,2M}$. Combining the previous proposition and that $\boldsymbol{z} \mapsto a(\boldsymbol{z})$ is holomorphic, we get that $\boldsymbol{z} \mapsto (\bar{\boldsymbol{\sigma}}, \bar{u})(\boldsymbol{z})$ is holomorphic in $\mathcal{O}$. □

Appendix B

Legendre coefficients summability and best s-term polynomial approximation rates

This appendix presents four lemmas on the summability of the coefficients of the Legendre polynomial expansion for the class of $(\boldsymbol{b}, 1)$-holomorphic functions introduced in §A.3. These will allow us to obtain upper bounds on the m-widths (6.1.3)–(6.1.5).

B.1 Setup

Consider the setup from Chapter 2 where $d = \infty$. Recall that $\mathcal{F}$ is the set of multi-indices with at most finitely-many nonzero entries. Let ϱ be the uniform probability measure on $\mathcal{U} = [-1, 1]^{\mathbb{N}}$ and $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\mathcal{F}}$ be the orthonormal Legendre basis of $L^2_\varrho(\mathcal{U})$ constructed via the tensorization

$$\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) = \prod_{k\in\mathbb{N}} \Psi_{\nu_k}(y_k), \quad \boldsymbol{y} \in \mathcal{U}, \boldsymbol{\nu} \in \mathcal{F},$$

where Ψ_ν is the univariate, orthonormal Legendre polynomial of degree ν. Then any $f \in L^2_\varrho(\mathcal{U}; \mathcal{V})$ has the convergent expansion (2.4.2). Now let $S \subset \mathcal{F}$ be a finite index set of size N. Then the truncated series of f is given by

$$f_S = \sum_{\boldsymbol{\nu}\in S} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}. \tag{B.1.1}$$

B.2 ℓ^p-summability and best s-term rates

Given a (multi-)index set $\Lambda \subseteq \mathcal{F}$. Recall the definition of the $\ell^p(\Lambda; \mathcal{V})$-norm from §2.2

$$\|\boldsymbol{v}\|_{p;\mathcal{V}} = \begin{cases} \left(\sum_{\boldsymbol{\nu}\in\Lambda} \|v_{\boldsymbol{\nu}}\|_{\mathcal{V}}^p\right)^{1/p}, & 0 < p < \infty, \\ \sup_{\boldsymbol{\nu}\in\Lambda} \|v_{\boldsymbol{\nu}}\|_{\mathcal{V}}, & p = \infty, \end{cases} \qquad \boldsymbol{v} = (v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}.$$

We shall typically use this in the case $\Lambda = \mathcal{F}$ or $\Lambda = [N]$.

The proof of the following lemma can be found in [12, Thm. 3.28] and provides a key summability estimate for the Legendre coefficients of a $(\boldsymbol{b}, 1)$-holomorphic function.

Lemma B.2.1. *Let $0 < p < 1$ and $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with $\boldsymbol{b} \geq \boldsymbol{0}$. Then the Legendre coefficients $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ in (2.4.2) satisfy*

$$\|\boldsymbol{c}\|_{p;\mathcal{V}} \leq C(\boldsymbol{b}, p), \quad \forall f \in \mathcal{H}(\boldsymbol{b}), \tag{B.2.1}$$

where $C(\boldsymbol{b}, p)$ depends on $\boldsymbol{b}$ and p only.

We are particularly interested in bounding the supremum of $C = C(\boldsymbol{b}, q)$ over $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ for $0 < p < q < 1$. This bound is used in the proof of part (b) of Theorems 6.3.3 and 6.3.4. The following result is obtained by modifying the proof of [12, Thm 3.28].

Lemma B.2.2. *Let $0 < p < q < 1$. Then*

$$\sup_{\|\boldsymbol{b}\|_{p,\mathsf{M}}\leq 1} C(\boldsymbol{b}, q) \leq c_{p,q},$$

where $C = C(\boldsymbol{b}, q)$ is the constant in (B.2.1) and $c_{p,q}$ is a positive constant depending on p and q only.

Proof. Consider $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1$. Notice from [12, Eq. (3.48)] that the constant C in (B.2.1) can be taken to be

$$C(\boldsymbol{b}, q) = \xi(\tilde{\kappa})^d \left(\sum_{n=0}^{\infty} \frac{(2n+1)^{q/2}}{\tilde{\kappa}^{qn}} \right)^{d/q} \|\boldsymbol{g}(\boldsymbol{b})\|_q, \tag{B.2.2}$$

where $\xi(t) = \min\{2t, \frac{\pi}{2}(t + t^{-1})\}/(t-1)$ for every $t > 1$, the term $\tilde{\kappa} = \tilde{\kappa}(\boldsymbol{b}) > 1$ is defined as the unique solution to

$$\frac{\tilde{\kappa} + \tilde{\kappa}^{-1}}{2} = 1 + \frac{1}{2\|\boldsymbol{b}\|_1},$$

and

$$\begin{aligned} g(\boldsymbol{b})_{\boldsymbol{\nu}} &= \frac{\|\boldsymbol{\nu}\|_1!}{\boldsymbol{\nu}!} \boldsymbol{h}(\boldsymbol{b})^{\boldsymbol{\nu}} \prod_{j\in\mathbb{N}} (\xi(\tilde{\kappa})\sqrt{3\nu_j} + 1), \quad \forall \boldsymbol{\nu} \in \mathcal{F}, \\ h(\boldsymbol{b})_j &= 2e b_{j+d}, \end{aligned} \tag{B.2.3}$$

where $d \in \mathbb{N}$ is a truncation parameter.

We aim to bound (B.2.2) by a constant depending on p and q. First, we show that there is a convergent sequence $\tilde{\boldsymbol{h}}$ independent of $\boldsymbol{b}$ that can replace $\boldsymbol{h}(\boldsymbol{b})$ in (B.2.3). Then, we proceed using similar arguments to those in the proof of [12, Thm. 3.28] to get the result.

Let $\tilde{\boldsymbol{b}}$ be the minimal monotone majorant (2.4.15) of $\boldsymbol{b}$. Using Stechkin's inequality we get

$$\sum_{j=n+1}^{\infty} \tilde{b}_j = \sigma_n(\tilde{\boldsymbol{b}})_1 \leq n^{1-1/p} \|\tilde{\boldsymbol{b}}\|_p = n^{1-1/p} \|\boldsymbol{b}\|_{p,\mathsf{M}} \leq c_p n^{1-1/p},$$

for some constant c_p depending only on p. Also by monotonicity,

$$n\tilde{b}_{2n} \leq \tilde{b}_{n+1} + \cdots + \tilde{b}_{2n} \leq \sum_{j=n+1}^{\infty} \tilde{b}_j \leq c_p n^{1-1/p}.$$

Hence $b_n \leq \tilde{b}_n \leq \tilde{c}_p n^{-1/p}$ for a possible different constant depending on p. Note that the inequality for odd values of n can be established using a similar argument. Keeping this in mind, we define the sequence $\tilde{\boldsymbol{h}}(p) = (\tilde{h}(p)_j)_{j\in\mathbb{N}}$ by

$$\tilde{h}(p)_j = 2\tilde{c}_p \mathrm{e}(j+d)^{-1/p},$$

where $d \in \mathbb{N}$ is a parameter that will be chosen in the next step. Thus, we get the bound

$$h(\boldsymbol{b})_j = 2\mathrm{e}b_{j+d} \leq 2\tilde{c}_p \mathrm{e}(j+d)^{-1/p} = \tilde{h}(p)_j, \quad \forall j \in \mathbb{N}.$$

Observe that $\tilde{\boldsymbol{h}}(p) \in \ell^1(\mathbb{N})$. Moreover, using the fact that $q/p > 1$ and a simple convergence argument, we obtain that

$$\|\tilde{\boldsymbol{h}}(p)\|_q = 2\tilde{c}_p \mathrm{e} \left(\sum_{j\in\mathbb{N}} (j+d)^{-q/p} \right)^{1/q} < \infty, \tag{B.2.4}$$

which implies that $\tilde{\boldsymbol{h}}(p) \in \ell^q(\mathbb{N})$. We now choose $d = d(p)$ as the minimum $d \in \mathbb{N}$ such that

$$\|\tilde{\boldsymbol{h}}(p)\|_1 = \sum_{j\in\mathbb{N}} 2\tilde{c}_p \mathrm{e}(j+d)^{-1/p} < 1. \tag{B.2.5}$$

On the other hand, since $\|\boldsymbol{b}\|_1 \leq \|\boldsymbol{b}\|_{1,\mathsf{M}} \leq \|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1$ and examining the solution of equation

$$\frac{\tilde{\kappa} + \tilde{\kappa}^{-1}}{2} = 1 + \frac{1}{2\|\boldsymbol{b}\|_1},$$

we deduce that $\tilde{\kappa} > 2.6$ through a straightforward inspection. Hence, from the definition of ξ and the lower bound on $\tilde{\kappa}$ we get that $\xi(\tilde{\kappa}) \leq 2\tilde{\kappa}/(\tilde{\kappa}-1) \leq 4$. Note that this upper bound is independent of $\tilde{\kappa}$, and therefore independent of $\boldsymbol{b}$. Keeping this in mind, we get

$$g(\boldsymbol{b})_{\boldsymbol{\nu}} = \frac{\|\boldsymbol{\nu}\|_1!}{\boldsymbol{\nu}!} \boldsymbol{h}(\boldsymbol{b})^{\boldsymbol{\nu}} \prod_{j\in\mathbb{N}} (\xi(\tilde{\kappa})\sqrt{3\nu_j} + 1) \leq \frac{\|\boldsymbol{\nu}\|_1!}{\boldsymbol{\nu}!} \tilde{\boldsymbol{h}}(p)^{\boldsymbol{\nu}} \prod_{j\in\mathbb{N}} (4\sqrt{3\nu_j} + 1) =: \tilde{g}(p)_{\boldsymbol{\nu}}, \quad \forall \boldsymbol{\nu} \in \mathcal{F}, \tag{B.2.6}$$

which implies that $\|\boldsymbol{g}(\boldsymbol{b})\|_q \leq \|\tilde{\boldsymbol{g}}(p)\|_q$. Therefore, we can bound (B.2.2) by

$$C(\boldsymbol{b}, q) = \xi(\tilde{\kappa})^d \left(\sum_{n=0}^{\infty} \frac{(2n+1)^{q/2}}{\tilde{\kappa}^{qn}} \right)^{d/q} \|\boldsymbol{g}(\boldsymbol{b})\|_q \leq \xi(\tilde{\kappa})^d \left(\sum_{n=0}^{\infty} \frac{(2n+1)^{q/2}}{\tilde{\kappa}^{qn}} \right)^{d/q} \|\tilde{\boldsymbol{g}}(p)\|_q. \tag{B.2.7}$$

To show that $\|\tilde{\boldsymbol{g}}(p)\|_q < \infty$ we combine (B.2.4) with (B.2.5) and apply [12, Lem. 3.29]. It remains to bound the other term in the previous inequality. Recall that $\xi(\tilde{\kappa}) \leq 4$ and

$\tilde{\kappa} > 2.6$. Then,

$$\xi(\tilde{\kappa})^d \left(\sum_{n=0}^{\infty} \frac{(2n+1)^{q/2}}{\tilde{\kappa}^{qn}} \right)^{d/q} \leq 4^d \left(\sum_{n=0}^{\infty} \frac{(2n+1)^{q/2}}{(2.6)^{qn}} \right)^{d/q} \leq \bar{c}_{p,q} < \infty,$$

where $\bar{c}_{p,q}$ is a positive constant depending on p and q only. Note that $\bar{c}_{p,q}$ depends on p due to the dependence of d on p. In this way, by taking supremum over $\|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1$ in (B.2.7) we get the result. □

We now present a best s-term approximation rate for the Legendre polynomials. For non-sparse vectors in $\ell^p(\mathcal{F};\mathcal{V})$, we recall from (2.4.8) the definition of the ℓ^p-norm best s-term approximation error as

$$\sigma_s(\boldsymbol{x})_{p;\mathcal{V}} = \inf_{\boldsymbol{z} \in \ell^p(\mathcal{F};\mathcal{V})} \{\|\boldsymbol{x} - \boldsymbol{z}\|_{p;\mathcal{V}} : |\mathrm{supp}(\boldsymbol{z})| \leq s\}, \quad \boldsymbol{x} \in \ell^p(\mathcal{F};\mathcal{V}), \tag{B.2.8}$$

where $\mathrm{supp}(\boldsymbol{z})$ is the support of the vector $\boldsymbol{z}$ as in (2.4.7). The following result can be deduced from Stechkin's inequality and Lemma B.2.1. Note that the proof of Lemma B.2.1 (see [12, Thm. 3.28]) involves establishing the summability of a bounding sequence for the $\mathcal{V}$-norms of the polynomial coefficients in (2.4.2). This bound is equal to $\|f\|_{L^\infty(\mathcal{R}(\boldsymbol{b});\mathcal{V})}$ multiplied by a factor that is independent of f and depending on $\boldsymbol{b}$ only. Consequently, the index set in the following result is independent of f.

Corollary B.2.3. *Let $0 < p < 1$, $q \geq p$, $\boldsymbol{b} \in \ell^p(\mathbb{N})$ with $\boldsymbol{b} \geq \boldsymbol{0}$ and $s \in \mathbb{N}$. Then, there exists a set $S \subset \mathcal{F}$ of size $|S| \leq s$ depending on $\boldsymbol{b}$ and p only such that*

$$\sigma_s(\boldsymbol{c})_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q;\mathcal{V}} \leq C(\boldsymbol{b},p) \cdot s^{1/q-1/p}, \quad \forall f \in \mathcal{H}(\boldsymbol{b}), \tag{B.2.9}$$

where $C(\boldsymbol{b},p) > 0$ is the constant in (B.2.1) and $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ are the Legendre coefficients in (2.4.2).

B.3 ℓ^p_{A}-summability and best s-term rates in anchored sets

In the last part of this appendix we require the notion of lower and anchored sets and the *minimal anchored majorant* of a sequence and the ℓ^p_{A} space. A sequence $\boldsymbol{c} \in \ell^\infty(\mathcal{F};\mathcal{V})$ belongs to $\ell^p_{\mathsf{A}}(\mathcal{F};\mathcal{V})$ if its minimal anchored majorant $\tilde{\boldsymbol{c}} = (\tilde{c}_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$, defined by

$$\tilde{c}_{\boldsymbol{\nu}} = \begin{cases} \sup\{\|c_{\boldsymbol{\mu}}\|_{\mathcal{V}} : \boldsymbol{\mu} \geq \boldsymbol{\nu}\} & \text{if } \boldsymbol{\nu} \neq \boldsymbol{e}_j \text{ for any } j \in \mathbb{N}, \\ \sup\{\|c_{\boldsymbol{\mu}}\|_{\mathcal{V}} : \boldsymbol{\mu} \geq \boldsymbol{e}_i \text{ for some } i \geq j\} & \text{if } \boldsymbol{\nu} = \boldsymbol{e}_j \text{ for some } j \in \mathbb{N}, \end{cases} \tag{B.3.1}$$

belongs to $\ell^p(\mathcal{F})$. In particular, we define its $\ell^p_{\mathsf{A}}(\mathcal{F};\mathcal{V})$-norm by

$$\|\boldsymbol{c}\|_{p,\mathsf{A};\mathcal{V}} = \|\tilde{\boldsymbol{c}}\|_{p;\mathcal{V}}. \tag{B.3.2}$$

For further details we refer to [12, Def. 3.31].

Similar to Lemma B.2.1, the following result shows summability of the Legendre coefficients of a $(\boldsymbol{b},1)$-holomorphic function in the ℓ^p_{A}-norm.

Lemma B.3.1. *Let $0 < p < 1$ and $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\boldsymbol{b} \geq \boldsymbol{0}$. Then the Legendre coefficients $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ in (2.4.2) satisfy*

$$\|\boldsymbol{c}\|_{p,\mathsf{A};\mathcal{V}} \leq C_{\mathsf{A}}(\boldsymbol{b}, p), \quad \forall f \in \mathcal{H}(\boldsymbol{b}), \tag{B.3.3}$$

where $C_{\mathsf{A}}(\boldsymbol{b}, p)$ depends on $\boldsymbol{b}$ and p only.

Proof. Let $\tilde{\boldsymbol{b}}$ be the minimal monotone majorant of $\boldsymbol{b}$, defined in (2.4.15). Following the same argument as in Corollary 2.4.15 we get that $\mathcal{R}(\tilde{\boldsymbol{b}}) \subseteq \mathcal{R}(\boldsymbol{b})$ where $\mathcal{R}(\boldsymbol{b})$ is as in (2.3.2). Therefore, $\mathcal{H}(\boldsymbol{b}) \subseteq \mathcal{H}(\tilde{\boldsymbol{b}})$. Since $\tilde{\boldsymbol{b}} \in \ell^p(\mathbb{N})$ is monotonically nonincreasing [12, Thm. 3.33] implies the result. □

As in Lemma B.2.2, we are interested in bounding the supremum of $C_{\mathsf{A}} = C_{\mathsf{A}}(\boldsymbol{b}, q)$ over $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$. This bound will be useful for the proof of part (b) of Theorem 6.3.4. The following result is obtained by modifying the proof of [12, Thm 3.33].

Lemma B.3.2. *Let $0 < p < q < 1$. Then*

$$\sup_{\|\boldsymbol{b}\|_{p,\mathsf{M}}\leq 1} C_{\mathsf{A}}(\boldsymbol{b}, q) \leq c_{p,q},$$

where $C_{\mathsf{A}}(\boldsymbol{b}, p)$ is the constant in (B.3.3) and $c_{p,q}$ is positive constant depending on p and q only.

Proof. Let $\tilde{\boldsymbol{b}}$ be the minimal monotone majorant (2.4.15) of $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ with $\|\tilde{\boldsymbol{b}}\|_p = \|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1$ and $\tilde{\kappa} = \tilde{\kappa}(\tilde{\boldsymbol{b}}) > 1$ be the unique solution to

$$\frac{\tilde{\kappa} + \tilde{\kappa}^{-1}}{2} = 1 + \frac{1}{4\|\tilde{\boldsymbol{b}}\|_1}.$$

Observe that $\tilde{\kappa} \geq 2$. Then, a simple inspection reveals that

$$\sqrt{2n+1} \leq \frac{6}{5}\left(\frac{3}{2}\right)^n \leq \frac{6}{5}\left(\frac{1+\tilde{\kappa}}{2}\right)^n, \forall n \in \mathbb{N}. \tag{B.3.4}$$

Also, define $\eta = \eta(\tilde{\kappa}) := (1+\tilde{\kappa})/(2\tilde{\kappa}) < 1$. Now, notice from [12, Eq. (3.62)] and the last paragraph in the proof of [12, Thm. 3.33] , that the constant C_{A} in (B.3.3) can be taken to be

$$C_{\mathsf{A}}(\boldsymbol{b}, q) = \widetilde{C}_{\mathsf{A}}(\tilde{\boldsymbol{b}}, q) = D_1^d \left(\sum_{n=0}^{\infty} \eta^{qn}\right)^{d/q} \|\boldsymbol{g}(\tilde{\boldsymbol{b}})\|_q, \tag{B.3.5}$$

where $D_1 = D_1(\tilde{\kappa}) := \max\{1, 6/5\xi(\tilde{\kappa})\}$ with $\xi(t) = \min\{2t, \frac{\pi}{2}(t + t^{-1})\}/(t-1)$ for every $t > 1$, and

$$\begin{aligned} g(\tilde{\boldsymbol{b}})_{\boldsymbol{\nu}} &= \frac{\|\boldsymbol{\nu}\|_1!}{\boldsymbol{\nu}!} \boldsymbol{h}(\tilde{\boldsymbol{b}})^{\boldsymbol{\nu}} \prod_{j\in\mathbb{N}} (\xi(\tilde{\kappa})\sqrt{3\nu_j} + 1), \quad \forall \boldsymbol{\nu} \in \mathcal{F}, \\ h(\tilde{\boldsymbol{b}})_j &= 4\mathrm{e}\tilde{b}_{j+d}, \end{aligned} \tag{B.3.6}$$

where $d \in \mathbb{N}$ is a truncation parameter.

Now, following the same arguments as in (B.2.3)–(B.2.7) we deduce that

$$\widetilde{C}_{\mathsf{A}}(\tilde{\boldsymbol{b}},q) = D_1^d \left(\sum_{n=0}^{\infty} \eta^{qn}\right)^{d/q} \|\boldsymbol{g}(\tilde{\boldsymbol{b}})\|_q \leq D_1^d \left(\sum_{n=0}^{\infty} \eta^{qn}\right)^{d/q} \|\tilde{\boldsymbol{g}}(p)\|_q, \tag{B.3.7}$$

where, using [12, Lem. 3.29] once more, we see that the sequence $\tilde{\boldsymbol{g}}(p)$, defined in (B.2.6) satisfies $\|\tilde{\boldsymbol{g}}(p)\|_q < \infty$. It remains to bound the other term in the previous inequality. As in the last steps in the proof of Lemma B.2.2, with $\xi(\tilde{\kappa}) \leq 4$ and $\tilde{\kappa} \geq 2$, we deduce that

$$D_1^d \left(\sum_{n=0}^{\infty} \eta^{qn}\right)^{d/q} \leq 4^d \left(\frac{6}{5}\right)^d \left(\sum_{n=0}^{\infty} \left(\frac{4}{5}\right)^{qn}\right)^{d/q} \leq c_{p,q} < \infty, \tag{B.3.8}$$

where $c_{p,q}$ is a positive constant depending on p and q only. Finally, by taking the supremum over $\|\boldsymbol{b}\|_{p,\mathsf{M}} \leq 1$ in (B.3.7) we get the result. □

Let $0 < p \leq \infty$. We now introduce the concept of the *ℓ^p-norm best s-term approximation error in anchored sets*. This is defined as

$$\sigma_{s,\mathsf{A}}(\boldsymbol{x})_{p;\mathcal{V}} = \inf_{\boldsymbol{z} \in \ell^p(\mathcal{F};\mathcal{V})} \{\|\boldsymbol{x} - \boldsymbol{z}\|_{p;\mathcal{V}} : |\mathrm{supp}(\boldsymbol{z})| \leq s, \mathrm{supp}(\boldsymbol{z}) \text{ anchored}\}, \quad \boldsymbol{x} \in \ell^p(\mathcal{F};\mathcal{V}). \tag{B.3.9}$$

Now we provide an algebraic s-term rate in anchored sets. The following result follows similar arguments to those used in Corollary B.2.3 and it is deduced by applying [12, Lemma 3.32] and Lemma B.3.1 to the Legendre coefficients $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ in (2.4.2). Observe that, to prove Lemma B.3.1 the $\mathcal{V}$-norm of the coefficients in (2.4.2) are bounded by a monotonically nonincreasing sequence (see [12, Eq. (3.55)]) that only depends on $\boldsymbol{b}$. Therefore, the anchored set in the following corollary is independent of f.

Corollary B.3.3. *Let $0 < p < 1$, $q \geq p$, $\boldsymbol{b} \in \ell^p_{\mathsf{M}}(\mathbb{N})$ and $s \in \mathbb{N}$. Then, there exists an anchored set $S \subset \mathcal{F}$ of size $|S| \leq s$ such that*

$$\sigma_{s,\mathsf{A}}(\boldsymbol{c})_{q;\mathcal{V}} \leq \|\boldsymbol{c} - \boldsymbol{c}_S\|_{q;\mathcal{V}} \leq C_{\mathsf{A}}(\boldsymbol{b},p) \cdot s^{1/q-1/p}, \quad \forall f \in \mathcal{H}(\boldsymbol{b}), \tag{B.3.10}$$

where $C_{\mathsf{A}}(\boldsymbol{b},p)$ is the constant in (B.3.3) and $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ are the Legendre coefficients in (2.4.2).

Appendix C

Widths of weighted ℓ^p-norm balls in $\mathbb{R}^N$

In this appendix, we present a proposition providing a lower bound to the Gelfand m-width in terms of m and $N \in \mathbb{N}$, an equality theorem proving the connection between Gelfand m-widths and Kolmogorov widths for a particular set of spaces, a duality result by Stesin [241] and a lemma that establishes the relationship between the unit balls in these spaces using the Kolmogorov m-widths.

First, recall the definition of the Gelfand m-width from §6.5.2. The following proposition can be obtained from [111, Prop. 2.1] by an inspection of the proof.

Proposition C.0.1 (Lower bound)**.** *Let $N \in \mathbb{N}$. For $0 < p \leq 1$, $m < N$ and $p < q \leq \infty$,*

$$d^m(B_N^p, \ell_N^q) \geq \left(\frac{1}{2}\right)^{\frac{2}{p}-\frac{1}{q}} \min\left\{1, \frac{\frac{2p}{\log(3^8 \mathrm{e})}\log(\mathrm{e}N/m)}{m}\right\}^{\frac{1}{p}-\frac{1}{q}}. \tag{C.0.1}$$

We now give the proof of an equality result based on the methodology described in [112, Lem. 10.15]. Specifically, we show that the Kolmogorov widths of ℓ_N^p-balls in the weighted ℓ_N^q space are equivalent to certain Gelfand widths.

First, for $1 \leq p, p^*, q, q^* \leq \infty$ we recall the definitions of the Gelfand and Kolmogorov widths for this particular case, see §6.5.2. The Gelfand m-width of the subset $B_N^{q^*}(1/\boldsymbol{w})$ of $\ell_N^{p^*}$ is

$$d^m(B_N^{q^*}(1/\boldsymbol{w}), \ell_N^{p^*}) = \inf\left\{\sup_{\boldsymbol{x} \in B_N^{q^*}(1/\boldsymbol{w}) \cap \mathcal{X}^m} \|\boldsymbol{x}\|_{p^*},\ \mathcal{X}^m \text{ a subspace of } \ell_N^{p^*} \text{ with } \operatorname{codim}(\mathcal{X}^m) \leq m\right\},$$

and the Kolmogorov m-width of a subset B_N^p of the space $\ell_N^q(1/\boldsymbol{w})$ is

$$d_m(B_N^p, \ell_N^q(\boldsymbol{w})) = \inf\left\{\sup_{\boldsymbol{x} \in B_N^p} \inf_{\boldsymbol{z} \in \mathcal{X}_m} \|\boldsymbol{x} - \boldsymbol{z}\|_{q,\boldsymbol{w}},\ \mathcal{X}_m \text{ a subspace of } \ell_N^q(\boldsymbol{w}) \text{ with } \dim(\mathcal{X}_m) \leq m\right\}.$$

Theorem C.0.2 (Stesin). *Let $N \in \mathbb{N}$ with $N > m$, $1 \leq q < p \leq \infty$, and $\boldsymbol{w} \in \mathbb{R}^N$ be a vector of positive weights. Then*

$$d_m(B_N^p(\boldsymbol{w}), \ell_N^q) = \left(\max_{\substack{i_1,\dots,i_{N-m} \in [N] \\ i_k \neq i_j}} \left(\sum_{j=1}^{N-m} w_{i_j}^{pq/(p-q)} \right)^{1/p-1/q} \right)^{-1}.$$

Theorem C.0.3 (Equality). *For $1 \leq p, q \leq \infty$, let $\boldsymbol{w} \in \mathbb{R}^N$ be a vector of positive weights and p^*, q^* be such that $1/p^* + 1/p = 1$ and $1/q^* + 1/q = 1$. Then*

$$d_m(B_N^p, \ell_N^q(\boldsymbol{w})) = d^m(B_N^{q^*}(1/\boldsymbol{w}), \ell_N^{p^*}). \tag{C.0.2}$$

Proof. First, given $\boldsymbol{x} \in B_N^p$ and a subspace X_m of $\ell_N^q(\boldsymbol{w})$, we follow the same arguments as those in [112, Lem. 10.15] to obtain

$$\inf_{\boldsymbol{z} \in X_m} \|\boldsymbol{x} - \boldsymbol{z}\|_{q,\boldsymbol{w}} = \langle \phi, \boldsymbol{x} \rangle,$$

for some linear bounded functional $\phi \in X_m^\circ$, with $\|\phi\|_{(\ell_N^q(\boldsymbol{w}))^*} \leq 1$, where

$$X_m^\circ := \{\phi \in (\ell_N^q(\boldsymbol{w}))^* : \phi(\boldsymbol{x}) = 0, \quad \forall \boldsymbol{x} \in X_m\}.$$

Now, by the definition in (6.5.1) we get

$$\inf_{\boldsymbol{z} \in X_m} \|\boldsymbol{x} - \boldsymbol{z}\|_{q,\boldsymbol{w}} \leq \sup_{\phi \in B_N^{q^*}(1/\boldsymbol{w}) \cap X_m^\circ} \langle \phi, \boldsymbol{x} \rangle.$$

On the other hand, for all $\phi \in B_N^{q^*}(1/\boldsymbol{w}) \cap X_m^\circ$, and $\boldsymbol{z} \in X_m$ we have

$$\langle \phi, \boldsymbol{x} \rangle = \langle \phi, \boldsymbol{x} - \boldsymbol{z} \rangle \leq \|\phi\|_{(\ell_N^q(\boldsymbol{w}))^*} \|\boldsymbol{x} - \boldsymbol{z}\|_{q,\boldsymbol{w}}.$$

Then we deduce the following equality

$$\inf_{\boldsymbol{z} \in X_m} \|\boldsymbol{x} - \boldsymbol{z}\|_{q,\boldsymbol{w}} = \sup_{\phi \in B_N^{q^*}(1/\boldsymbol{w}) \cap X_m^\circ} \langle \phi, \boldsymbol{x} \rangle. \tag{C.0.3}$$

Taking supremum on both sides over $\boldsymbol{x} \in B_N^p$, we have

$$\begin{aligned}
\sup_{\boldsymbol{x} \in B_N^p} \inf_{\boldsymbol{z} \in X_m} \|\boldsymbol{x} - \boldsymbol{z}\|_q &= \sup_{\boldsymbol{x} \in B_N^p} \sup_{\phi \in B_N^{q^*}(1/\boldsymbol{w}) \cap X_m^\circ} \langle \phi, \boldsymbol{x} \rangle \\
&= \sup_{\phi \in B_N^{q^*}(1/\boldsymbol{w}) \cap X_m^\circ} \sup_{\boldsymbol{x} \in B_N^p} \langle \phi, \boldsymbol{x} \rangle \\
&= \sup_{\phi \in B_N^{q^*}(1/\boldsymbol{w}) \cap X_m^\circ} \|\phi\|_{p^*} \sup_{\boldsymbol{x} \in B_N^p} \|\boldsymbol{x}\|_p \\
&= \sup_{\phi \in B_N^{q^*}(1/\boldsymbol{w}) \cap X_m^\circ} \|\phi\|_{p^*}.
\end{aligned}$$

Taking the infimum over all subspaces X_m with $\dim(X_m) \leq m$ and noticing the one-to-one correspondence between the subpaces X_m° and the subspaces $\mathcal{X}^m$ with $\operatorname{codim}(\mathcal{X}^m) \leq m$, we

obtain

$$d_m(B_N^p, \ell_N^q(\boldsymbol{w})) = d^m(B_N^{q^*}(1/\boldsymbol{w}), \ell_N^{p^*}),$$

as required. □

The following result establishes a connection between the unit ball in the weighted space $\ell_N^q(\boldsymbol{w})$ and the unit weighted ball in ℓ_N^q, using the Kolmogorov m-width

$$d_m(B_N^p(1/\boldsymbol{w}), \ell_N^q) = \inf \left\{ \sup_{\boldsymbol{x} \in B_N^p(1/\boldsymbol{w})} \inf_{\boldsymbol{z} \in \mathcal{X}_m} \|\boldsymbol{x} - \boldsymbol{z}\|_q,\ \mathcal{X}_m \text{ a subspace of } \ell_N^q \text{ with } \dim(\mathcal{X}_m) \leq m \right\}.$$

Lemma C.0.4. *Let $\boldsymbol{w} \in \mathbb{R}^N$ be a vector of positive weights and $1 \leq p, q \leq \infty$. Then*

$$d_m(B_N^p, \ell_N^q(\boldsymbol{w})) = d_m(B_N^p(1/\boldsymbol{w}), \ell_N^q). \tag{C.0.4}$$

Proof. Let $\boldsymbol{x} \in B_N^p$ and $\boldsymbol{z} \in X_m$, where X_m is a m-dimensional subspace of $X = \ell_N^q(\boldsymbol{w})$. Then

$$\inf_{\boldsymbol{z} \in X_m} \|\boldsymbol{x} - \boldsymbol{z}\|_{q,\boldsymbol{w}} = \inf_{\boldsymbol{z} \in X_m} \left(\sum_{i \in [N]} \left(\frac{|x_i - z_i|}{w_i} \right)^q \right)^{1/q} = \inf_{\boldsymbol{z}' \in X_m'} \|(1/\boldsymbol{w}) \odot \boldsymbol{x} - \boldsymbol{z}'\|_q.$$

Notice there is a one-to-one correspondence between subspaces X_m and subspaces $X_m' = \{(1/\boldsymbol{w}) \odot \boldsymbol{z} : \boldsymbol{z} \in X_m\}$. Also, there is a one-to-one correspondence between $\boldsymbol{x} \in B_N^p$ and $(1/\boldsymbol{w}) \odot \boldsymbol{x} \in B_N^p(1/\boldsymbol{w})$. Thus,

$$d_m(B_N^p, \ell_N^q(\boldsymbol{w})) = d_m(B_N^p(1/\boldsymbol{w}), \ell_N^q).$$

as required. □