

# **An Application of Kolmogorov's Superposition Theorem to Function Reconstruction in Higher Dimensions**

Dissertation  
zur  
Erlangung des Doktorgrades (Dr. rer. nat.)  
der  
Mathematisch–Naturwissenschaftlichen Fakultät  
der  
Rheinischen Friedrich–Wilhelms–Universität Bonn  
vorgelegt von  
Jürgen Braun  
aus  
Bornheim  
Bonn 2009

Angefertigt mit Genehmigung der Mathematisch–Naturwissenschaftlichen Fakultät  
der Rheinischen Friedrich–Wilhelms–Universität Bonn

1. Gutachter: Prof. Dr. Michael Griebel
  2. Gutachter: Prof. Dr. Mario Bebendorf
- Tag der Promotion: 20. November 2009  
Erscheinungsjahr: 2009

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn  
[http://hss.ulb.uni-bonn.de/diss\\_online](http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

## Zusammenfassung

In der vorliegenden Arbeit wird ein Regularisierungsnetzwerk zur Rekonstruktion von stetigen Funktionen  $f : [0, 1]^n \rightarrow \mathbb{R}$  vorgestellt, welches direkt auf einer neuen konstruktiven Version von Kolmogorovs Superpositionen Theorem basiert. Dabei sind lediglich die Funktionswerte  $f(\mathbf{x}_j)$  an diskreten Datenpunkten  $\mathbf{x}_j, j = 1, \dots, P$  bekannt.

Typischerweise leidet die numerische Lösung mathematischer Probleme unter dem sogenannten *Fluch der Dimension*. Dieser Begriff beschreibt das exponentielle Wachstum der Komplexität des verwendeten Verfahrens mit der Dimension  $n$ . Um dies zumindest teilweise zu vermeiden, werden üblicherweise höhere Regularitätsannahmen an die Lösung des Problems gemacht, was allerdings häufig unrealistisch ist. Daher wird in dieser Arbeit eine Darstellung der Funktion  $f$  als Superposition eindimensionaler Funktionen verwendet, welche keiner höheren Regularitätsannahmen als Stetigkeit bedarf. Zu diesem Zweck wird eine konstruktive Variante des Kolmogorov Superpositionen Theorems, welche auf D. Sprecher zurückgeht, so angepasst, dass nur eine äußere Funktion  $\Phi$  sowie eine universelle innere Funktion  $\psi$  zur Darstellung von  $f$  notwendig ist. Die Funktion  $\psi$  ist nach einer Definition von M. Köppen explizit und unabhängig von  $f$  als Fortsetzung einer Funktion, welche auf einer Dichten Teilmenge der reellen Achse definiert ist, gegeben. Der fehlende Beweis von Existenz, Stetigkeit und Monotonie von  $\psi$  wird in dieser Arbeit geführt. Zur Berechnung der äußeren Funktion  $\Phi$  wird ein iterativer Algorithmus von Sprecher so modifiziert, dass jeder Iterationsschritt, abhängig von  $f$ , ein Element einer Folge univariater Funktionen  $\{\Phi^r\}_r$  liefert. Es wird gezeigt werden, dass die Folge gegen einen stetigen Grenzwert  $\Phi : \mathbb{R} \rightarrow \mathbb{R}$  konvergiert. Dies liefert einen konstruktiven Beweis einer neuen Version des Kolmogorov Superpositionen Theorems mit einer äußeren und einer inneren Funktion.

Da die numerische Komplexität des Algorithmus zur Berechnung von  $\Phi$  exponentiell mit der Dimension wächst, stellen wir alternativ ein Regularisierungsnetzwerk, basierend auf dieser Darstellung, vor. Dabei wird die äußere Funktion aus gegebenen Daten  $(\mathbf{x}_j, f(\mathbf{x}_j)), j = 1, \dots, P$  berechnet. Das Modell zur Rekonstruktion von  $f$  wird in zwei Schritten eingeführt. Zunächst wird zur Definition eines vorläufigen Modells die äußere Funktion, bzw. eine Approximation an  $\Phi$ , in einer endlichen Basis mit unbekannten Koeffizienten dargestellt. Diese werden dann durch eine Variationsformulierung bestimmt, d.h. durch die Minimierung eines regularisierten empirischen Fehlerfunktionalen. Eine detaillierte numerische Analyse zeigt dann, dass Kolmogorovs Darstellung die Dimensionalität von  $f$  in Oszillationen von  $\Phi$  transformiert. Somit ist die Verwendung von Basisfunktionen mit lokalem Träger nicht geeignet, da die räumliche Auflösung der Approximation die starken Oszillationen erfassen muss. Des Weiteren zeigt eine Analyse der Fouriertransformation von  $\Phi$ , dass die relevanten Frequenzen, unabhängig von  $f$ , a priori bestimmbar sind, und

dass die äußere Funktion Produktstruktur aufweist. Dies motiviert die Definition des endgültigen Modells. Dazu wird  $\Phi$  nun durch ein Produkt von Funktionen ersetzt und jeder Faktor in einer Fourierbasis entwickelt. Die Koeffizienten werden ebenfalls durch Minimierung eines regularisierten empirischen Fehlerfunktional bestimmt.

Für beide Modelle wird ein theoretischer Rahmen in Form von Hilberträumen mit reproduzierendem Kern entwickelt. Die zugehörigen Normen bilden dabei jeweils den Regularisierungsterm der entsprechenden Fehlerfunktionale. Somit können beide Ansätze als Regularisierungsnetzwerke interpretiert werden. Allerdings ist zu beachten, dass das Fehlerfunktional für den Produktansatz nicht konvex ist und nichtlineare Minimierungsverfahren zur Berechnung der Koeffizienten notwendig sind. Weitere ausführliche numerische Tests zeigen, dass dieses Modell in der Lage ist Funktionen zu rekonstruieren welche von bis zu zehn Variablen abhängen.

## Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich bei der Fertigstellung dieser Arbeit unterstützt haben. Dieser Dank gebührt an erster Stelle meinem Betreuer Prof. Dr. Michael Griebel für die Möglichkeit zur Erstellung dieser Arbeit und die Bereitstellung des Themas. Darüber hinaus hat er durch zahlreiche Diskussionen, mit wertvollen Ideen und Ratschlägen sowie einer ausgezeichneten Infrastruktur ein hervorragendes Arbeitsumfeld bereitgestellt. Bei Prof. Dr. Mario Bebendorf möchte ich mich herzlich für die Übernahme des Zweitgutachtens bedanken. Ebenso gebührt ein großer Dank meinen Kollegen Dr. Sven Groß und Dr. Christian Rieger für das Korrekturlesen und die stetige Bereitschaft zur Diskussion inhaltlicher Fragen. Des Weiteren möchte ich mich bei allen Kollegen und Mitarbeitern des Instituts für Numerische Simulation für die freundliche, hilfsbereite und inspirierende Arbeitsatmosphäre bedanken. Nicht zuletzt gilt mein Dank auf persönlicher Ebene meiner Frau Bettina sowie meiner Familie für das entgegengebrachte Verständnis und die Geduld während meiner Promotionszeit.

Bonn, im Oktober 2009

Jürgen Braun

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Kolmogorov's superposition theorem</b>	<b>11</b>
2.1	Variants of Kolmogorov's theorem . . . . .	13
2.2	Geometric interpretation . . . . .	16
2.3	Sprecher's constructive version . . . . .	20
2.3.1	Discontinuity of the inner function . . . . .	21
2.3.2	Correction of the inner function . . . . .	23
2.3.3	Existence of Köppen's function . . . . .	27
2.3.4	Continuity of $\psi$ . . . . .	29
2.3.5	Monotonicity of $\psi$ . . . . .	31
2.3.6	Definitions and reformulation with a single outer function . . . . .	34
2.3.7	Modification of Sprecher's algorithm . . . . .	36
2.4	Relation to Space Filling Curves . . . . .	48
<b>3</b>	<b>Approximative versions of Kolmogorov's superposition theorem</b>	<b>53</b>
3.1	Relations to neural networks and other approximation schemes .	54
3.2	Two models based on Sprecher's theorem . . . . .	60
3.2.1	First model . . . . .	61
3.2.2	Second model . . . . .	62
<b>4</b>	<b>Regularization networks</b>	<b>63</b>
4.1	Reproducing kernel Hilbert spaces . . . . .	65
4.2	Statistical learning theory . . . . .	70
4.3	Structural risk minimization . . . . .	75
4.4	Maximum A Posteriori Interpretation . . . . .	78
4.5	Approximation spaces and norms . . . . .	79
4.5.1	Regularization Network approach for the first model	87
4.5.2	Regularization Network approach for the second model	90
<b>5</b>	<b>Implementational details</b>	<b>95</b>

---

5.1	Nonlinear minimizers . . . . .	95
5.1.1	Numerical complexities for evaluations of $E^{(n)}, \nabla E^{(n)}$	99
5.2	Multiple precision arithmetic . . . . .	102
5.3	Boundary values . . . . .	105
5.4	Further remarks . . . . .	106
<b>6</b>	<b>Numerical results</b>	<b>107</b>
6.1	First model . . . . .	108
6.1.1	B-splines . . . . .	108
	Simple example . . . . .	110
	General example . . . . .	112
	Boundary values . . . . .	114
6.1.2	Trigonometric polynomials . . . . .	116
	Simple example . . . . .	122
	General example . . . . .	122
6.1.3	Locations of the relevant frequencies . . . . .	125
6.2	Second model . . . . .	130
6.2.1	Trigonometric polynomials . . . . .	130
	Simple example . . . . .	135
	Numerical analysis of the nonlinear minimizers . . . . .	137
	Numerical analysis of the model parameters . . . . .	142
	Higher dimensional examples . . . . .	146
<b>7</b>	<b>Conclusions</b>	<b>153</b>
<b>A</b>	<b>Appendix</b>	<b>157</b>
A.1	Duality between RKHS's and stochastic processes . . . . .	157
A.2	Real valued formulation of first model . . . . .	159
A.3	Real valued formulation of second model . . . . .	162
A.4	Further definitions . . . . .	166
	Ill-posed problems . . . . .	166
	Definition of $VC$ -dimension and $V_\gamma$ -dimension . . . . .	167
	<b>Bibliography</b>	<b>169</b>

# Chapter 1

## Introduction

The computation or reconstruction of unknown functions is a task that arises in many mathematical problems of various kinds. Depending on the application, the function has to be computed from different sources like e.g. large discrete data sets, but also from the right hand side of a partial differential equation, the sampling of a random process, or as the description of a hypersurface. Moreover, in many practical applications the dimensionality of the problem is high, see below for examples. For most problems the solution is not available in a closed form or its evaluation is not feasible and one has to resort for approximations which are then computed by adequate numerical methods. Here, the *numerical costs* to compute the solution play an important role. This term describes the number of involved algorithmic operations, the storage requirements of the method itself, i.e. its complexity, and the storage requirements to represent the solution. Now, due to a limitation of these quantities the numerical solution of most mathematical problems is practically difficult to compute in more than three or four dimensions. The reason is the so-called *curse of dimensionality*. This term, which was first coined by Bellman in [8], meanwhile is well known and is used to express the exponential dependency of the involved numerical costs on the dimensionality.

Now, while typically partial differential equations (PDE's) describe a physical process evolving in time in three space dimensions, an example in this context which involves more variables is the solution of the Schrödinger equation, [83]. It describes how the state of a physical system in quantum mechanics changes with time and requires for each electron in the system at least three space variables. Thus the total dimension of the system increases with the number of electrons. More examples for PDE's in higher dimensions stem from stochastic processes like the modelling of a mechanical system with random oscillations or the pricing of financial derivatives. Here, the statistics of the system can be described by the parabolic Fokker–Planck equation, [32, 77, 98, 99, 103]. Clearly, the solution of a PDE requires both the inversion of a partial differential operator and the representation of the solution.

It turns out that in higher dimensions even the latter task is a hard to implement.

To explain this, we follow the argumentation from [45]. Consider a higher-dimensional function  $f : [0, 1]^n \rightarrow \mathbb{R}$  which has to be approximated by a function  $f_a$  with a prescribed accuracy  $\varepsilon > 0$ . Then, the numerical costs to compute  $f_a$  depend exponentially on the dimensionality  $n$  of  $f$ . In fact, we encounter complexities of the order  $\mathcal{O}(\varepsilon^{-n/r})$ . Here, the damping factor  $1/r$  in the exponent results from some isotropic smoothness parameter  $r > 0$  which depends on the respective approach, the smoothness of the function under consideration, the polynomial degree of the ansatz functions and the details of the implementation. For example, let the approximand  $f_a$  be defined on a simple uniform grid as piecewise  $n$ -polynomial function over a bounded domain, and let  $\varepsilon = N^{-r}$ . In this case, one achieves accuracies of the order  $\mathcal{O}(N^{-r})$  with  $\mathcal{O}(N^n)$  grid points or degrees of freedom for  $f_a$ . Clearly, the computational costs and storage requirements for the approximand grow exponentially with the dimensionality  $n$ , and for this reason, the numerical treatment of mathematical problems is often restricted to three or four dimensions, even on the most powerful machines presently available.

To circumvent the curse of dimensionality, we could make the assumption that  $r = c \cdot n$  for some  $c$  independent of  $n$ . In this case we directly see that the numerical costs are of the order  $\mathcal{O}(\varepsilon^{-n/(c \cdot n)}) = \mathcal{O}(\varepsilon^{-1/c})$ , and thus independent of  $n$  while the accuracy is  $\mathcal{O}(\varepsilon^c)$ . This way, the curse of dimensionality could be broken easily. However, this statement has to be handled with care since the order constants still may (exponentially) depend on the dimension  $n$ . Furthermore, such a smoothness assumption is somewhat unrealistic.

While for lower (two or three) dimensional physical applications the shape of the underlying domain is important, the domains become simpler in higher dimensions. Here, typically they have tensor product structure for which one may exemplarily consider the  $n$ -dimensional unit cube  $[0, 1]^n$ .

An important tool for the development of numerical methods in high dimension is the analysis of variance (ANOVA) representation of a function, [127, 135]. There, the function  $f$  which depends on  $n$  variables is decomposed into a sum over  $2^n$  terms such that each term depends only on a unique subset of variables. To explain this, consider a splitting of one-dimensional spaces  $V^{(d)}$ ,  $d = 1, \dots, n$  into the space of constants  $\mathbf{1}^{(d)}$  and the remainder subspace  $W^{(d)}$ , i.e.  $V^{(d)} = \mathbf{1}^{(d)} \oplus W^{(d)}$ . This introduces a natural decomposition of their tensor product space  $V := V^{(1)} \otimes \dots \otimes V^{(n)}$  into  $2^n$  subspaces by

$$V = \bigoplus_{\mathbf{u} \subset \{1, \dots, n\}} \left( \bigotimes_{d \in \{1, \dots, n\} \setminus \mathbf{u}} \mathbf{1}^{(d)} \right) \otimes \left( \bigotimes_{d \in \mathbf{u}} W^{(d)} \right) =: \bigoplus_{\mathbf{u} \subset \{1, \dots, n\}} W_{\mathbf{u}},$$

such that  $V \ni f = \sum_{\mathbf{u} \subset \{1, \dots, n\}} f_{\mathbf{u}}$ , and each function  $f_{\mathbf{u}} \in W_{\mathbf{u}}$  depends only on the

variables  $x_d$ ,  $d \in \mathbf{u}$ . Depending on the initial splitting<sup>1</sup>, this leads to different types of the ANOVA decomposition.

This representation of the function then reveals the relative importance of the respective dimensions and their interactions. It turns out that there exist applications for which the terms  $f_{\mathbf{u}}$  are not of equal importance and decay for increasing  $|\mathbf{u}|$ . Here, see also [45], we mention simulations from molecular dynamics where only potential functions up to the order four, i.e. restrict the sum to terms with  $|\mathbf{u}| \leq 4$ , suffice to represent the potential energy hypersurface of a system. Another example are statistical applications where usually only the covariance of variables is considered and higher order correlations are neglected. This corresponds to the case that  $|\mathbf{u}| \leq 2$ . Furthermore, for data mining, see below, it is found in [35] that even for really high-dimensional data the terms  $f_{\mathbf{u}}$  with  $|\mathbf{u}| > 7$  are not significant. Finally, problems in mathematical finance like option pricing, bond valuation or the pricing of collateralized mortgage backed securities can be formulated as high-dimensional integrals. For these integrals it was shown in [128] that the importance of each dimension is naturally weighted by certain hidden weights, and higher-order terms tend to play a less significant role.

Further results on the curse of dimensionality in this context deal with the tractability of general high-dimensional (and even infinite-dimensional) approximation and integration problems. This was investigated in a series of papers [24, 56, 57, 104–106, 129–131] where the notion of weighted spaces was introduced. In these spaces the importance of successive coordinate directions is quantified by a decaying, or even finite<sup>2</sup>, sequence of weights for the terms  $W_{\mathbf{u}}$ .

This directly leads to the effective dimension of a function  $f$ , see [15]. Here, based on the ANOVA decomposition of  $f$ , one differs between two different types of effective dimension. A function  $f : [0, 1]^n \rightarrow \mathbb{R}$  is said to have superposition dimension  $n_s$  if the sum of the partial variances of the ANOVA terms with  $|\mathbf{u}| \leq n_s$  exceeds 99 percent of the total variance of  $f$ . Alternatively, it has truncation dimension  $n_s$  if the sum over the variances of the terms  $f_{\mathbf{u}}$  with  $\mathbf{u} \subset \{1, \dots, n_s\}$  exceeds this bound. For high-dimensional financial problems it was argued in [128] that they are often of low effective dimension.

A multilevel discretization of the one-dimensional remainder subspaces  $W^{(d)}$ ,  $d = 1, \dots, n$  in the tensor product construction of the ANOVA decomposition leads to sparse grids, see [14], which require smoothness assumptions of bounded mixed derivatives. Then, an involved cost–benefit approach, which is based on the energy–norm to determine the degree of the anisotropic multilevel refinements, results in

---

<sup>1</sup>The splitting is defined by a mapping  $P : V^{(d)} \rightarrow \mathbb{1}^{(d)}$  with  $Pf(x) = \int_0^1 f(x)d\mu(x)$  and some measure  $\mu$ . For example, the Lebesgue measure  $d\mu(x) = dx$  leads to the well known ANOVA decomposition used in statistics, [26]. Another example is the Dirac measure  $d\mu(x) = \delta(x - a)$  located at a point  $a$ . This leads to a representation which is considered in [97] as cut–HDMR.

<sup>2</sup>Here, by finite we mean that for some  $q < n$ , and  $|\mathbf{u}| \geq q$  the corresponding weights are zero.

a sparse grid space, see [45]. Here, the number of involved degrees of freedom as well as the order of the approximation error are independent of the dimension  $n$ . The so-called dimension adaptive sparse grid method has been recently introduced in [40] for integration of higher dimensional functions. There, the placement of integration points for the underlying quadrature rule is automatically controlled depending on the importance of the respective dimension. Recently, sparse grids have also been applied in [31] for principal manifold learning. The task here is to find lower dimensional structures that are embedded in higher dimensional spaces.

We now turn to the previously mentioned concept of data mining which denotes the process of finding hidden patterns, relations and trends in large data sets, see [16]. One problem which arises in this rather comprehensive field is the parameterization of a higher dimensional function which is only known on a discrete data set. Namely, it results from questions in this context which can be interpreted as scattered data approximation problems. It deals with the reconstruction of an unknown function from given scattered data points. These applications arise from various fields such as computer science, geology, biology, meteorology, engineering, stock analysis, or business studies, [38, 39, 121, 132].

In this thesis, we consider the general formulation of this problem, i.e. we assume that data points

$$\mathcal{Z} := \{(\mathbf{x}_j, y_j) : j = 1, \dots, P\} \subset [0, 1]^n \times \mathbb{R}$$

are given, where the values  $y_j$  are the samples  $y_j = f(\mathbf{x}_j)$  of an unknown function  $f : [0, 1]^n \rightarrow \mathbb{R}$  on the set  $\{\mathbf{x}_j\}_{j=1}^P \subset [0, 1]^n$ . The problem is now, to reconstruct the high dimensional function  $f$  from the discrete data  $\mathcal{Z}$ .

There exist various algorithms to solve this problem like multivariate adaptive regression splines (MARS) [35], additive models [48, 49], support vector machine regression [102], projection pursuit algorithms [34, 117], neural networks [49, 84], or radial basis functions [49, 132]. In a series of papers [28, 41, 42, 96] it was shown that these methods can be unified in the general framework of Regularization Networks. Here, the ill-posed problem of reconstructing  $f$  from discrete data is regularized by assuming an appropriate prior on the class of approximating functions. Typically, regularization techniques [118, 119] impose additional smoothness constraints on the approximand  $f_a$  which then is the minimizer of an error functional of the form

$$E(f_a) = \frac{1}{P} \sum_{j=1}^P V(f_a(\mathbf{x}_j), y_j) + \nu \mathcal{R}(f_a).$$

Here, the loss function  $V : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  enforces closeness to the data while the regularization term  $\mathcal{R}(f_a)$  enforces smoothness of  $f_a$ . The regularization parameter  $\nu > 0$  allows for a tradeoff between the two terms. In the Bayesian interpretation of this formulation the stabilizer corresponds to a smoothness prior, and the error

term to a model of the noise in the data which is usually assumed to be Gaussian and additive, see [28, 42, 96, 127].

Typically, smoothness is imposed by a regularization term which has the form  $\mathcal{R}(f) = \|\mathcal{S}f\|_{L_2(\Omega)}$  with some differential operator  $\mathcal{S}$ . It is a well known fact that this can be formalized by means of reproducing kernel Hilbert spaces  $H(K)$  with reproducing kernel  $K$ , see Chapter 4 or [28, 38, 41, 42, 96, 102, 127] for details. In fact, if  $K$  is the Green's function of the operator  $(\mathcal{S}^*\mathcal{S})$ , then the norm  $\|\cdot\|_{H(K)}^2$  in  $H(K)$  coincides with  $\|\mathcal{S}\cdot\|_{L_2(\Omega)}$ . Conversely, it is the flexibility in the choice of the kernel function  $K$  and the loss function  $V$  which allows the interpretation of many well known approximation techniques as Regularization Networks, as mentioned above.

For this reason, we will employ the Regularization Network approach and the reproducing kernel Hilbert setting in this thesis to reconstruct the function from its samples. However, we aim at the use of an efficient representation of the approximand.

To explain this, note that under certain assumptions on  $K$  and  $V$  the minimizer  $f_a$  of  $E(\cdot)$  has the form

$$f_a(\mathbf{x}) := \sum_{j=1}^P c_j K(\mathbf{x}, \mathbf{x}_j)$$

where the coefficient vector  $\mathbf{c} := (c_1, \dots, c_P)^T$  is given as solution of the system of linear equations  $(\mathbf{K} + \nu \mathbf{I}) \mathbf{c} = \mathbf{y}$ , where  $\mathbf{K} := (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^P$ ,  $\mathbf{y} := (y_1, \dots, y_P)^T$ , and  $\mathbf{I}$  denotes the identity matrix. Clearly, the complexity to compute  $f_a$  is of the order  $\mathcal{O}(P^3)$  which leads to high computational costs if the data set  $\mathcal{Z}$  is large. However note that it is independent of the dimension  $n$ .

Thus, to deal with large data sets  $\mathcal{Z}$ , it would be advantageous if the minimizer  $f_a \in H(K)$  could be represented in a more efficient way, which does not depend on  $P$ , whereas the computation of the norm  $\|f_a\|_{H(K)}^2$  still has to be feasible. This would enable us to directly benefit from a comprehensive information on  $f$  in form of many function samples. The relation between the kernel approach and a sparse grid representation of the approximand was investigated in [38].

However, we have already seen that the representation of a function in general suffers from the curse of dimensionality and higher regularity has to be assumed for  $f$  to circumvent it at least to some extend. In this thesis, instead of increasing the regularity assumptions on  $f$  or assuming a lower effective dimension, we will use a representation of the higher-dimensional function as superposition of lower-dimensional functions. In 1957 the Russian mathematician A. N. Kolmogorov showed in [69] the following

**Theorem.** *Let  $f : [0, 1]^n \rightarrow \mathbb{R}$  be an arbitrary multivariate continuous function.*

Then it has the representation

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} g_q \left( \sum_{p=1}^n \psi_{q,p}(x_p) \right), \quad (1.1)$$

with continuous one-dimensional outer and inner functions  $g_q$  and  $\psi_{q,p}$ . All these functions  $g_q, \psi_{q,p}$  are defined on the real line. The inner functions  $\psi_{q,p}$  are independent of the function  $f$ .

Thus, with this result rather than directly approximating the  $n$ -dimensional function  $f$  we have to find  $2n + 1$  one-dimensional functions which approximate the outer functions  $g_1, \dots, g_{2n+1}$ . We will see later that this can even be reduced to the approximation of one single outer function  $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ .

Kolmogorov's theorem is closely related to Hilbert's 13th problem which involves the study of solutions of algebraic equations. It is a well known fact that the solution of an algebraic equation of degree  $\leq 4$  can be computed by formulae that only contain radicals and arithmetic operations and therefore by functions of one or two variables. In his Paris lecture in 1900 Hilbert conjectured [58, 126] that “*A solution of the general equation of degree 7 cannot be represented as a superposition of continuous functions of two variables.*” The question was answered by Kolmogorov in 1957, [69]. He showed that any continuous function of several variables can be represented as a superposition of continuous functions of one variable and the operation of addition. Kolmogorov's student Arnold showed even before in [1–3] that any  $f \in C([0, 1]^3)$  can be represented as a superposition of continuous functions in two variables, and thus refuted Hilbert's conjecture.

Several improvements of Kolmogorov's original version were published in the following years. Lorentz showed that the outer functions  $g_q$  can be chosen to be the same [81, 82] while Sprecher proved that the inner functions  $\psi_{q,p}$  can be replaced by  $\lambda_p \psi_q$  with appropriate constants  $\lambda_p$  [107, 109]. A proof of Lorentz's version with one outer function that is based on the Baire category theorem was given by Hedberg [55] and Kahane [68]. A further improvement was made by Friedman [33], who showed that the inner functions can be chosen to be Lipschitz continuous. A geometric interpretation of the theorem is that the  $2n + 1$  inner sums  $\sum_{p=1}^n \psi_{q,p}$  map the unit cube  $[0, 1]^n$  homeomorphically onto a compact set  $\Gamma \subset \mathbb{R}^{2n+1}$ . The fact that any compact set  $K \subset \mathbb{R}^n$  can be homeomorphically embedded into  $\mathbb{R}^{2n+1}$  was already known from the Menger–Nöbeling theorem [61]. Ostrand [92] and Tikhomirov [70] extended Kolmogorov's theorem to arbitrary  $n$ -dimensional metric compact sets.

More recently, Kolmogorov's superposition theorem found attention in neural network computation by Hecht–Nielsen's interpretation as a feed-forward network with an input layer, one hidden layer and an output layer [52, 53, 107]. However, the inner functions in all these versions of Kolmogorov's theorem are highly non-smooth. Also, the outer functions depend on the specific function  $f$  and hence are

---

not representable in a parameterized form. Moreover, all one-dimensional functions are the limits or sums of some infinite series of functions, which cannot be computed practically. Therefore Girosi and Poggio [43] made the criticism that such an approach is not applicable in neurocomputing. The reason for this is that the original proof of Kolmogorov's theorem is not constructive, i.e. one can show the existence of a representation (1.1) but it cannot be used in an algorithm for numerical calculations. Kurkova [75, 76] partly eliminated these difficulties by substituting the exact representation in (1.1) with an approximation of the function  $f$ . She replaced the one-variable functions with finite linear combinations of affine transformations of a single arbitrary sigmoidal function  $\psi$ . Her direct approach also enabled an estimation of the number of hidden units (neurons) as a function of the desired accuracy and the modulus of continuity of  $f$  being approximated. In [84] a constructive algorithm is proposed that approximates a function  $f$  to any desired accuracy with one single design, which means that no additional neurons have to be added. There, also a short overview of the history of Kolmogorov's superposition theorem in neural network computing is given. Other approximative, but constructive approaches to function approximation by generalizations of Kolmogorov's superposition theorem can be found in [22, 64, 85].

Recently, Sprecher derived in [111, 112] a numerical algorithm for the implementation of the external univariate functions, which promises to constructively prove Kolmogorov's superposition theorem. In these articles, the inner functions  $\psi_q$  are defined as translations of a single function  $\psi$  that is explicitly defined as an extension of a function which is defined on a dense subset of the real line. Sprecher proved convergence of this algorithm in [111, 112]. Throughout this proof, he relied on continuity and monotonicity of the resulting  $\psi$ . It can however be shown that his  $\psi$  does not possess these important properties. This was already observed by Köppen in [73] where a modified inner function  $\psi$  was suggested. Köppen claims, but does not prove the continuity of his  $\psi$  and merely comments on the termination of the recursion which defines his corrected function  $\psi$ . These gaps will be closed here, see also [12], i.e. existence as well as the essential properties of continuity and monotonicity of Köppen's  $\psi$  are shown. Following the lines of [111, 112], convergence of Sprecher's algorithm was proved in [12] which finally confirmed his constructive result.

Our goal is now to develop an approach to function approximation that is capable to brake the curse of dimensionality at least to some extend without higher regularity assumptions on the function  $f$ . To this end, we first modify the constructive proof of Kolmogorov's superposition theorem due to Sprecher [12, 73, 111, 112] to a numerically preferable new version with only one outer function which has the

following form:

$$f(x_1, \dots, x_n) = \sum_{q=0}^m \Phi \left( \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q \right).$$

Then, we alter the original algorithm such that the new version computes a sequence of functions  $\{\Phi^r\}_r$  for which then convergence to the outer function  $\Phi$  is shown when  $r \rightarrow \infty$ . This constructively proofs the new version of Kolmogorov's theorem.

We suggest a Regularization Network approach which is based on this representation to efficiently reconstruct an unknown  $n$ -dimensional continuous function  $f$  from known sampled points. In a preliminary step, we replace the outer function in the new representation by an expansion in a finite basis with unknown coefficients. This results in a linear representation of the approximand. The coefficients are then determined via a variational formulation, i.e. the minimization of the regularized convex cost functional  $E^{(1)}(\cdot)$  which measures the empirical error on a random sample set. Further numerical investigations of this first model and its Fourier transform will show that Sprecher's inner function causes an unfolding of the dimensions into a product of functions. Moreover, the relevant frequency numbers depend only on model parameters which are independent of the function  $f$  and can be computed a priori. This motivates our second model in which we replace the outer function by a product of functions and expand each factor in a Fourier series with unknown coefficients. The second model takes the form

$$f_L(\mathbf{x}) := \sum_{q=0}^m \prod_{d=0}^n \left( \sum_{j=-N_d}^{N_d} c_j^d \exp \left[ 2\pi i k_j^d \left( \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q \right) \right] \right),$$

for known frequency numbers  $k_j^d \in \mathbb{Z}$ . The previously described procedure will be formulated by means of reproducing kernel Hilbert spaces which then form the basis of the respective models. This directly enables the definition of corresponding feasible norms in the regularization term.

Altogether, we thus obtain an efficient algorithm for function reconstruction from sample points in the spirit of a Regularization Network [22, 43, 52, 53, 62–64, 75, 76, 84, 85]. The number of degrees of freedom of our algorithm depends linearly on the dimension of the function to be approximated. Also, and in contrast to the general Regularization Networks, the complexity depends linearly on the number  $P$  of learning points. However, the resulting cost functional  $E^{(n)}(\cdot)$  for the second model will no longer be convex and has to be minimized with a nonlinear line search method. Additionally, special care has to be taken in the implementation of the algorithm. It turns out that the dimensionality of the function  $f$  is transformed by Kolmogorov's representation into an oscillation of the outer function  $\Phi$ . Thus, to resolve the strong oscillations and high frequency numbers the arithmetic precision

---

has to be increased beyond machine precision in an implementation. This is where the curse of dimensionality still might be present, even though we are approximating only a one-dimensional function.

To sum up, the contributions of this thesis<sup>3</sup> are the mathematical proofs of existence, continuity and monotonicity of Köppen's inner function  $\psi$ , as well as the formulation of a new version of Kolmogorov's superposition theorem with one single outer function  $\Phi$  which results from modifications of Sprecher's constructive theorem. Accordingly, Sprecher's algorithm is adapted in a suitable way and its convergence towards the outer function  $\Phi$  is proved. Furthermore, two models for an approximation of functions from discrete data samples are presented which directly use the new version of Kolmogorov's theorem. The corresponding Regularization Network approaches are developed by means of reproducing kernel Hilbert spaces. Finally, results of numerical experiments for the new models are presented.

The remainder of this thesis is organized as follows: In Chapter 2 we consider the Kolmogorov superposition theorem and present its various improvements by different authors. This includes Sprecher's constructive version for which we show that his inner function  $\psi$  is discontinuous while existence, continuity, and monotonicity can be shown for Köppen's corrected function  $\psi$ . A reformulation of Sprecher's results leads to a new version with only one single outer function. It requires a modification of the algorithm for which then convergence is shown. Furthermore, we present a geometric interpretation of Kolmogorov's superposition theorem and comment on the unfolding of dimensions that is caused by the inner functions. Chapter 3 gives some examples of approximation schemes that are closely related to Kolmogorov's superposition theorem. Here, also our regression models will be briefly introduced. In Chapter 4, the Regularization Network approach will be developed. To this end, basic facts on reproducing kernel Hilbert spaces, Statistical learning theory, Structural risk minimization, and a Bayesian interpretation of Regualrization Networks are presented. Finally, the underlying reproducing kernel Hilbert spaces with the corresponding norms are derived for our approach. Chapter 5 deals with details on the implementation of the method. Here, after a brief description of the iterative nonlinear minimizers, the numerical complexities for each iteration step are derived. Furthermore, we introduce a nested iteration scheme and explain the necessity of a multiple precision arithmetic in more detail. Then, a transformation of the problem to a sub cube of  $[0, 1]^n$  is introduced which is due to numerical problems at the boundary of the unit cube. Finally, some remarks on a parallelization of the algorithm are made. In Chapter 6 the results for various numerical experiments are given. For the first model, the results will show that the use of locally supported basis functions like B-splines is disadvantageous. The numerical results for the Fourier basis will then provide important insight into the structure of the outer function  $\Phi$ .

---

<sup>3</sup>The mathematical proofs of existence, continuity and monotonicity of Köppen's inner function  $\psi$  have already been published in [12].

In particular, they reveal its product structure and the dependency of the relevant frequency numbers from several model parameters. The product structure of  $\Phi$  in combination with the fact that the locations of the important frequency numbers in Fourier space are independent of the function  $f$ , i.e. they can be computed a priori, then motivates the definition of the second model. A detailed numerical analysis for this model will show that it is capable of approximating functions up to dimension ten and that the product ansatz for the outer function in our new version of Kolmogorov's superposition theorem is an appropriate choice.

# Chapter 2

## Kolmogorov's superposition theorem

This section introduces Kolmogorov's superposition theorem and subsequent versions that were developed by different authors. Therefore, as a starting point we will define what is meant by **superpositions of functions** in this context. We say that a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a superposition of functions  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  and  $h_j : \mathbb{R}^{\ell_j} \rightarrow \mathbb{R}$ ,  $j = 1, \dots, k$  for some values  $k, \ell_1, \dots, \ell_k \in \mathbb{N}$ , if it has the representation

$$f(x_1, \dots, x_n) = g(y_1, \dots, y_k) \quad \text{with} \quad y_j = h_j(x_{i_1^j}, \dots, x_{i_{\ell_j}^j}), \quad j = 1, \dots, k,$$

where for fixed  $j \in \{1, \dots, k\}$ ,  $\ell_j \leq n$ , and the indices

$$i_k^j \in \{1, \dots, n\}, \quad k = 1, \dots, \ell_j,$$

are pairwise distinct. Here, we explicitly allow the functions  $h_1, \dots, h_k$  to be superpositions again. To make the reader familiar with this definition we give two simple examples.

- (a) The function  $f(x, y) = x^y$  is a superposition of  $g(s, t) = \exp(s \cdot t)$  and  $h(s) = \ln(s)$ . We have  $f(x, y) = g(h(x), y)$ . Here in turn,  $g$  can be represented as superposition of  $v(s) = \exp(s)$  and  $w(s, t) = s \cdot t$  which leads to  $f(x, y) = v(w(h(x), y))$ .
- (b) Clearly, the addition of three numbers  $f(x, y, z) = x + y + z$  can be represented as a superposition of the addition of two numbers  $g(s, t) = s + t$ . By the associative law we can write  $f(x, y, z) = g(x, g(y, z)) = g(g(x, y), z)$ .

These simple examples already show that a representation by superpositions is not unique. We remark that the understanding of a superposition of functions sometimes differs in literature, e.g. when speaking about the superposition of plane waves.

Kolmogorov's theorem deals with the representation of continuous functions of  $n$  variables as superposition of functions that only depend on one variable and

addition. This is closely related to Hilbert's 13th problem. In his Paris lecture at the International Conference of Mathematicians in 1900, David Hilbert formulated 23 problems which in his opinion were important for the further development of mathematics, [58]. The 13th of these problems dealt with the solution of general equations of higher degrees. It is known that for algebraic equations of degree  $\leq 4$  the solution can be computed by formulae that only contain radicals and arithmetic operations. For linear equations this fact is known since ancient times while the solution of quadratic equations was found by Babylonians. In the 16th century, Tartaglia, Cardan, and Ferrari showed how to solve cubic and quartic equations. For higher orders, Galois' theory shows us that the solutions of algebraic equations cannot be expressed in terms of basic algebraic operations.

The roots of polynomial equations can be interpreted as functions of their coefficients. In the following we mean by solution just a single valued branch of the general solution. It should be noted that it is not clear that this is precisely what Hilbert meant. Now, we already know that the solutions of algebraic equations up to degree four are representable by functions of one or two variables. For higher degrees it follows from the so-called Tschirnhausen transform that the general algebraic equation  $x^n + a_1x^{n-1} + \dots + a_n = 0$  can be translated to the form  $y^n + b_{n-4}y^{n-4} + \dots + b_1y + 1 = 0$ . The Tschirnhausen transform is given by a formula containing only radicals and arithmetic operations and transforms. Therefore, the solution of an algebraic equation of degree  $n$  can be represented as a superposition of functions of two variables if  $n < 7$  and as a superposition of functions of  $n - 4$  variables if  $n \geq 7$ . For  $n = 7$  the solution is a superposition of arithmetic operations, radicals, and the solution of the equation  $z^7 + c_3z^3 + c_2z^2 + c_1z + 1 = 0$ . A further simplification with algebraic transformations seems to be impossible which led to Hilbert's conjecture [58, 126] that "*A solution of the general equation of degree 7 cannot be represented as a superposition of continuous functions of two variables.*" This explains the relation of Hilbert's 13th problem to the representation of a higher-dimensional function as superposition of lower-dimensional functions. In this context, it has stimulated many studies in the theory of functions and other related problems by different authors. It was due to the fact that no one doubted the validity of Hilbert's conjecture that one studied superpositions of *continuous* functions at first. Hilbert himself studied the superposition of algebraic functions [59] while Wiman and Chebotarev (see [126] and the references therein) thought of possible developments in Galois theory. Kolmogorov first studied the theory of analytic functions and then focussed on smooth functions. We refer to [126] for a more detailed description on relations of Hilbert's 13th problem to other topics and more references.

Finally, Hilbert's conjecture was proved wrong by Kolmogorov in 1957, [69]. He showed that any continuous function of several variables can be represented as a superposition of continuous functions of one variable and the operation of addition. Kolmogorov's student Arnold showed even before in [1–3] that any  $f \in C([0, 1]^3)$

can be represented as a superposition of continuous functions in two variables. Kolmogorov proved in his paper [69] that for any arbitrary continuous  $n$ -dimensional function  $f : [0, 1]^n \rightarrow \mathbb{R}$  there exist one-dimensional functions  $g_q$ ,  $q = 1, \dots, 2n + 1$  and increasing functions  $\psi_{q,p} : \mathbb{R} \rightarrow \mathbb{R}$ ,  $q = 1, \dots, 2n + 1$ ,  $p = 1, \dots, n$  such that

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} g_q \left( \sum_{p=1}^n \psi_{q,p}(x_p) \right). \quad (2.1)$$

Here, the inner functions  $\psi_{q,p}$  are independent of  $f$  while the outer functions  $g_q$  depend on  $f$ . This result settled Hilbert's 13th problem for *continuous* functions. In the following years, various versions of Kolmogorov's theorem were proved. We will show these results in the next section but do not claim that this list is exhaustive.

## 2.1 Variants of Kolmogorov's theorem

In Kolmogorov's theorem and all subsequent versions that will follow, the representation of an  $n$ -dimensional function  $f$  by superpositions of one-dimensional functions is similar to the form in (2.1). Therefore we will always denote by **outer functions** the functions  $g_q$  which take a sum over **inner functions**  $\psi_{q,p}$  as argument. Note that in all versions of Kolmogorov's theorem which will be presented in the following the respective inner functions are independent of the  $n$ -dimensional function while the outer functions depend on  $f$ .

We start with a variant of Kolmogorov's theorem that reduces the number of outer functions and is due to Lorentz. He showed in 1962 [80] that the outer functions  $g_q$  in (2.1) can be replaced by a single function  $g$ . More precisely, Lorentz proved the existence of functions  $g$  and  $\psi_{q,p}$ ,  $q = 1, \dots, 2n + 1$ ,  $p = 1, \dots, n$  such that

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} g \left( \sum_{p=1}^n \psi_{q,p}(x_p) \right). \quad (2.2)$$

Additionally, Lorentz stated, as it was remarked in [108], that the  $q$ -dependence of the outer functions can always be eliminated by a shift of their argument. Since this insight will be crucial for one main result of this thesis we explain this in more detail in Section 2.3.6.

In [107] Sprecher additionally replaced the inner functions  $\psi_q$  in (2.2) by one single inner function with an appropriate shift in its argument. He proved that there exist real values  $\eta$ ,  $\lambda_1, \dots, \lambda_n$ ,  $\lambda_{p,q}$ ,  $p = 1, \dots, n$ ,  $q = 0, \dots, 2n$ , a function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , and a real, monotonic increasing function  $\psi : [0, 1] \rightarrow [0, 1]$  with  $\psi \in \text{Lip}(\ln 2 / \ln(2N + 2))$ ,  $N \geq n \geq 2$ , such that

$$f(x_1, \dots, x_n) = \sum_{q=0}^{2n} g \left( \sum_{p=1}^n \lambda_p \psi(x_p + \eta q) + q \right), \quad (2.3)$$

and

$$f(x_1, \dots, x_n) = \sum_{q=0}^{2n} g \left( \sum_{p=1}^n \lambda_{p,q} \psi(x_p + \eta q) \right). \quad (2.4)$$

Note that the values  $\eta$ , and the functions  $g, \psi$  can differ in the two representations. In this paper, also a necessary condition on functions  $h_0, \dots, h_m$  is derived for a general representation of the form

$$f(\mathbf{x}) = \sum_{q=0}^m g(h_q(\mathbf{x}))$$

to hold for fixed  $h_q$  which are independent of  $f$ . The property is called **separation of points** and reads as follows: For any two points  $\mathbf{x} \neq \mathbf{y}$  the set  $(h_0(\mathbf{x}), \dots, h_m(\mathbf{x}))$  is not a permutation of  $(h_0(\mathbf{y}), \dots, h_m(\mathbf{y}))$ . Clearly, if two points  $\mathbf{x}_1$  and  $\mathbf{x}_2$  were not separated by  $h_0, \dots, h_m$ , then any function with  $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$  would not be representable in this form. Note that this is not a sufficient condition and additional properties are needed to guarantee the existence of a representation of any function in the special form under consideration here.

The versions (2.2), (2.3) and (2.4) with one single outer function relate Kolmogorov's theorem to Milutin's theorem, see [133], which states that the space  $\mathcal{C}(K)$  is isomorphic to  $\mathcal{C}([0, 1])$  if  $K$  is a compact, metric, uncountable space. Clearly, if  $K := [0, 1]^n$  all representations map the single outer function  $g \in \mathcal{C}([0, 1])$  to an element in  $\mathcal{C}(K)$ . However, they do not provide an inverse mapping since uniqueness of the outer function is not guaranteed in any version of Kolmogorov's theorem.

We have seen that the number of functions in Kolmogorov's theorem is not minimal and that there exist different variants of (2.1) where the number of inner and outer functions is significantly reduced. Another question is, whether the number of  $2n+1$  outer terms in any of the representations can be reduced? This was answered in the negative by Sternfeld in 1985, [116]. He showed that any basis family for compact metric spaces contains at least  $2n+1$  elements.

A result concerning the smoothness of the one-dimensional functions in Kolmogorov's superposition theorem is due to Fridman. In 1967 [33] he could show that the inner functions  $\psi_{q,p}$  in the original version (2.1) can be chosen to be in the class  $\text{Lip}(1)$ . Then, in a further paper Sprecher could prove that  $\psi \in \text{Lip}(1)$  is also admissible in his representations (2.3) and (2.4) with only one inner function, [109]. A negative result by Vitushkin from 1964 [125, 126] dealed with the differentiability of the functions  $g_q$  and  $\psi_{q,p}$  in (2.1). He showed for  $r \geq 2$  and  $n \geq 2$  that there exist  $r$  times continuously differentiable functions in  $n$  variables which cannot be represented by  $r$  times continuously differentiable functions with less variables. A consequence of this result is that the inner functions  $\psi_q$ , since they are independent of  $f$ , cannot be continuously differentiable. This even holds if one wants to represent an analytic function  $f$  only, see [125].

A generalization of Kolmogorov's superposition theorem to compact metric spaces by Ostrand [92] from 1965 reads as follows: For  $p = 1, \dots, m$  let  $X^p$  be compact metric spaces of finite dimension  $n_p$ , and let  $n = \sum_{p=1}^m n_p$ . Then, there exist continuous functions  $\psi_{q,p} : X^p \rightarrow [0, 1]$ ,  $q = 1, \dots, 2n+1$ ,  $p = 1, \dots, n$  and real continuous functions  $g_q : [0, 1] \rightarrow \mathbb{R}$ ,  $q = 1, \dots, 2n+1$  such that any continuous function  $f : X^1 \times \dots \times X^m \rightarrow \mathbb{R}$  is representable in the form

$$f(\mathbf{x}_1, \dots, \mathbf{x}_m) = \sum_{q=1}^{2n+1} g_q \left( \sum_{p=1}^m \psi_{q,p}(\mathbf{x}_p) \right). \quad (2.5)$$

A more general form for the  $n$ -dimensional unit cube is due to Hedberg and Kahane. In [55, 66] they gave a completely different proof of the representation

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} g \left( \sum_{p=1}^n \lambda_p \psi_q(x_p) \right), \quad (2.6)$$

that is based on the Baire category theorem, see also [68, 82]. They showed that for rationally independent<sup>1</sup>, positive numbers  $\lambda_1, \dots, \lambda_n$  with  $\sum_{p=1}^n \lambda_p = 1$  and quasi-all<sup>2</sup> tuples  $(\psi_1, \dots, \psi_{2n+1})$  with continuous, non-decreasing entries  $\psi_q : [0, 1] \rightarrow [0, 1]$ , there exists a continuous function  $g : [0, 1]^n \rightarrow [0, 1]$ , such that for any continuous function  $f : [0, 1]^n \rightarrow [0, 1]$  equation (2.6) holds. Employing Kahane's ideas, Doss proved in 1976 [25] that the inner sum can also be replaced by a product of functions.

Using duality arguments, Sternfeld presents in [115] relatively short and simple proofs of Kolmogorov's original theorem (2.1) and its generalization (2.5) by Ostrand.

In 1978 Sternfeld showed a result for compact metric spaces  $X$  and bounded functions, [114]. He proved that if any continuous function  $f$  on  $X$  can be represented as

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} g_q(\Psi_q(x_1, \dots, x_n)) \quad (2.7)$$

with continuous functions  $g_q, \Psi_q$ ,  $q = 1, \dots, 2n+1$ , then also any bounded function  $f$  on  $X$  can be represented in the form (2.7) with bounded outer functions  $g_q$ , depending on  $f$ .

As a generalization of this result Ismailov studied in [65] superpositions of functions without involving any topology, like continuity or boundedness, on  $X$ . He established a necessary and sufficient condition on a general set  $X$  which is equivalent to the fact that any function  $f$  on  $X$  can be represented in the form (2.7). To be

---

<sup>1</sup>Real numbers  $\lambda_1, \dots, \lambda_n$  are called rationally independent, if for any rational numbers  $r_1, \dots, r_n$  it follows from  $r_1\lambda_1 + \dots + r_n\lambda_n = 0$  that  $r_1 = \dots = r_n = 0$ .

<sup>2</sup>A property holds for quasi-all points in a complete metric space  $X$ , if it holds for all points of a set  $U \subset X$  and  $X \setminus U$  is of the first Baire category.

more precise, there have to exist no closed paths in  $X$  with respect to  $\Psi_1, \dots, \Psi_{2n+1}$ .<sup>3</sup> Ismailov showed that for smooth functions with simple structure, like e.g. Ridge functions, the set  $X \subset \mathbb{R}^n$  always has closed paths while for continuous functions  $\Psi_q$ ,  $q = 1, \dots, 2n + 1$ , many sets of  $\mathbb{R}^n$ , like the unit cube, any compact subset, or even  $\mathbb{R}^n$  itself, may have no closed paths.

As we have seen, there exist many versions of Kolmogorov's superposition theorem that either improved its general complexity by reducing the number of involved functions, improved the smoothness of the one-dimensional functions, or generalized the proposition to other spaces or sets. However, all previously shown results have in common that they are not constructive: A concrete formula for any of the inner or outer functions is unknown. This deficiency was first partly eliminated by Sprecher in [110]. Here, he recursively defined a single inner function  $\psi : [0, 1 + (1/5!)] \rightarrow [0, 1 + (1/5!)]$  that does not depend on  $n$ . Then, depending on  $\psi$ , he showed the existence of  $2n + 1$  systems of disjoint  $n$ -dimensional sub cubes with decreasing diameters covering  $[0, 1]^n$  in a prescribed manner such that previous results from [107] guaranteed the existence of the outer functions. Finally, in [111, 112] Sprecher gave a completely constructive version of Kolmogorov's theorem by explicitly defining a function  $\psi : [0, 1] \rightarrow [0, 1]$ , different from that in [110]. Additionally he introduced a construction algorithm for the outer functions  $\Phi_q$  in a representation of the form

$$f(x_1, \dots, x_n) = \sum_{q=1}^m \Phi_q \left( \sum_{p=1}^n \lambda_p \psi(x_p + qa) \right),$$

see (2.9) below. Sprecher's proof relied, amongst other things, on the fact that this  $\psi$  is continuous and monotone increasing. It was Köppen in 2002, [73], who first remarked that this is actually not the case. Here, he corrected Sprecher's inner function  $\psi$  through a point wise, recursive definition on the dense set of terminating rational numbers  $\mathcal{D} \subset \mathbb{R}$  and claimed that his function exists, and that it is continuous and monotone increasing. However, Köppen did not prove these properties but argued that Sprecher's constructive representation holds with the new  $\psi$ . It is one intention of this thesis to show that this is indeed the case and that Sprecher's ideas in general hold, see also [12]. We present a detailed mathematical investigation of this fact in Section 2.3.

## 2.2 Geometric interpretation

At this point we rather continue with a geometric interpretation of (2.6). The contents of this section are taken from [68], see also [66]. Let  $\lambda_p$ ,  $p = 1, \dots, n$  and

---

<sup>3</sup>For a formal definition of closed paths in a set  $X$  with respect to a set of functions  $h_1, \dots, h_r$  we refer to [65].

$\psi_q$ ,  $q = 1, \dots, 2n + 1$  be the same as in (2.6). We remind the reader that these coefficients and the inner functions are independent from  $f$ . Now, we can define a continuous embedding  $E : [0, 1]^n \rightarrow \mathbb{R}^{2n+1}$  by the coordinate functions

$$X_q := E_q(x_1, \dots, x_n) := \sum_{p=1}^n \lambda_p \psi_q(x_p), \quad q = 1, \dots, 2n + 1.$$

The image of  $[0, 1]^n$  in  $\mathbb{R}^{2n+1}$  is a compact set, denoted by  $\Gamma$ , and from (2.6) we see that any  $f \in \mathcal{C}([0, 1]^n)$  has the representation

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} g(E_q(x_1, \dots, x_n)). \quad (2.8)$$

By definition, the embedding  $E : [0, 1]^n \rightarrow \Gamma$  is surjective. We have already seen, see Section 2.1 and [107], that a necessary condition for the representation (2.6) to hold is that the coordinate functions  $E_1, \dots, E_{2n+1}$  separate all points in  $[0, 1]^n$ , i.e. for  $\mathbf{x} \neq \mathbf{y}$  we can conclude that  $E(\mathbf{x}) \neq E(\mathbf{y})$ . Hence,  $E : [0, 1]^n \rightarrow \Gamma$  is injective and therefore it is a homeomorphism.

Now, for any given function  $F \in \mathcal{C}(\Gamma)$  we can define  $f := F \circ E : [0, 1]^n \rightarrow \mathbb{R}$  and see that  $f \in \mathcal{C}([0, 1]^n)$ . This implies that  $f$  has a representation of the form (2.8) what shows that  $F$  can be represented on  $\Gamma$  as

$$F(X_1, \dots, X_{2n+1}) = \sum_{q=1}^{2n+1} g(X_q).$$

In other words, the subspace

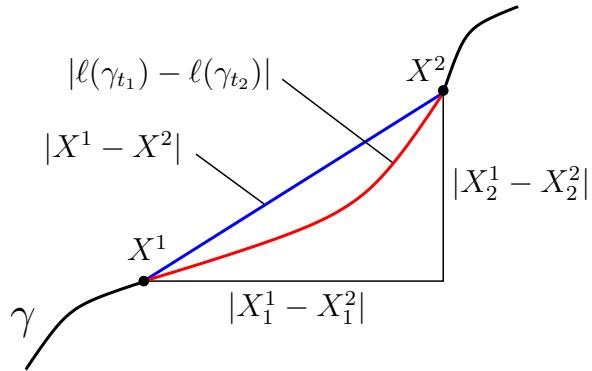
$$Z := \left\{ F \in \mathcal{C}(\mathbb{R}^{2n+1}) : F(X_1, \dots, X_{2n+1}) = \sum_{q=1}^{2n+1} g(X_q), g \in \mathcal{C}(\mathbb{R}) \right\}$$

interpolates freely<sup>4</sup> on  $\Gamma$ , i.e. the image of  $[0, 1]^n$  under the homeomorphism  $E$ .

In a further result, Ostrand [92] and Tikhomirov [70] extended Kolmogorov's theorem to arbitrary  $n$ -dimensional metric compact sets  $K$ , see also [68]. Namely, there exists a homeomorphic embedding  $\varphi : K \rightarrow \mathbb{R}^{2n+1}$  such that the subspace  $Z \subset \mathcal{C}(\mathbb{R}^{2n+1})$  of functions interpolates freely on the compact set  $\Gamma = \varphi(K)$ . A theorem of Menger–Nöbeling, see [61], states that any  $n$ -dimensional compact set  $K$  can be embedded homeomorphically into the  $2n + 1$  dimensional Euclidean space.

---

<sup>4</sup>Let  $Y$  be a subset of the space  $\mathcal{C}(K)$  of continuous functions on a compact set  $K$  and let  $E \subset K$  be a closed subset. Then  $Y$  is said to interpolate freely on  $E$  with a constant  $c \geq 1$ , if for each  $H(x) \in \mathcal{C}(E)$  there exists  $g(x) \in Y$  such that  $g(x) = H(x)$ ,  $x \in E$  and  $\|g\|_{\mathcal{C}(K)} \leq c\|H\|_{\mathcal{C}(E)}$ .



**Figure 2.1.** The Euclidean distance between the points  $X^1$  and  $X^2 \in \mathbb{R}^{2n+1}$  (blue line) is smaller than the arc length of their connection on the curve  $\gamma$  (red line). Clearly, for the coordinates  $|X_q^1 - X_q^2| \leq |X^1 - X^2|$  holds.

Therefore, Kolmogorov's theorem is a special case but this version gives more, since the compact set  $\Gamma$  has additional properties.

Next, we will comment on the geometric structure of the set  $\Gamma$ . We can assume, see [68], that the functions  $\psi_q$ ,  $q = 1, \dots, 2n+1$  in (2.6) are strictly increasing. Now, let  $\zeta : [0, 1] \rightarrow \mathbb{R}^{2n+1}$  with

$$\zeta(t) := (\psi_1(t), \dots, \psi_{2n+1}(t)), \quad t \in [0, 1]$$

be the parametric representation of a continuous curve  $\gamma$  in  $\mathbb{R}^{2n+1}$ . Since the functions  $\psi_q$  are increasing, they have bounded variation and the curve  $\gamma$  is rectifiable. We will denote by  $\gamma_t$  the partial curve with parameterization  $\zeta(\tau)$ ,  $\tau \in [0, t]$ . Let  $s(t) = \ell(\gamma_t)$  be the arc length of  $\gamma_t$ . Then,  $(\zeta \circ s^{-1})(\tau)$ ,  $\tau \in [0, s(1)]$  is the re-parameterization of  $\gamma$  with respect to the arc length parameter. With the rescaling  $p(\tau) = \tau/s(1)$  we can define  $\varphi_q := (\psi_q \circ s^{-1} \circ p^{-1})$ ,  $q = 1, \dots, 2n + 1$  and get by

$$\varphi(\sigma) := (\varphi_1(\sigma), \dots, \varphi_{2n+1}(\sigma)), \quad \sigma \in [0, 1]$$

a further parameterization of the curve  $\gamma$ . Now for  $i \in \{1, 2\}$ , let  $\sigma_i \in [0, 1]$  and  $X^i := \varphi(\sigma_i) \in \mathbb{R}^{2n+1}$  be the corresponding points on the curve. Then, the partial curve with endpoint  $X^i$  is given by  $\gamma_{t_i}$ , where  $t_i := (s^{-1} \circ p^{-1})(\sigma_i)$ , and its arc length is  $\ell(\gamma_{t_i}) = s(t_i) = s(1)\sigma_i$ . Since we know that the Euclidean distance between the points  $X^1$  and  $X^2$  in  $\mathbb{R}^{2n+1}$  is smaller than the arc length of their connection on the curve, see Figure 2.1, we can estimate

$$|\varphi_q(\sigma_1) - \varphi_q(\sigma_2)| = |X_q^1 - X_q^2| \leq |X^1 - X^2| \leq |\ell(\gamma_{t_1}) - \ell(\gamma_{t_1})| = s(1) |\sigma_1 - \sigma_2|.$$

This shows that the functions  $\varphi_q$ ,  $q = 1, \dots, 2n + 1$  satisfy the Lipschitz condition of order one.

Next, we will show that the functions  $\varphi_q$  are also admissible in Kolmogorov's representation. From (2.6) we see that the set  $\Gamma$  is a convex combination of  $n$  copies  $\gamma^p$ ,  $p = 1, \dots, n$  of the curve  $\gamma$ , i.e. we have  $\Gamma = \lambda_1\gamma^1 + \dots + \lambda_n\gamma^n$ . For  $(X_1, \dots, X_{2n+1}) \in \Gamma$ , there exist unique points  $(x_1, \dots, x_n), (\sigma_1, \dots, \sigma_n) \in [0, 1]^n$  such that

$$X_q = \sum_{p=1}^n \lambda_p \psi_q(x_p) = \sum_{p=1}^n \lambda_p \varphi_q(\sigma_p), \quad q = 1, \dots, 2n+1.$$

Therefore, we have the maps

$$\begin{array}{ccc} (x_1, \dots, x_n) & \longleftrightarrow & (\sigma_1, \dots, \sigma_n) \\ & \searrow & \swarrow \\ & (X_1, \dots, X_{2n+1}) & \end{array}$$

where the diagonal arrows are homeomorphisms of  $[0, 1]^n$  onto  $\Gamma$ , while the horizontal one is a homeomorphism of the  $n$ -dimensional unit cube onto itself. Now, let  $F : [0, 1]^n \rightarrow \mathbb{R}$  be an arbitrary continuous function. We can define  $f \in \mathcal{C}([0, 1]^n)$  by setting

$$f(x_1, \dots, x_n) := F(\sigma_1, \dots, \sigma_n), \quad \text{with } \sigma_i = (p \circ s)(x_i), \quad i = 1, \dots, n.$$

From (2.6) we know that

$$\begin{aligned} f(x_1, \dots, x_n) &= \sum_{q=1}^{2n+1} g \left( \sum_{p=1}^n \lambda_p \psi_q(x_p) \right) \\ &= \sum_{q=1}^{2n+1} g \left( \sum_{p=1}^n \lambda_p \psi_q((s^{-1} \circ p^{-1})(\sigma_p)) \right) \\ &= \sum_{q=1}^{2n+1} g \left( \sum_{p=1}^n \lambda_p \varphi_q(\sigma_p) \right). \end{aligned}$$

This means that for any function  $F \in \mathcal{C}([0, 1]^n)$  we have the representation

$$F(\sigma_1, \dots, \sigma_n) = \sum_{q=1}^{2n+1} g \left( \sum_{p=1}^n \lambda_p \varphi_q(\sigma_p) \right)$$

and the functions  $\varphi_q$  belong to the class  $\text{Lip}(1)$ . This insight was first established by Fridman with different arguments and required a significant improvement of Kolmogorov's construction, [33, 68]. Anyhow, this possibility also follows automatically from Kolmogorov's original statement, [66].

### 2.3 Sprecher's constructive version

We now turn in detail to Sprecher's constructive version of the Kolmogorov superposition theorem from [108, 109] that has been briefly introduced in Section 2.1. Here, Sprecher explicitly defined an inner function  $\psi : \mathbb{R} \rightarrow \mathbb{R}$ , and formulated a construction algorithm for the outer functions  $\Phi_q : \mathbb{R} \rightarrow \mathbb{R}$ . Then, he gave a convergence proof for this algorithm to show the following result.

**Theorem 2.1.** *Let  $n \geq 2$ ,  $m \geq 2n$  and  $\gamma \geq m + 2$  be given integers and  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  as in Definition 2.2 below. We set*

$$a := \frac{1}{\gamma(\gamma - 1)}, \quad \beta(r) := \frac{n^r - 1}{n - 1},$$

and

$$\lambda_p := \begin{cases} 1, & p = 1, \\ \sum_{r=1}^{\infty} \gamma^{-(p-1)\beta(r)}, & p > 1. \end{cases}$$

Then, for any arbitrary continuous function  $f : [0, 1]^n \rightarrow \mathbb{R}$ , there exist  $m + 1$  continuous functions  $\Phi_q : \mathbb{R} \rightarrow \mathbb{R}$ ,  $q = 0, \dots, m$ , such that

$$f(\mathbf{x}) = \sum_{q=0}^m \Phi_q \circ \tilde{\Psi}_q(\mathbf{x}), \quad \text{with} \quad \tilde{\Psi}_q(\mathbf{x}) = \sum_{p=1}^n \lambda_p \psi(x_p + qa). \quad (2.9)$$

However, to make this proof work, two fundamental properties of  $\psi$ , namely continuity and monotonicity, are needed. Unfortunately, the inner function  $\psi$  in [111, 112] is neither continuous nor monotone. This observation is due to Köppen who first remarked in 2002, [73] that Sprecher's  $\psi$  actually has neither of these properties. Therefore, he suggested a correction of the inner function  $\psi$  in terms of a recursively defined function. His new  $\psi$  was defined point wise on the dense set of terminating rational numbers  $\mathcal{D} \subset \mathbb{R}$ . Anyhow, Köppen only claimed that his function exists, and that it is continuous and monotone increasing. He did not prove these properties but argued, without proving it, that Sprecher's constructive representation holds with the new  $\psi$ . In this section we will show that this is indeed the case. However, this will be done for a slightly changed version of Theorem 2.1 that employs only one outer function. For the proof of Theorem 2.1 with Köppen's inner function we refer to [12]. The remainder of this section is organized as follows. First, we will review Sprecher's definition of the inner function  $\psi$  and show in Section 2.3.1 that it is not continuous. Finally, in Section 2.3.2 we will introduce Köppen's correction of the function for which we will then show existence (Section 2.3.3), continuity (Section 2.3.4), and monotonicity (Section 2.3.5). A new version of Kolmogorov's theorem, which is a modification of Sprecher's result, will

be introduced in Section 2.3.6. In Section 2.3.7 we adapt Sprecher's algorithm to the new version and show convergence. This will constructively prove our new version of Kolmogorov's superposition theorem.

### 2.3.1 Discontinuity of the inner function

We start with repeating Sprecher's definition of  $\psi$  from [109]. Then, we show that it indeed does not define a continuous and monotone increasing function.

First, we have to define for a fixed basis  $\gamma > 1$  and  $k \in \mathbb{N}$  the set of terminating rational numbers

$$\mathcal{D}_k = \mathcal{D}_k(\gamma) := \left\{ d_k \in \mathbb{Q} : d_k = \sum_{r=1}^k i_r \gamma^{-r}, i_r \in \{0, \dots, \gamma - 1\} \right\}. \quad (2.10)$$

Then the set

$$\mathcal{D} := \bigcup_{k \in \mathbb{N}} \mathcal{D}_k \quad (2.11)$$

is dense in  $[0, 1]$ .

In [111], the inner function  $\psi$  was defined point wise on the set  $\mathcal{D}$  in the following way.

**Definition 2.2.** Let  $\langle i_1 \rangle := 0$  and for  $r \geq 2$  let

$$\langle i_r \rangle := \begin{cases} 0, & i_r = 0, 1, \dots, \gamma - 2, \\ 1, & i_r = \gamma - 1. \end{cases}$$

Furthermore, we define  $[i_1] := 0$  and, for  $r \geq 2$ ,

$$[i_r] := \begin{cases} 0, & i_r = 0, 1, \dots, \gamma - 3, \\ 1, & i_r = \gamma - 2, \gamma - 1, \end{cases}$$

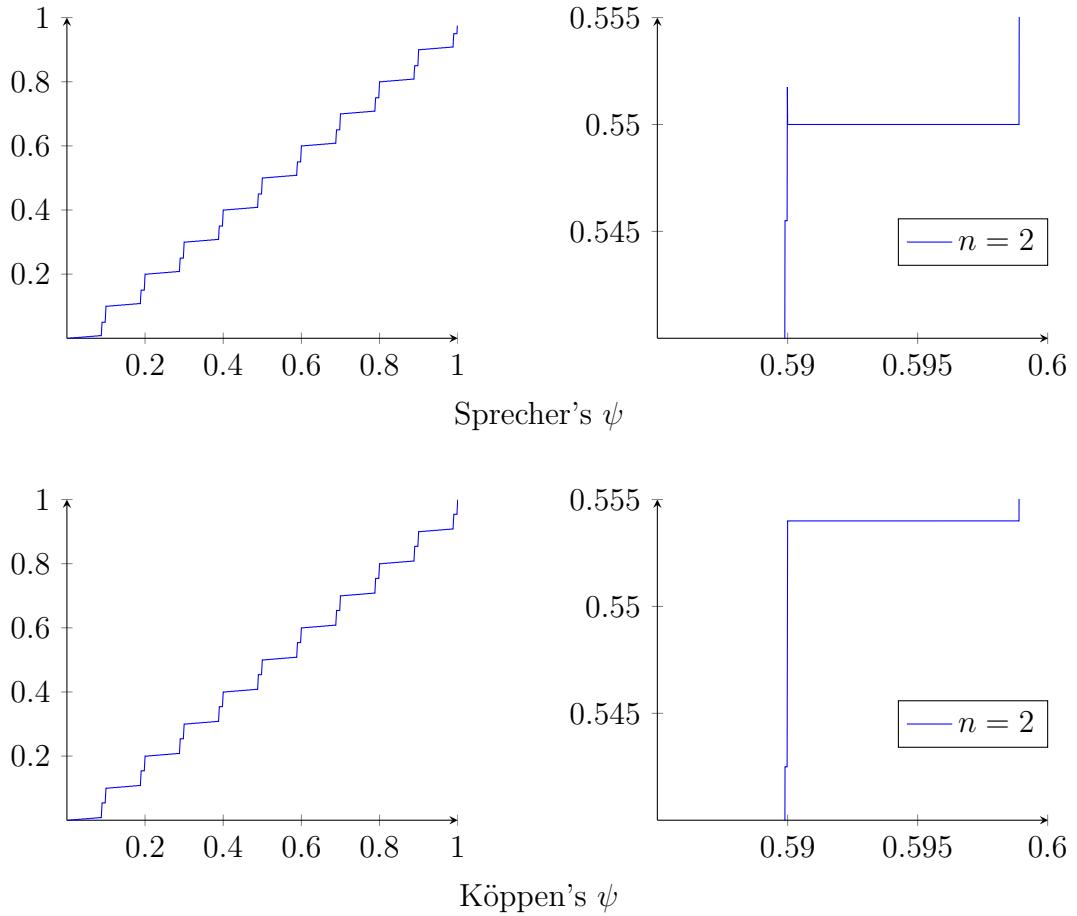
$$\tilde{i}_r := i_r - (\gamma - 2)\langle i_r \rangle,$$

and

$$m_r := \langle i_r \rangle \left( \sum_{s=1}^{r-1} ([i_s] \cdot \dots \cdot [i_{r-1}]) \right).$$

The function  $\psi$  is then defined on  $\mathcal{D}_k$  by

$$\psi(d_k) := \sum_{r=1}^k \tilde{i}_r 2^{-m_r} \gamma^{-\beta(r-m_r)}. \quad (2.12)$$



**Figure 2.2.** The top row shows the graph of Sprecher's  $\psi$  from (2.12) on the interval  $[0, 1]$  (left) and a zoom into a smaller interval (right), computed for the values of the set  $\mathcal{D}_5$ ,  $\gamma = 10$  and  $n = 2$ . One can clearly see the non-monotonicity and discontinuity near the value  $x = 0.59$ . The bottom row shows Köppen's version from (2.14) for the same parameters (left) and a zoom into the same region (right). Here, the discontinuity is no longer present.

Note that the definition of  $\psi$  depends on the dimension  $n$  since  $\beta(\cdot)$  depends on  $n$ . For a simpler notation we dispense with an additional index. The graph of the function  $\psi$  is depicted in Figure 2.2 for  $k = 5$ ,  $\gamma = 10$  and  $n = 2$ , i.e. it was calculated with the Definition 2.2 on the set of rational decimal numbers  $\mathcal{D}_k$ . Next, we assume that for any convergent sequence  $\{x_k\}_k \subset \mathcal{D}$ , with  $x = \lim_{k \rightarrow \infty} x_k$ , the limit

$$\psi(x) := \lim_{k \rightarrow \infty} \psi(d_k)$$

exists. In other words, we assume that there exists an extension of  $\psi$  to the real line,

which also will be denoted by  $\psi$ . Then, we show that this extension can neither be continuous nor monotone increasing.

The following calculation shows directly that the function is not continuous in contrast to the claim in [111]. With the choice  $\gamma = 10$  and  $n = 2$  one gets with Definition 2.2 the function values

$$\psi(0.58999) = 0.55175 \quad \text{and} \quad \psi(0.59) = 0.55.$$

This counter-example shows that the function  $\psi$  is not monotone increasing on  $\mathcal{D}$  and therefore its extension cannot be monotone increasing on the real line. We furthermore can see from the additive structure of  $\psi$  in (2.12) that

$$\psi(0.58999) < \psi(x_k) \quad \text{for all } x_k \in (0.58999, 0.59) \cap \mathcal{D}. \quad (2.13)$$

This implies that

$$\psi(0.58999) \leq \psi(x) \quad \text{for all } x \in (0.58999, 0.59).$$

Indeed, assuming for  $x \in (0.58999, 0.59)$  that  $\epsilon := \psi(0.58999) - \psi(x) > 0$ , we can find an integer  $k_0 \in \mathbb{N}$ , such that  $|\psi(x_k) - \psi(x)| < \epsilon$  and  $x_k \in (0.58999, 0.59)$  for all  $k > k_0$ . However, from this property it follows that  $\psi(x_k) < \psi(0.58999)$ , which is a conflict to (2.13). This shows that the function  $\psi$ , assumed that it exists, is also not continuous.

**Remark 2.3.** For  $\gamma = 10$ , discontinuities of  $\psi$  arise for all values

$$x = \frac{i}{10} + \frac{9}{100}, \quad i = 0, \dots, 9.$$

### 2.3.2 Correction of the inner function

Among other things, the convergence proof in [111, 112] is based on continuity and monotonicity of  $\psi$ . As the inner function defined by Sprecher does not provide these properties the convergence proof also becomes invalid unless the definition of  $\psi$  is properly modified. To this end, Köppen suggested in [73] a corrected version of the inner function and stated its continuity.

**Definition 2.4.** Let  $\gamma > 1$  be a fixed basis and  $\mathcal{D}$  be the set of terminating rational numbers as defined in (2.11). For  $k \in \mathbb{N}$  and  $d_k \in \mathcal{D}$  the function  $\psi$  is defined by  $\psi(d_k) := \psi_k(d_k)$ , where the latter value is recursively given by

$$\psi_k(d_k) = \begin{cases} d_k, & k = 1, \\ \psi_{k-1}(d_k - \frac{i_k}{\gamma^k}) + \frac{i_k}{\gamma^{\beta(k)}}, & k > 1, i_k < \gamma - 1, \\ \frac{1}{2} \left( \psi_{k-1}(d_k - \frac{i_k}{\gamma^k}) + \psi_{k-1}(d_k + \frac{1}{\gamma^k}) + \frac{i_k}{\gamma^{\beta(k)}} \right), & k > 1, i_k = \gamma - 1. \end{cases} \quad (2.14)$$

Note that the index of  $\psi_k(\cdot)$  indicates which value  $i_k$  is considered and is not redundant.

Köppen claimed that this recursion terminates. The following sections show that the so-defined  $\psi$  indeed can be extended to a function  $\psi : [0, 1] \rightarrow \mathbb{R}$ , and that this extension is continuous and monotone increasing.

For the calculations in the following sections it is advantageous to have an explicit representation of (2.14) as a sum. To this end, we need some further definitions. For simplicity, the values of  $\psi_{k-j}$  at the points  $d_{k-j}$  and  $d_{k-j} + \frac{1}{\gamma^{k-j}}$  are denoted as

$$\psi_{k-j} := \psi_{k-j}(d_{k-j}) \quad \text{and} \quad \psi_{k-j}^+ := \psi_{k-j}(d_{k-j} + \frac{1}{\gamma^{k-j}}).$$

Then, the recursion (2.14) takes for  $k - j > 1$  the form

$$\psi_{k-j} = \begin{cases} \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \psi_{k-j-1}, & i_{k-j} < \gamma - 1, \\ \frac{\gamma-2}{2\gamma^{\beta(k-j)}} + \frac{1}{2}\psi_{k-j-1} + \frac{1}{2}\psi_{k-j-1}^+, & i_{k-j} = \gamma - 1 \end{cases} \quad (2.15)$$

and

$$\psi_{k-j}^+ = \begin{cases} \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \frac{1}{\gamma^{\beta(k-j)}} + \psi_{k-j-1}, & i_{k-j} < \gamma - 2, \\ \frac{i_{k-j}}{2\gamma^{\beta(k-j)}} + \frac{1}{2}\psi_{k-j-1} + \frac{1}{2}\psi_{k-j-1}^+, & i_{k-j} = \gamma - 2, \\ \psi_{k-j-1}^+, & i_{k-j} = \gamma - 1. \end{cases} \quad (2.16)$$

Using the values

$$s_j := \begin{cases} 0, & i_{k-j+1} < \gamma - 2, \\ \frac{1}{2}, & i_{k-j+1} = \gamma - 2, \\ 1, & i_{k-j+1} = \gamma - 1 \end{cases} \quad (2.17)$$

and

$$\tilde{s}_j := \begin{cases} 0, & i_{k-j+1} < \gamma - 1, \\ \frac{1}{2}, & i_{k-j+1} = \gamma - 1, \end{cases} \quad (2.18)$$

we can define the matrix

$$\mathbf{M}_j := \begin{pmatrix} (1 - \tilde{s}_{j+1}) & \tilde{s}_{j+1} \\ (1 - s_{j+1}) & s_{j+1} \end{pmatrix}$$

and the vector

$$\mathbf{b}_j := \begin{pmatrix} (1 - 2\tilde{s}_{j+1}) \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1} \frac{\gamma-2}{\gamma^{\beta(k-j)}} \\ (1 - s_{j+1}) \left[ \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1 - 2s_{j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right] \end{pmatrix}.$$

With their help, the recursions (2.15) and (2.16) can be brought into the more compact form

$$\begin{pmatrix} \psi_{k-j} \\ \psi_{k-j}^+ \end{pmatrix} = \mathbf{M}_j \begin{pmatrix} \psi_{k-j-1} \\ \psi_{k-j-1}^+ \end{pmatrix} + \mathbf{b}_j. \quad (2.19)$$

The next definitions allow for a transformation of this recursive form into a direct representation of (2.19). To this end, we set  $\theta_0 := 1$ ,  $\theta_0^+ := 0$ ,  $\theta_1 := 1 - \tilde{s}_1$ ,  $\theta_1^+ := \tilde{s}_1$  and define for  $j = 1, \dots, k-1$

$$\begin{pmatrix} \theta_{j+1} \\ \theta_{j+1}^+ \end{pmatrix} = \mathbf{M}_j^T \begin{pmatrix} \theta_j \\ \theta_j^+ \end{pmatrix}. \quad (2.20)$$

By induction we can directly deduce from (2.20), (2.17), and (2.18) the useful properties

$$\theta_j + \theta_j^+ = 1 \quad \text{and} \quad \theta_j, \theta_j^+ > 0. \quad (2.21)$$

Now, with these definitions we can show the following lemmas that give us a direct point-wise representation of  $\psi$ .

**Lemma 2.5.** *The  $\xi$ -th step of the recursion (2.14) can be written as the sum*

$$\begin{aligned} \psi_k = & \sum_{j=0}^{\xi-1} \theta_j \left[ (1 - 2\tilde{s}_{j+1}) \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \\ & + \theta_\xi^+ \left[ (1 - s_{j+1}) \left( \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1 - 2s_{j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] + \theta_\xi \psi_{k-\xi} + \theta_\xi^+ \psi_{k-\xi}^+. \end{aligned} \quad (2.22)$$

**Proof.** [Proof by induction]  $\xi = 1$ : From (2.19) for  $j = 0$  we directly get

$$\psi_k = \left[ (1 - 2\tilde{s}_1) \frac{i_k}{\gamma^{\beta(k)}} + \tilde{s}_1 \frac{\gamma - 2}{\gamma^{\beta(k)}} \right] + (1 - \tilde{s}_1) \psi_{k-1} + \tilde{s}_1 \psi_{k-1}^+.$$

$\xi \rightarrow \xi + 1$ : We abbreviate  $\ominus(s) := (1 - s)$ . With (2.19) and (2.20) we have

$$\begin{aligned}
\psi_k &= \sum_{j=0}^{\xi-1} \theta_j \left[ \ominus(2\tilde{s}_{j+1}) \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \\
&\quad + \theta_j^+ \left[ \ominus(s_{j+1}) \left( \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \ominus(2s_{j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] \\
&\quad + \theta_\xi \left[ \ominus(\tilde{s}_{\xi+1}) \psi_{k-(\xi+1)} + \tilde{s}_{\xi+1} \psi_{k-(\xi+1)}^+ + \ominus(2\tilde{s}_{\xi+1}) \frac{i_{k-\xi}}{\gamma^{\beta(k-\xi)}} + \tilde{s}_{\xi+1} \frac{\gamma - 2}{\gamma^{\beta(k-\xi)}} \right] \\
&\quad + \theta_\xi^+ \left[ \ominus(s_{\xi+1}) \psi_{k-(\xi+1)} + s_{\xi+1} \psi_{k-(\xi+1)}^+ + \ominus(s_{\xi+1}) \left[ \frac{i_{k-\xi}}{\gamma^{\beta(k-\xi)}} + \ominus(2s_{\xi+1}) \frac{1}{\gamma^{\beta(k-\xi)}} \right] \right] \\
&= \sum_{j=0}^{\xi} \theta_j \left[ \ominus(2\tilde{s}_{j+1}) \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \\
&\quad + \theta_j^+ \left[ \ominus(s_{j+1}) \left( \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \ominus(2s_{j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] + \theta_{\xi+1} \psi_{k-(\xi+1)} + \theta_{\xi+1}^+ \psi_{k-(\xi+1)}^+.
\end{aligned}$$

□

**Lemma 2.6.** *The function  $\psi$  from Definition 2.4 has the representation*

$$\begin{aligned}
\psi(d_k) &= \sum_{j=0}^{k-2} \theta_j \left[ (1 - 2\tilde{s}_{j+1}) \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \\
&\quad + \theta_j^+ \left[ (1 - s_{j+1}) \left( \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1 - 2s_{j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] \\
&\quad + \theta_{k-1} \frac{i_1}{\gamma} + \theta_{k-1}^+ \frac{i_1 + 1}{\gamma}.
\end{aligned} \tag{2.23}$$

**Proof.** Choose  $\xi = k - 1$  in Lemma 2.5. □

Köppen assumed that there exists an extension from the dense set  $\mathcal{D}$  to the real line as in Sprecher's construction and that this extended  $\psi$  is monotone increasing and continuous but did not give a proof for it. In the following, we provide such a proof. The function  $\psi_k$  is depicted in Figure 2.2 for the same parameters  $k = 5$ ,  $\gamma = 10$ , and  $n = 2$  as before. A zoom into a smaller interval containing the value  $x = 0.59$  shows that the discontinuity from Sprecher's Definition 2.2 is no longer present.

### 2.3.3 Existence of Köppen's function

We first consider the existence of an extension of  $\psi$  to the real line and begin with the remark that for any basis  $\gamma > 1$ , every real number  $x \in [0, 1] \setminus \mathcal{D}$  has a representation

$$x = \sum_{r=1}^{\infty} \frac{i_r}{\gamma^r} = \lim_{k \rightarrow \infty} \sum_{r=1}^k \frac{i_r}{\gamma^r} = \lim_{k \rightarrow \infty} d_k.$$

For such a value  $x$ , we define the inner function

$$\psi(x) := \lim_{k \rightarrow \infty} \psi_k(d_k) = \lim_{k \rightarrow \infty} \psi_k \left( \sum_{r=1}^k \frac{i_r}{\gamma^r} \right) \quad (2.24)$$

and show the existence of this limit. To this end, we consider the behavior of the function values  $\psi_k$  and  $\psi_k^+$  as  $k$  tends to infinity. In the following, we restrict ourselves to the requirements  $n \geq 2$  and  $\gamma \geq 2n + 2$  from Theorem 2.1. For the following calculations it is useful to remind the definition

$$\beta(j) := \frac{n^j - 1}{n - 1}.$$

**Lemma 2.7.** *For growing values of  $k$  one has for  $\psi_k$  defined in (2.15) and  $\psi_k^+$  from (2.16)*

$$\psi_k^+ = \psi_k + \mathcal{O}(2^{-k}).$$

**Proof.** With (2.19), the fact that  $\gamma^{\beta(j)} = \gamma^{\beta(j-1)}\gamma^{n^{j-1}} > \gamma^{n^{j-2}}$  and  $\gamma^n > 2$ , we have

$$\begin{aligned}
|\psi_k^+ - \psi_k| &\leq |s_1 - \tilde{s}_1| |\psi_{k-1}^+ - \psi_{k-1}| + \frac{\gamma - 2}{\gamma^{\beta(k)}} |\tilde{s}_1 - s_1 + (1 - s_1)(1 - 2s_1)| \\
&\leq \frac{1}{2} |\psi_{k-1}^+ - \psi_{k-1}| + \frac{\gamma - 2}{\gamma^{\beta(k)}} \\
&\leq \frac{1}{2} \left| \frac{1}{2} |\psi_{k-2}^+ - \psi_{k-2}| + \frac{\gamma - 2}{\gamma^{\beta(k-1)}} \right| + \frac{\gamma - 2}{\gamma^{\beta(k)}} \leq \dots \\
&\leq \left( \frac{1}{2} \right)^{k-1} |\psi_1^+ - \psi_1| + \left[ \left( \frac{1}{2} \right)^{k-2} \frac{1}{\gamma^{\beta(2)}} + \dots + \left( \frac{1}{2} \right)^{k-k} \frac{1}{\gamma^{\beta(k)}} \right] \\
&= \left( \frac{1}{2} \right)^{k-1} |\psi_1^+ - \psi_1| + \left( \frac{1}{2} \right)^{k-2} (\gamma - 2) \left( \sum_{j=2}^k \frac{2^{j-2}}{\gamma^{\beta(j)}} \right) \\
&\leq \left( \frac{1}{2} \right)^{k-1} |\psi_1^+ - \psi_1| + \left( \frac{1}{2} \right)^{k-2} (\gamma - 2) \left( \sum_{j=0}^{\infty} \left( \frac{2}{\gamma^n} \right)^j \right) \\
&= \left( \frac{1}{2} \right)^{k-1} |\psi_1^+ - \psi_1| + \left( \frac{1}{2} \right)^{k-2} (\gamma - 2) \frac{\gamma^n}{\gamma^n - 2} \\
&= \left( \frac{1}{2} \right)^{k-2} \left[ \frac{1}{2\gamma} + \frac{(\gamma - 2)\gamma^n}{\gamma^n - 2} \right],
\end{aligned}$$

and the assertion is proved. Here, the second estimate holds since  $s_j - \tilde{s}_j \in \{0, \frac{1}{2}\}$  and  $\tilde{s}_j - s_j + (1 - s_j)(1 - 2s_j) \in \{1, \frac{1}{2}, -\frac{1}{2}\}$ .  $\square$

Now, the fact that the values  $\psi_k$  and  $\psi_k^+$  approach by the rate  $\mathcal{O}(2^{-k})$  for increasing values of  $k$  enables us to show the following lemma:

**Lemma 2.8.** Let  $x \in [0, 1] \setminus \mathcal{D}$  be an arbitrary value and  $\{d_k\}_k \subset \mathcal{D}$  a convergent sequence with limit  $x = \lim_{k \rightarrow \infty} d_k$ . Then, the sequence  $\{\psi_k\}_k \subset \mathbb{R}$ , with values  $\psi_k := \psi_k(d_k)$  defined by (2.14), is a Cauchy sequence.

**Proof.** For  $k, k' \in \mathbb{N}$ , where without loss of generality  $k > k'$ , we set  $\xi := k - k'$  and apply (2.22) to represent  $\psi_k$ . Then, since  $(1 - 2\tilde{s}_{j+1})i_{k-j} + \tilde{s}_{j+1}(\gamma - 2) < 2(\gamma - 2)$ , and  $(1 - s_{j+1})(i_{k-j} + (1 - 2s_{j+1})) < 2(\gamma - 2)$ , we observe that each term in the sum (2.22) can be bounded from above by  $2(\gamma - 2)\gamma^{-\beta(j)}$ . Together with the fact that

$\gamma^{\beta(j)} > \gamma^{n^{j-1}}$ , Lemma 2.7, and (2.21) this leads to the following estimate:

$$\begin{aligned}
|\psi_k - \psi_{k'}| &\leq |\theta_{k-k'}\psi_{k'} + \theta_{k-k'}^+\psi_{k'}^+ - \psi_{k'}| + 2(\gamma - 2) \sum_{j=k'+1}^k \frac{1}{\gamma^{\beta(j)}} \\
&\leq |\theta_{k-k'}\psi_{k'} + \theta_{k-k'}^+\psi_{k'}^+ - \psi_{k'}| + 2(\gamma - 2) \sum_{j=k'+1}^k \left(\frac{1}{\gamma^n}\right)^{j-1} \\
&= \left| \theta_{k-k'}\psi_{k'} + \theta_{k-k'}^+ \left( \psi_{k'} + \mathcal{O}(2^{-k'}) \right) - \psi_{k'} \right| \\
&\quad + \frac{2\gamma^n(\gamma - 2)}{1 - \gamma^n} \left( \left(\frac{1}{\gamma^n}\right)^k - \left(\frac{1}{\gamma^n}\right)^{k'} \right) \\
&= \mathcal{O}(2^{-k'}) + \frac{2\gamma^n(\gamma - 2)}{1 - \gamma^n} \left( \left(\frac{1}{\gamma^n}\right)^k - \left(\frac{1}{\gamma^n}\right)^{k'} \right).
\end{aligned}$$

The right hand side tends to 0 when  $k, k' \rightarrow \infty$ .  $\square$

We have shown that  $\{\psi_k\}_k$  is a Cauchy sequence in  $\mathcal{D}$ . Since we know that the real numbers are complete, we can infer the existence of a limit in  $\mathbb{R}$ . Thus the function  $\psi$  from (2.24) is well defined what provides us with a function value for all  $x \in [0, 1]$ . It remains to show that this  $\psi$  is continuous and monotone increasing. This will be the topic of the following subsections.

### 2.3.4 Continuity of $\psi$

We now show the continuity of the inner function  $\psi$ . To this end we first recall some properties of the representations of real numbers. Let

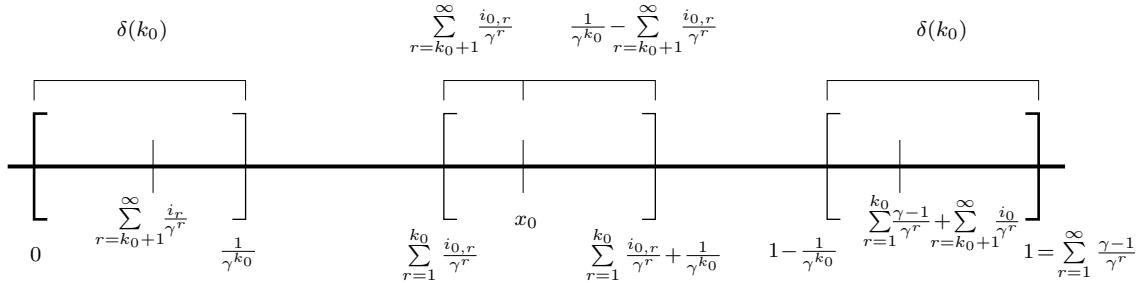
$$x := \sum_{r=1}^{\infty} \frac{i_r}{\gamma^r} \quad \text{and} \quad x_0 := \sum_{r=1}^{\infty} \frac{i_{0,r}}{\gamma^r}$$

be the representation of the values  $x$  and  $x_0$  in the basis  $\gamma$ , respectively.

**Remark 2.9.** For simplicity, this notation also includes values  $x, x_0 \in \mathcal{D}$  with finite representations: If  $x \in \mathcal{D}_k \subset \mathcal{D}$  for some finite  $k$ , we set  $i_r = 0$  for  $r > k$ . Finite representations are always preferred over infinite ones. For example, with  $\gamma = 10$  we take  $x = 0.1$  instead of  $x = 0.\bar{0}\bar{9}$ .

Now, let  $x_0 \in (0, 1)$ ,  $k_0 \in \mathbb{N}$  be given and

$$\delta(k_0) := \min \left\{ \sum_{r=k_0+1}^{\infty} \frac{i_{0,r}}{\gamma^r}, \frac{1}{\gamma^{k_0}} - \sum_{r=k_0+1}^{\infty} \frac{i_{0,r}}{\gamma^r} \right\}. \quad (2.25)$$



**Figure 2.3.** The figure shows the interval  $[0, 1]$ . For any two values  $x_1$  and  $x_2$  that both lie in one of the depicted small intervals it holds that  $i_{1,r} = i_{2,r}$  for  $r = 1, \dots, k_0$ . The three intervals represent the possible cases that occur in the proof of Theorem 2.10.

For any  $x \in (x_0 - \delta(k_0), x_0 + \delta(k_0))$  it follows that

$$i_r = i_{0,r}, \quad r = 1, \dots, k_0. \quad (2.26)$$

Special attention has to be paid to the values  $x_0 = 0$  and  $x_0 = 1$ . In both cases, we can choose  $\delta(k_0) = \gamma^{-k_0}$ . Then (2.26) holds for all  $x \in [0, \delta(k_0))$  if  $x_0 = 0$  and all  $x \in (1 - \delta(k_0), 1]$  if  $x_0 = 1$ . The three different cases are depicted in Figure 2.3. Altogether we thus can find for any given arbitrary  $x_0 \in [0, 1]$  and  $k_0 \in \mathbb{N}$  a  $\delta$ -neighborhood

$$U := (x_0 - \delta(k_0), x_0 + \delta(k_0)) \cap [0, 1] \quad (2.27)$$

in which (2.26) holds. With this observation we are able to prove the following theorem:

**Theorem 2.10.** *The function  $\psi : [0, 1] \rightarrow \mathbb{R}$ , whose values on  $\mathcal{D}$  are given by Definition 2.4 and which is extended to  $\mathbb{R}$  by (2.24), is continuous on  $[0, 1]$ .*

**Proof.** We show that  $\psi$  is continuous in  $x_0 \in [0, 1]$ . To this end, let  $\varepsilon > 0$  be given and  $k_0 \in \mathbb{N}$  such that

$$\left( \frac{1}{\gamma^n} \right)^{k_0} < \frac{\varepsilon (1 - \gamma^n)}{4\gamma^n(\gamma - 2)}.$$

Furthermore, let  $\delta(k_0) > 0$  be defined by (2.25) if  $x_0 \in (0, 1)$ , and  $\delta(k_0) := \gamma^{-k_0}$  else.

With (2.23) and (2.27) we see for  $x, x_0 \in U$ :

$$\begin{aligned}
|\psi(x) - \psi(x_0)| &= \lim_{k \rightarrow \infty} |\psi(d_k) - \psi(d_{0,k})| \\
&= \lim_{k \rightarrow \infty} \left| \sum_{j=0}^{k-k_0-1} \theta_j \left[ (1 - 2\tilde{s}_{j+1}) \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \right. \\
&\quad + \theta_j^+ \left[ (1 - s_{j+1}) \left( \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1 - 2s_{j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] \\
&\quad - \sum_{j=0}^{k-k_0-1} \theta_{0,j} \left[ (1 - 2\tilde{s}_{0,j+1}) \frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{0,j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \\
&\quad \left. + \theta_{0,j}^+ \left[ (1 - s_{0,j+1}) \left( \frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + (1 - 2s_{0,j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] \right| \\
&\leq \lim_{k \rightarrow \infty} \sum_{j=0}^{k-k_0-1} \left| \theta_j \left[ (1 - 2\tilde{s}_{j+1}) \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \right| \\
&\quad + \left| \theta_j^+ \left[ (1 - s_{j+1}) \left( \frac{i_{k-j}}{\gamma^{\beta(k-j)}} + (1 - 2s_{j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] \right| \\
&\quad + \lim_{k \rightarrow \infty} \sum_{j=0}^{k-k_0-1} \left| \theta_{0,j} \left[ (1 - 2\tilde{s}_{0,j+1}) \frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + \tilde{s}_{0,j+1} \frac{\gamma - 2}{\gamma^{\beta(k-j)}} \right] \right| \\
&\quad + \left| \theta_{0,j}^+ \left[ (1 - s_{0,j+1}) \left( \frac{i_{0,k-j}}{\gamma^{\beta(k-j)}} + (1 - 2s_{0,j+1}) \frac{1}{\gamma^{\beta(k-j)}} \right) \right] \right| \\
&\leq \lim_{k \rightarrow \infty} \frac{4\gamma^n(\gamma - 2)}{1 - \gamma^n} \left| \left( \frac{1}{\gamma^n} \right)^k - \left( \frac{1}{\gamma^n} \right)^{k_0} \right| = \frac{4\gamma^n(\gamma - 2)}{1 - \gamma^n} \left( \frac{1}{\gamma^n} \right)^{k_0} < \varepsilon.
\end{aligned} \tag{2.28}$$

Note that the estimation of the last two sums was derived similar to that in the proof of Lemma 2.8. In conclusion we found for given  $\varepsilon > 0$  a  $\delta(k_0) > 0$  such that  $|\psi(x) - \psi(x_0)| < \varepsilon$  whenever  $|x - x_0| < \delta(k_0)$ . This is just the definition of continuity of  $\psi$  in  $x_0 \in (0, 1)$ . Since the interval  $U$  is only open to the right if  $x_0 = 0$  and open to the left if  $x_0 = 1$ , the inequality (2.28) also shows for these two cases continuity from the right and from the left, respectively.  $\square$

### 2.3.5 Monotonicity of $\psi$

A further crucial property of the function  $\psi$  is its monotonicity. We show this first on the dense subset  $\mathcal{D} \subset \mathbb{R}$  of terminating rational numbers.

**Lemma 2.11.** *For every  $k \in \mathbb{N}$ , and values  $\psi_k$  defined by (2.15) and  $\psi_k^+$  from*

(2.16) there holds

$$\psi_k^+ \geq \psi_k + \frac{1}{\gamma^{\beta(k)}}.$$

**Proof.** [Proof by induction]  $k = 1$ :

$$\psi_1^+ - \psi_1 = \psi_1(d_1 + \frac{1}{\gamma}) - \psi_1(d_1) = d_1 + \frac{1}{\gamma} - d_1 = \frac{1}{\gamma} = \frac{1}{\gamma^{\beta(1)}}$$

$k \rightarrow k + 1$ :

$$\begin{aligned} \psi_{k+1}^+ - \psi_{k+1} &= (s_0 - \tilde{s}_0)(\psi_k^+ - \psi_k) \\ &\quad + \frac{1}{\gamma^{\beta(k+1)}}((2\tilde{s}_0 - s_0)i_{k+1} + (1 - s_0)(1 - 2s_0) - \tilde{s}_0(\gamma - 2)) \\ &= \begin{cases} \frac{1}{\gamma^{\beta(k+1)}}, & i_{k+1} < \gamma - 2 \quad (s_0 = \tilde{s}_0 = 0), \\ \frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{i_{k+1}}{\gamma^{k+1}}, & i_{k+1} = \gamma - 2 \quad (s_0 = \frac{1}{2}, \tilde{s}_0 = 0), \\ \frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{\gamma - 2}{\gamma^{\beta(k+1)}}, & i_{k+1} = \gamma - 1 \quad (s_0 = 1, \tilde{s}_0 = \frac{1}{2}). \end{cases} \end{aligned}$$

For the first case  $i_{k+1} < \gamma - 2$ , the assertion is trivial. For the other two cases, we have

$$\begin{aligned} \frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{i_{k+1}}{\gamma^{k+1}} &\geq \frac{1}{2}(\psi_k^+ - \psi_k) - \frac{1}{2}\frac{\gamma - 2}{\gamma^{\beta(k+1)}} \\ &\geq \frac{1}{2} \left( \frac{1}{\gamma^{\beta(k)}} - \frac{\gamma - 2}{\gamma^{\beta(k+1)}} \right) \geq \frac{1}{\gamma^{\beta(k+1)}}. \end{aligned}$$

Here, the validity of the last estimate can be obtained from

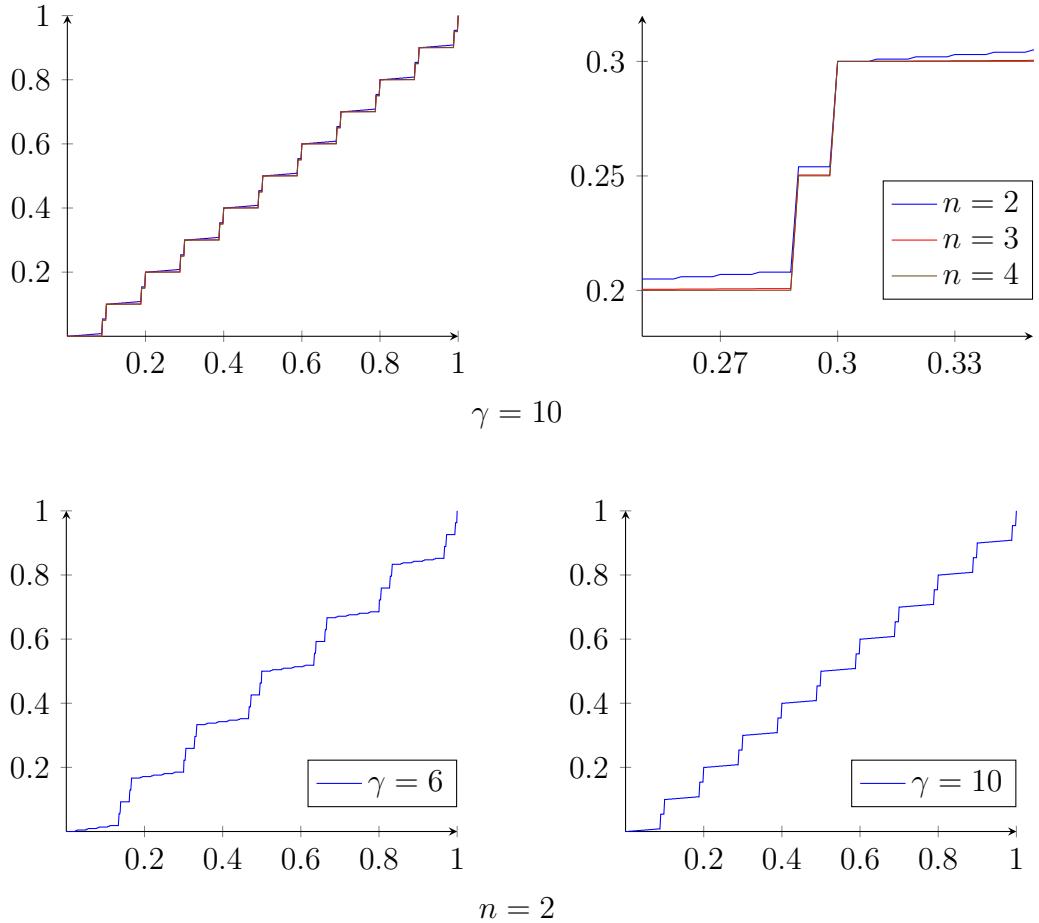
$$\begin{aligned} \frac{1}{2} \left( \frac{1}{\gamma^{\beta(k)}} - \frac{\gamma - 2}{\gamma^{\beta(k+1)}} \right) \geq \frac{1}{\gamma^{\beta(k+1)}} &\Leftrightarrow \frac{1}{2} \left( \frac{\gamma^{n^{k-1}}}{\gamma^{\beta(k+1)}} - \frac{\gamma - 2}{\gamma^{\beta(k+1)}} \right) \geq \frac{1}{\gamma^{\beta(k+1)}} \\ \Leftrightarrow \gamma^{n^{k-1}} - \gamma + 2 \geq 2 &\Leftrightarrow \gamma^{n^{k-1}} \geq \gamma \Leftrightarrow n^{k-1} \geq 1. \end{aligned}$$

This shows the assertion.  $\square$

We can now easily prove the following theorem:

**Theorem 2.12.** *The function  $\psi : [0, 1] \rightarrow \mathbb{R}$ , whose values on  $\mathcal{D}$  are given by Definition 2.4 and which is extended to  $\mathbb{R}$  by (2.24), is monotone increasing on  $[0, 1]$ .*

**Proof.** By Lemma 2.11 we know that  $\psi$  is strictly monotone increasing on a dense subset of  $[0, 1]$ . Since the function is continuous, this holds for the whole interval  $[0, 1]$ .  $\square$



**Figure 2.4.** The top row shows the graph of  $\psi$  for  $n = 2, 3, 4$  and constant basis  $\gamma = 10$ . The top right picture is a zoom into its left neighbour to show the slight difference of the functions. The bottom row shows plots for  $n = 2$  and  $\gamma = 6$  (left) and  $\gamma = 10$  (right).

In summary, we have demonstrated that the inner function  $\psi$  defined by Sprecher (c.f. [111, 112]) is neither continuous nor monotone increasing, whereas the definition (2.24) of  $\psi$  for Köppen's recursion (2.14) from [73] possesses these properties.

We conclude these sections on the inner function with some remarks on  $\psi$  and its dependency on the dimension  $n$  and the choice of the basis  $\gamma$ . In Figure 2.4 we show the graphs for different choices of these parameters. First, the function is plotted for different values of  $n$ . For  $n = 2, 3, 4$  and fixed  $\gamma = 10$ , the graphs of  $\psi$  are very similar and can only be distinguished if we zoom into a detailed view. This is shown in the top row of Figure 2.4. This changes for variable values of  $\gamma$  as it is shown in the bottom row. For fixed  $n = 2$ , the non-smoothness of  $\psi$

becomes stronger for larger  $\gamma$ . Therefore, the smallest possible choice for  $\gamma$  seems to be advantageous. However, the effect cannot be avoided for increasing dimension  $n$  due to the restriction  $2n+2 \leq \gamma$ . Altogether we can say that the non-smoothness of the inner function  $\psi$  becomes stronger for increasing dimension  $n$ . Now, comparing Köppen's definition with the result from [110], where a universal inner function was defined independently from  $n$ , we observe that both functions were constructed in a similar way. This leads to the conclusion that, loosely speaking, the  $n$ -independent function  $\psi$  from [110] incorporates more non-smoothness than it is necessary to represent  $n$ -dimensional functions  $f$ .

### 2.3.6 Definitions and reformulation with a single outer function

We are now equipped with a continuous, monotone increasing function  $\psi : [0, 1] \rightarrow \mathbb{R}$  for which Köppen claimed in [73], without proving it, that Sprecher's constructive Theorem 2.1 holds. Following the lines of Sprecher, the proof was finally given in [12]. Here, we dispense with a demonstration of this result and rather introduce a new constructive version of Kolmogorov's theorem with only one outer function. Using a remark by Lorentz, [108], who stated that the  $q$ -dependence of the outer functions can always be eliminated by a shift of their argument, the new theorem will be derived from Theorem 2.1 by slight changes in Sprecher's definitions and arguments. Accordingly, the algorithm from [112] will be adapted to prove the result.

First, we repeat some definitions and introduce new parameters. As before,  $n \in \mathbb{N}$  always denotes the dimension while  $\gamma \in \mathbb{N}$  is the basis in (2.10). We use Köppen's function  $\psi : [0, 1] \rightarrow \mathbb{R}$ , given by Definition 2.4 and (2.24) and define the function  $\psi_e : \mathbb{R} \rightarrow \mathbb{R}$  by

$$\psi_e(x) := i + \psi(x - i), \quad x \in [i, i + 1), \quad i \in \mathbb{Z}. \quad (2.29)$$

Due to the previous results and  $\psi(0) = 0$ ,  $\psi(1) = 1$ , this function is continuous and monotone increasing. Since  $\psi \equiv \psi_e$  on the interval  $[0, 1]$  we can drop the subscript of  $\psi_e$  in the following.

We keep the values

$$a := \frac{1}{\gamma(\gamma - 1)}, \quad \beta(r) := \frac{n^r - 1}{n - 1},$$

and

$$\lambda_p := \begin{cases} 1, & p = 1, \\ \sum_{r=1}^{\infty} \gamma^{-(p-1)\beta(r)}, & p > 1, \end{cases}$$

as in Theorem 2.1. Then for  $k \in \mathbb{N}$ , the  $k$ -th remainder of  $\lambda_p$  is denoted by

$$\varepsilon_{k,p} := \sum_{r=k+1}^{\infty} \gamma^{-(p-1)\beta(r)}. \quad (2.30)$$

From these values that stem from Sprecher's constructions in [111, 112] we now derive new parameters for our representation.

**Definition 2.13.** Let  $\tilde{\Gamma}_k \in \mathbb{R}$  be chosen such that

$$\tilde{\Gamma}_k < \frac{1}{2} \left( \gamma^{-n\beta(k)} - (\gamma - 2) \varepsilon_{k,2} \left( \sum_{p=1}^n \lambda_p \right) \right). \quad (2.31)$$

For  $\delta > 0$ , let the value  $\Lambda_\delta \in \mathbb{R}$  be given by

$$\Lambda_\delta^{-1} \left[ \left( \sum_{p=1}^n \lambda_p \right) (1 + (\gamma - 2)\varepsilon_{k,2}) + 2\tilde{\Gamma}_k \right] = 1 - \delta. \quad (2.32)$$

Since  $\sum_{p=1}^n \lambda_p > 1$  and  $1 + (\gamma - 2)\varepsilon_{k,2} > 1$ , this implies  $\Lambda_\delta > 1$ . With these values we now define for  $p = 1, \dots, n$ , the new coefficients

$$\alpha_p := \Lambda_\delta^{-1} \lambda_p, \quad (2.33)$$

and for  $q = 0, \dots, m$ , the shifts

$$\Delta_q := -\Lambda_\delta^{-1} \left[ \left( \sum_{p=1}^n \lambda_p \right) \psi \left( q \sum_{r=2}^k \gamma^{-r} \right) - \tilde{\Gamma}_k \right] + q. \quad (2.34)$$

For subsequent calculations we also define

$$b_k := \varepsilon_{k,2} \left( \sum_{p=1}^n \alpha_p \right) \quad \text{and} \quad \Gamma_k := \Lambda_\delta^{-1} \tilde{\Gamma}_k. \quad (2.35)$$

**Theorem 2.14.** Let  $n \geq 2$ ,  $m \geq 2n$ , and  $\gamma \geq m+2$  be given integers and  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  be given by (2.29). Additionally, define  $\alpha_p \in \mathbb{R}$ ,  $p = 1, \dots, n$  by (2.33) and  $\Delta_q \in \mathbb{R}$ ,  $q = 0, \dots, m$  by (2.34). Then, for any arbitrary continuous function  $f : [0, 1]^n \rightarrow \mathbb{R}$ , there exists a continuous functions  $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ , such that

$$f(\mathbf{x}) = \sum_{q=0}^m \Phi \circ \Psi_q(\mathbf{x}), \quad \text{with} \quad \Psi_q(\mathbf{x}) = \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q. \quad (2.36)$$

Furthermore,  $\Phi$  can be computed by Algorithm 2.1 which is defined below.

The relation between this theorem and Theorem 2.1 is very simple. Originally, for each  $q = 0, \dots, m$  the inner sum  $\tilde{\Psi}_q$  in (2.9) maps the  $n$ -dimensional unit cube into an interval on the real line on which  $f$  is approximated by an outer function  $\Phi_q$ ,



**Figure 2.5.** For varying values  $q \in \{0, \dots, m\}$  the images of the  $n$ -dimensional unit cube under the mappings  $\tilde{\Psi}_q$  from (2.9) overlap on the real line, while the image-intervals of the functions  $\Psi_q$ , defined in (2.36), are disjoint. The latter allows for the construction of a single outer function  $\Phi$  to represent  $f$ .

see (2.36). Due to an overlap of the image-intervals for different  $q$ 's the functions  $\Phi_0, \dots, \Phi_m$  have to be constructed separately. This can be avoided by the additional shift  $\Delta_q$  in the inner sums  $\Psi_q$ . Practically, this shift places the image-intervals for  $q = 0, \dots, m$  next to each other which allows for the construction of a single outer function  $\Phi$ . Figure 2.5 illustrates this idea that is due to Lorentz, see [108].

The proof of Theorem 2.14 will be the topic of Section 2.3.7.

### 2.3.7 Modification of Sprecher's algorithm

We will now present an adaption of Sprecher's constructive algorithm from [112] which computes the outer function  $\Phi$  in (2.36) and show that it is convergent. This will prove Theorem 2.14.

To this end, let  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  be an arbitrary continuous monotone increasing function with  $\sigma(x) \equiv 0$  when  $x \leq 0$  and  $\sigma(x) \equiv 1$  when  $x \geq 1$ . For given  $k \in \mathbb{N}$ ,  $q \in \{0, \dots, m\}$  and  $\mathbf{d}_k := (d_{k,1}, \dots, d_{k,n}) \in \mathcal{D}_k^n$ , we define the values

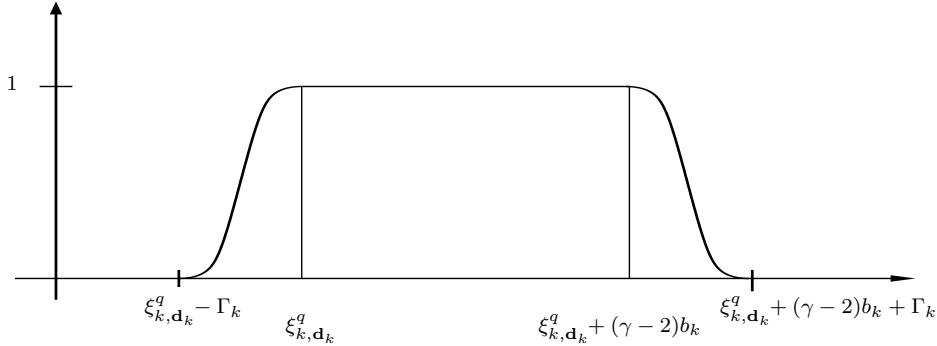
$$\xi_{k,\mathbf{d}_k}^q := \sum_{p=1}^n \alpha_p \psi \left( d_{k,p} + q \sum_{r=2}^k \gamma^{-r} \right) + \Delta_q \in \mathbb{R}. \quad (2.37)$$

This gives  $(m+1)(\gamma^k + 1)^n$  real values  $\xi_{k,\mathbf{d}_k}^q$  for which we next define functions  $\omega_{k,\mathbf{d}_k}^q : \mathbb{R} \rightarrow \mathbb{R}$  by

$$\omega_{k,\mathbf{d}_k}^q(y) := \sigma \left( \Gamma_k^{-1} \left( y - \xi_{k,\mathbf{d}_k}^q \right) + 1 \right) - \sigma \left( \Gamma_k^{-1} \left( y - \xi_{k,\mathbf{d}_k}^q - (\gamma - 2) b_k \right) \right). \quad (2.38)$$

As depicted in Figure 2.6, the function  $\omega_{k,\mathbf{d}_k}^q$  is constant on the interval  $[\xi_{k,\mathbf{d}_k}^q, \xi_{k,\mathbf{d}_k}^q + (\gamma - 2)b_k]$  and has compact support.

We are now in the position to present the algorithm which implements the representation of an arbitrary multivariate function  $f$  as superposition of single variable functions. Let  $\|\cdot\|$  denote the usual maximum norm of functions and let  $f : [0, 1]^n \rightarrow \mathbb{R}$  be a given continuous function with known maximum norm  $\|f\|$ .



**Figure 2.6.** The compactly supported function  $\omega_{k,d_k}^q : \mathbb{R} \rightarrow \mathbb{R}$ .

Furthermore, let  $\varepsilon > 0$  be a fixed real number such that

$$0 < \frac{m-n+1}{m+1}\varepsilon + \frac{2n}{m+1} =: \eta < 1 \quad (2.39)$$

which can always be achieved by choosing

$$0 < \varepsilon < 1 - \frac{n}{m-n+1} < 1. \quad (2.40)$$

The outer function in Theorem 2.14 can now be computed by the following algorithm:

**ALGORITHM 2.1** (to compute the outer function  $\Phi$ ). Starting with  $f_0 \equiv f$ , for  $r = 1, 2, 3, \dots$ , iterate the following steps:

- I. Given the function  $f_{r-1}(\mathbf{x})$ , determine an integer  $k_r$  such that for any two points  $\mathbf{x}, \mathbf{x}' \in [0, 1]^n$  with  $\|\mathbf{x} - \mathbf{x}'\| \leq \gamma^{-k_r}$  it holds that  $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \varepsilon \|f_{r-1}\|$ . This determines rational coordinate points  $\mathbf{d}_{k_r} = (d_{k_r,1}, \dots, d_{k_r,n}) \in \mathcal{D}_{k_r}^n$ .
- II. For all  $\mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n$  and  $q = 0, 1, \dots, m$ :
  - II-1 Compute the values  $\xi_{k,d_{k_r}}^q$  given by (2.37).
  - II-2 Compute the functions  $\omega_{k,d_{k_r}}^q(y)$  given by (2.38).
- III. III-1 Compute the function

$$\phi^r(y) = \frac{1}{m+1} \sum_{q=0}^m \sum_{\mathbf{d}_{k_r}} f_{r-1}(\mathbf{d}_{k_r}) \omega_{k,d_{k_r}}^q(y), \quad (2.41)$$

where for each  $q$  the sum  $\sum_{\mathbf{d}_{k_r}}$  is taken over all values  $\mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n$ .

III–2 Substitute the transfer functions  $\Psi_q(\mathbf{x})$  and compute

$$\phi^r \circ \Psi_q(\mathbf{x}) := \frac{1}{m+1} \sum_{q=0}^m \sum_{\mathbf{d}_{k_r}} f_{r-1}(\mathbf{d}_{k_r}) \omega_{k_r}^q (\Psi_q(\mathbf{x})).$$

Again, the second sum  $\sum_{\mathbf{d}_{k_r}}$  is built for each  $q$  over all values  $\mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n$ .

III–3 Compute the function

$$f_r(\mathbf{x}) := f(\mathbf{x}) - \sum_{q=0}^m \sum_{j=1}^r \phi^j \circ \Psi_q(\mathbf{x}). \quad (2.42)$$

This completes the  $r$ -th iteration loop and gives the  $r$ -th approximation to  $f$ . Now replace  $r$  by  $r + 1$  and go to step I.

The convergence of the series  $\{f_r\}_r$  for  $r \rightarrow \infty$  to the limit  $\lim_{r \rightarrow \infty} f_r =: g \equiv 0$  is equivalent to the validity of Theorem 2.14. To show this, the following convergence proof adapts the arguments from [12] which proved Theorem 2.1. Both proofs essentially follow [111,112] but differ in the arguments that refer to the inner function  $\psi$  which is now replaced by Köppen's definition.

The main argument for convergence is the validity of the following theorem:

**Theorem 2.15.** *For the approximations  $f_r$ ,  $r = 1, 2, \dots$  defined in step III–3 of Algorithm 2.1 there holds the estimate*

$$\|f_r\| = \left\| f_{r-1}(\mathbf{x}) - \sum_{q=0}^m \phi^r \circ \Psi_q(\mathbf{x}) \right\| \leq \eta \|f_{r-1}\|.$$

To prove this theorem, some preliminary work is necessary. To this end, the following lemmas are taken from the proof of Theorem 2.1 in [12]. The arguments to show Theorem 2.15 are then derived from these results. One key to the numerical implementation of Algorithm 2.1 is the minimum distance of images of rational grid points  $\mathbf{d}_k$  under the mapping  $\Psi_q$  from (2.9) for  $q = 0$ . This distance can be bounded from below. The estimate can be derived from the following lemma.

**Lemma 2.16.** *For each integer  $k \in \mathbb{N}$ , set*

$$\mu(\mathbf{d}_k, \mathbf{d}'_k) := \sum_{p=1}^n \lambda_p [\psi(d_{k,p}) - \psi(d'_{k,p})],$$

where  $d_{k,p}, d'_{k,p} \in \mathcal{D}_k$ . Then

$$\min |\mu(\mathbf{d}_k, \mathbf{d}'_k)| \geq \gamma^{-n\beta(k)},$$

where the minimum is taken over all pairs  $\mathbf{d}_k \neq \mathbf{d}'_k \in \mathcal{D}_k^n$ , i.e.

$$\sum_{p=1}^n |d_{k,p} - d'_{k,p}| \neq 0.$$

**Proof.** Since for each  $k$  the set  $\mathcal{D}_k$  is finite, a unique minimum exists. For each  $k \in \mathbb{N}$ , let  $d_{k,p}, d'_{k,p} \in \mathcal{D}_k$  and  $A(d_{k,p}, d'_{k,p}) := \psi(d_{k,p}) - \psi(d'_{k,p})$  for  $p = 1, \dots, n$ . Since  $\psi$  is monotone increasing, we know that  $A(d_{k,p}, d'_{k,p}) \neq 0$  for all admissible values of  $p$ . Now from Lemma 2.11 it follows directly that

$$\min_{\mathcal{D}_k} |A(d_{k,p}, d'_{k,p})| = \gamma^{-\beta(k)}, \quad (2.43)$$

where for each fixed  $p, k$  the minimum is taken over the decimals for which  $|d_{k,p} - d'_{k,p}| \neq 0$ . The upper bound

$$\min |\mu(\mathbf{d}_k, \mathbf{d}'_k)| \leq \lambda_n \gamma^{-\beta(k)} \quad (2.44)$$

can be gained from the definition of the  $\mu(\mathbf{d}_k, \mathbf{d}'_k)$  and the fact that  $1 = \lambda_1 > \lambda_2 > \dots > \lambda_n$  as follows: Since  $|\mu(\mathbf{d}_k, \mathbf{d}'_k)| \leq \sum_{p=1}^n \lambda_p |A(d_{k,p}, d'_{k,p})|$  we can see from (2.43) and (2.44) that a minimum of  $|\mu(\mathbf{d}_k, \mathbf{d}'_k)|$  can only occur if  $A(d_{k,T}, d'_{k,T}) \neq 0$  for some  $T \in \{2, \dots, n\}$ .

We have with (2.30) that

$$\lambda_p - \varepsilon_{k,p} = \sum_{r=1}^k \gamma^{-(p-1)\beta(r)} \quad (2.45)$$

and consider the expression

$$A(d_{k,1}, d'_{k,1}) + \sum_{p=2}^T (\lambda_p - \varepsilon_{k,p}) A(d_{k,p}, d'_{k,p}). \quad (2.46)$$

We claim the following:

$$\text{If } A(d_{k,T}, d'_{k,T}) \neq 0 \text{ then } A(d_{k,1}, d'_{k,1}) + \sum_{p=2}^T (\lambda_p - \varepsilon_{k,p}) A(d_{k,p}, d'_{k,p}) \neq 0,$$

i.e. the term  $(\lambda_T - \varepsilon_{k,T}) A(d_{k,T}, d'_{k,T})$  cannot be annihilated by the preceding terms in the sum. To show this, an application of (2.45) leads to

$$\lambda_T - \varepsilon_{k,T} = \gamma^{-(T-1)} + \gamma^{-(T-1)\beta(2)} + \dots + \gamma^{-(T-1)\beta(k)}.$$

Also note that, for the choice  $k = 1$  and  $i_{1,T} = \gamma - 1$  as well as  $i'_{1,T} = 0$  in (2.23), the largest possible term in the expansion of  $|A(d_{k,T}, d'_{k,T})|$  in powers of  $\gamma^{-1}$  is

$$\frac{\gamma - 1}{\gamma}.$$

Therefore,  $(\lambda_T - \varepsilon_{k,T})|A(d_{k,T}, d'_{k,T})|$  contains at least one term  $\tau$  such that

$$0 < \tau \leq \gamma^{-(T-1)\beta(k)} \frac{\gamma - 1}{\gamma}.$$

But according to (2.43) and (2.45) the smallest possible term of  $(\lambda_p - \varepsilon_{k,p})|A(d_{k,p}, d'_{k,p})|$  for  $p < T$  is

$$\gamma^{-(T-2)\beta(k)} \gamma^{-\beta(k)} = \gamma^{-(T-1)\beta(k)}$$

so that the assertion holds and (2.46) indeed does not vanish.

If  $|i_{k,T} - i'_{k,T}| = 1$ , we have without loss of generality in the representation (2.23) the values

$i'_k$	$i_k$	$\tilde{s}'_1$	$s'_1$	$\theta'_0$	$\theta'^+_0$	$\tilde{s}_1$	$s_1$	$\theta_0$	$\theta^+_0$
$\gamma - 2$	$\gamma - 1$	0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1	1	0
$\gamma - 3$	$\gamma - 2$	0	0	1	0	0	$\frac{1}{2}$	1	0
$\gamma - 4$	$\gamma - 3$	0	0	1	0	0	0	1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

and we can directly infer that the expansion of (2.46) in powers of  $\gamma^{-1}$  contains the term

$$\gamma^{-(T-1)\beta(k)} \gamma^{-\beta(k)} = \gamma^{-T\beta(k)}. \quad (2.47)$$

We now show that this is the smallest term in the sum (2.46). To this end, we use the representation (2.23) for  $A(d_{k,p}, d'_{k,p})$  and factor out  $\gamma^{-\beta(k-j)}$  for each  $j$ . Since  $\theta_j$  and  $\theta_j^+$  become smaller than  $2^{-j}$ , we can bound each term in the sum (2.46) from below by  $\gamma^{-\beta(k-j)} 2^{-j}$ . The further estimation  $\gamma^{-\beta(k-j)} 2^{-j} > \gamma^{-\beta(k)}$  shows that (2.47) is indeed the smallest term in the sum and hence cannot be annihilated by other terms in (2.46). Therefore,

$$\left| A(d_{k,1}, d'_{k,1}) + \sum_{p=2}^T (\lambda_p - \varepsilon_{k,p}) A(d_{k,p}, d'_{k,p}) \right| \geq \gamma^{-T\beta(k)}.$$

But this implies that also

$$\left| A(d_{k,1}, d'_{k,1}) + \sum_{p=2}^T \lambda_p A(d_{k,p}, d'_{k,p}) \right| \geq \gamma^{-T\beta(k)}$$

since all possible terms in the expansion of  $\sum_{p=2}^T \varepsilon_{k,p} A(d_{k,p}, d'_{k,p})$  in powers of  $\gamma^{-1}$  are too small to annihilate  $\gamma^{-T\beta(k)}$ . Thus, choosing  $T = n$  the lemma is proven.  $\square$

Note that this lemma, which is taken from [12], estimates the minimum distance of grid points under the mapping  $\tilde{\Psi}_q$ ,  $q = 0$ . An estimate for  $\Psi_q$ ,  $q = 0$  directly follows from (2.33).

The linear combinations  $\Psi_q(\mathbf{x})$  of the inner functions serve for each  $q = 0, \dots, m$  as a mapping from the hypercube  $[0, 1]^n$  to  $\mathbb{R}$ . Therefore, further knowledge on the structure of these mappings is necessary. To this end, we need the following lemma which is also taken from [12]:

**Lemma 2.17.** *For each integer  $k \in \mathbb{N}$ , let*

$$\delta_k := \frac{\gamma - 2}{(\gamma - 1) \gamma^k}. \quad (2.48)$$

*Then for all  $d_k \in \mathcal{D}_k$  and  $\varepsilon_{k,2}$  as given in (2.30) we have*

$$\psi(d_k + \delta_k) = \psi(d_k) + (\gamma - 2) \varepsilon_{k,2}.$$

**Proof.** *The proof relies mainly on the continuity of  $\psi$  and some direct calculations. If we express  $\delta_k$  as an infinite sum we have*

$$d_k + \delta_k = \lim_{k_0 \rightarrow \infty} \left\{ d_k + \sum_{r=1}^{k_0} \frac{\gamma - 2}{\gamma^{k+r}} \right\} =: \lim_{k_0 \rightarrow \infty} d_{k_0}.$$

*Since  $\psi$  is continuous we get*

$$\psi \left( \lim_{k_0 \rightarrow \infty} d_{k_0} \right) = \lim_{k_0 \rightarrow \infty} \psi(d_{k_0})$$

*and since  $i_{k+r} = \gamma - 2$  for  $r = 1, \dots, k_0$ , it follows directly that  $\tilde{s}_r = 0$  for  $j = 0, \dots, k_0 - k$ . Therefore  $\theta_j^+ = 0$  and  $\theta_j^- = 1$  for  $j = 0, \dots, k_0 - k$ . With the representation (2.22) and the choice  $\xi = k_0 - k$ , the assertion follows.  $\square$*

As a direct consequence of this lemma, we have the following corollary, see [12], in which the one-dimensional case is treated.

**Corollary 2.18.** *For each integer  $k \in \mathbb{N}$  and  $d_k \in \mathcal{D}_k$ , the pairwise disjoint intervals*

$$E_k(d_k) := [d_k, d_k + \delta_k] \quad (2.49)$$

are mapped by  $\psi$  into the pairwise disjoint image intervals

$$H_k(d_k) := [\psi(d_k), \psi(d_k) + (\gamma - 2)\varepsilon_{k,2}]. \quad (2.50)$$

**Proof.** From their definition it follows directly that the intervals  $E_k(d_k)$  are pairwise disjoint. The corollary then follows from Lemma 2.16 and Lemma 2.17.  $\square$

We now generalize this result to the multidimensional case.

**Lemma 2.19.** For each fixed integer  $k \in \mathbb{N}$  and  $\mathbf{d}_k \in \mathcal{D}_k^n$ , the pairwise disjoint cubes

$$S_k(\mathbf{d}_k) := \prod_{p=1}^n [d_{k,p}, d_{k,p} + \delta_k] \quad (2.51)$$

are mapped by  $\Psi_q(\mathbf{x})$  from (2.36) for  $q = 0$  into the pairwise disjoint intervals

$$T_k(\mathbf{d}_k) := \left[ \sum_{p=1}^n \alpha_p \psi(d_{k,p}), \sum_{p=1}^n \alpha_p \psi(d_{k,p}) + \left( \sum_{p=1}^n \alpha_p \right) (\gamma - 2) \varepsilon_{k,2} \right]. \quad (2.52)$$

**Proof.** The fact that  $S_k(\mathbf{d}_k)$  is mapped into the interval  $T_k(\mathbf{d}_k)$  follows from Corollary 2.18 and the monotonicity of  $\psi$ . To prove that the intervals are disjoint we have to show that

$$\left( \sum_{p=1}^n \alpha_p \right) (\gamma - 2) \varepsilon_{k,2} \leq \left| \sum_{p=1}^n \alpha_q (\psi(d_{k,p}) - \psi(d'_{k,p})) \right|$$

for any two values  $\mathbf{d}_k \neq \mathbf{d}'_k$  in  $\mathcal{D}_k^n$ . From the proof of Lemma 4 in [111] we know that

$$\left( \sum_{p=1}^n \lambda_p \right) (\gamma - 2) \varepsilon_{k,2} \leq \gamma^{-n\beta(k)}.$$

With (2.33) and Lemma 2.16 we can directly estimate

$$\left( \sum_{p=1}^n \alpha_p \right) (\gamma - 2) \varepsilon_{k,2} < \Lambda_\delta^{-1} \gamma^{-n\beta(k)} < \left| \sum_{p=1}^n \alpha_q (\psi(d_{k,p}) - \psi(d'_{k,p})) \right|,$$

what shows the assertion.  $\square$

We now consider Algorithm 2.1 again. We need one more ingredient:

**Lemma 2.20.** For each  $r = 1, 2, \dots$ , the following estimate holds:

$$\|\phi^r(y)\| \leq \frac{1}{m+1} \|f_{r-1}\|.$$

**Proof.** The support of the functions  $\omega_{k,\mathbf{d}_k}^q(y)$ ,  $q = 0, \dots, m$ ,  $\mathbf{d}_k \in \mathcal{D}_k$  are the open intervals

$$U_k^q(\mathbf{d}_k) := (\xi_{k,\mathbf{d}_k}^q - \Gamma_k, \xi_{k,\mathbf{d}_k}^q + (\gamma - 2)b_k + \Gamma_k).$$

These intervals have empty intersections,

$$U_k^q(\mathbf{d}_k^1) \cap U_k^{q'}(\mathbf{d}_k^2) = \emptyset,$$

if either  $\mathbf{d}_k^1 \neq \mathbf{d}_k^2$  or  $q \neq q'$ . To show this, we first consider the case where  $q \in \{0, \dots, m\}$  is fixed. Here, we have to show that for  $\xi_{k,\mathbf{d}_k^2}^q > \xi_{k,\mathbf{d}_k^1}^q$  it holds

$$\xi_{k,\mathbf{d}_k^1}^q + (\gamma - 2)b_k + \Gamma_k < \xi_{k,\mathbf{d}_k^2}^q - \Gamma_k$$

or equivalently

$$(\gamma - 2)b_k + 2\Gamma_k < \xi_{k,\mathbf{d}_k^2}^q - \xi_{k,\mathbf{d}_k^1}^q.$$

With (2.35), (2.31) and Lemma 2.16 we can estimate

$$(\gamma - 2)b_k + 2\Gamma_k < \Lambda_\delta^{-1}\gamma^{-n\beta(k)} < \xi_{k,\mathbf{d}_k^2}^q - \xi_{k,\mathbf{d}_k^1}^q$$

what shows the assumption. Next, we define the values

$$a_q := \min_{\mathbf{d}_k \in \mathcal{D}_k^n} \{\xi_{k,\mathbf{d}_k}^q - \Gamma_k\} \quad (2.53)$$

and

$$b_q := \max_{\mathbf{d}_k \in \mathcal{D}_k^n} \{\xi_{k,\mathbf{d}_k}^q + (\gamma - 2)b_k + \Gamma_k\}. \quad (2.54)$$

Now, let  $q_1, q_2 \in \{0, \dots, m\}$ , such that without loss of generality  $q_1 < q_2$ . Simple computations, monotonicity of  $\psi$ , and the definition of  $\Lambda_\delta$  then show that

$$b_{q_1} = q_1 + 1 - \delta < q_2 = a_{q_2}. \quad (2.55)$$

Finally, this shows the assertion that all intervals  $U_k^q(\mathbf{d}_k)$ ,  $q = 0, \dots, m$ ,  $\mathbf{d}_k \in \mathcal{D}_k$ , have empty intersections. Now, to show Lemma 2.20 we see from (2.38) that  $0 \leq \omega_{k,\mathbf{d}_k}^q(y) \leq 1$  and with (2.41) we derive:

$$\|\phi^r(y)\| = \left\| \frac{1}{m+1} \sum_{q=0}^m \sum_{\mathbf{d}_{k_r}} f_{r-1}(\mathbf{d}_{k_r}) \omega_{k,\mathbf{d}_{k_r}}^q(y) \right\| = \frac{1}{m+1} \max_{\mathbf{d}_{k_r}} |f_{r-1}(\mathbf{d}_{k_r})|.$$

Here both, the sum and the maximum are built over all values  $\mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n$ . The lemma then follows from the definition of the maximum norm.  $\square$

For the following computations we define for given  $k \in \mathbb{N}$ ,  $q \in \{0, \dots, m\}$  and  $\mathbf{d}_k := (d_{k,1}, \dots, d_{k,n}) \in \mathcal{D}_k^n$  the values

$$d_{k,p}^q := d_{k,p} + q \sum_{r=2}^k \gamma^{-r}.$$

Note that all statements that are valid for points  $\mathbf{d}_k \in \mathcal{D}_k^n$  also hold for  $\mathbf{d}_k^q := (d_{k,1}^q, \dots, d_{k,n}^q) \in \mathcal{D}_k^n$ . We are now ready to prove Theorem 2.15.

**Proof.** [Proof of Theorem 2.15] For simplicity, we include the value  $d_k = 1$  in the definition of the rational numbers  $\mathcal{D}_k$ . Consider now for each integer  $q$  and  $a = [\gamma(\gamma - 1)]^{-1}$  the family of closed intervals

$$E_k^q(d_k^q) := [d_k^q - q a, d_k^q - q a + \delta_k]. \quad (2.56)$$

With  $\delta_k = (\gamma - 2)(\gamma - 1)^{-1}\gamma^{-k}$  we can see that

$$E_k^q(d_k^q) = \left[ d_k - \frac{q}{\gamma - 1}\gamma^{-k}, d_k - \frac{q}{\gamma - 1}\gamma^{-k} + \frac{\gamma - 2}{\gamma - 1}\gamma^{-k} \right]$$

and that these intervals are separated by gaps

$$G_k^q(d_k^q) := (d_k^q - q a + \delta_k, d_k^q - q a + \gamma^{-k})$$

of width  $(\gamma - 1)^{-1}\gamma^{-k}$ . With the intervals  $E_k^q$  we obtain for each  $k$  and  $q = 0, \dots, m$  the closed (Cartesian product) cubes

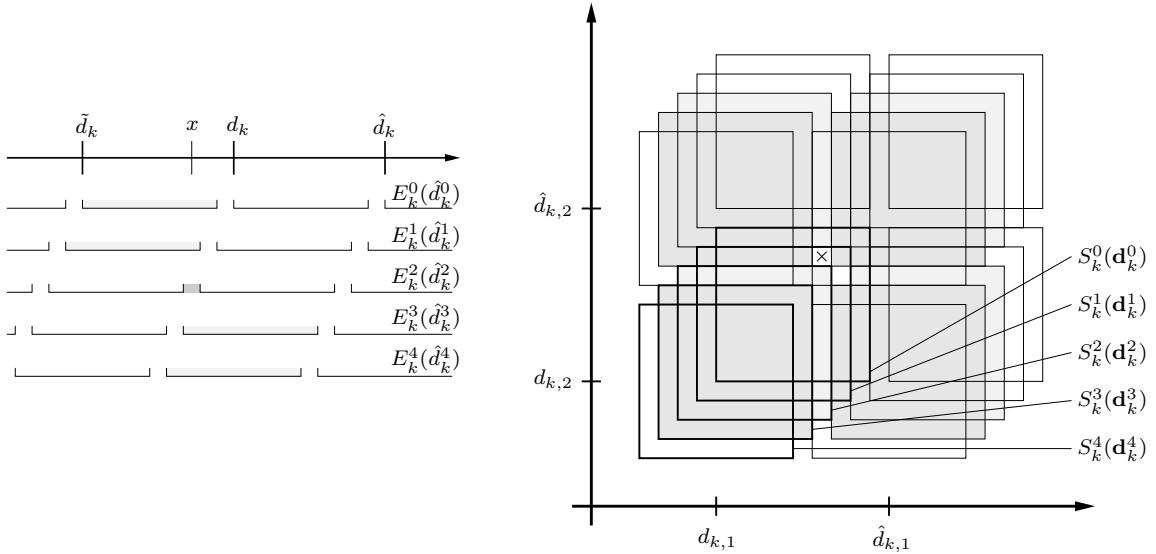
$$S_k^q(\mathbf{d}_k^q) := \prod_{p=1}^n [d_{k,p}^q - q a, d_{k,p}^q + \delta_k - q a],$$

whose images under  $\Psi_q(\mathbf{x}) = \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q$  are the disjoint closed intervals

$$\begin{aligned} T_k^q(\mathbf{d}_k^q) &:= \left[ \sum_{p=1}^n \alpha_p \psi(d_{k,p}^q) + \Delta_q, \sum_{p=1}^n \alpha_p \psi(d_{k,p}^q) + \Delta_q + \left( \sum_{p=1}^n \alpha_p \right) (\gamma - 2) \varepsilon_{k,2} \right] \\ &= \left[ \xi_{k,\mathbf{d}_k}^q, \xi_{k,\mathbf{d}_k}^q + (\gamma - 2) b_k \right], \end{aligned}$$

as derived in Lemma 2.19. For the two-dimensional case, the cubes  $S_k^q(\mathbf{d}_k^q)$  are depicted in Figure 2.7.

Now let  $k$  be fixed. The mapping  $\Psi_q(\mathbf{x})$  associates to each cube  $S_k^q(\mathbf{d}_k^q)$  from the coordinate space a unique image  $T_k^q(\mathbf{d}_k^q)$  on the real line. For fixed  $q$  the images of any two cubes from the set  $\{S_k^q(\mathbf{d}_k^q) : \mathbf{d}_k \in \mathcal{D}_k^n\}$  have empty intersections. This allows a local approximation of the target function  $f(\mathbf{x})$  on these images  $T_k^q(\mathbf{d}_k^q)$  for



**Figure 2.7.** Let  $k$  be a fixed integer,  $m = 4$ ,  $\gamma = 10$  and  $\tilde{d}_{k,i} := d_{k,i} - \gamma^{-k}$ ,  $\hat{d}_{k,i} := d_{k,i} + \gamma^{-k}$ ,  $i \in \{1, 2\}$ . The left figure depicts the intervals  $E_k^q(\tilde{d}_k)$  for  $q = 1, \dots, m$ . The subscript  $i$  indicating the coordinate direction is omitted for this one-dimensional case. The point  $x$  is contained in the intervals  $E_k^0(\tilde{d}_k^0)$ ,  $E_k^1(\tilde{d}_k^1)$ ,  $E_k^3(\tilde{d}_k^3)$ ,  $E_k^4(\tilde{d}_k^4)$  (shaded) and in the gap  $G_k^2(\tilde{d}_k^2)$  (dark shaded). The figure on the right shows the cubes  $S_k^q(\mathbf{d}_k)$  for  $n = 2$ ,  $q = 1, \dots, m$  and different values  $\mathbf{d}_k \in \mathcal{D}_k^n$ . For  $q \in \{2, 3\}$ , the marked point is not contained in any of the cubes from the set  $\{S_k^q(\mathbf{d}_k) : \mathbf{d}_k \in \mathcal{D}_k^n\}$ .

$\mathbf{x} \in S_k^q(\mathbf{d}_k^q)$ . However, as the outer function  $\phi^r$  has to be continuous, these images have to be separated by gaps in which  $f(\mathbf{x})$  cannot be approximated. Thus an error is introduced that cannot be made arbitrarily small. This deficiency is eliminated by the affine translations of the cubes  $S_k^q(\mathbf{d}_k^q)$  through the variation of the  $q$ 's. To explain this in more detail, let  $x \in [0, 1]$  be an arbitrary point. With (2.56) we see that the gaps  $G_k^q(d_k^q)$  which separate the intervals do not intersect for variable  $q$ . Therefore, there exists only one value  $q_*$  such that  $x \in G_k^{q_*}(d_k^{q_*})$ . This implies that for the remaining  $m$  values of  $q$  there holds  $x \in E_k^q(d_k^q)$  for some  $d_k$ . If we now consider an arbitrary point  $\mathbf{x} \in [0, 1]^n$ , we see that there exist at least  $m - n + 1$  different values  $q_j$ ,  $j = 1, \dots, m - n + 1$  for which  $\mathbf{x} \in S_k^{q_j}(\mathbf{d}_k^{q_j})$  for some  $\mathbf{d}_k$ . Note that the points  $\mathbf{d}_k$  can differ for different values  $q_j$ . From (2.56) we see that  $\mathbf{d}_k \in S_k^{q_j}(\mathbf{d}_k^{q_j})$ .

Now we consider step I of Algorithm 2.1. To this end, remember that  $\varepsilon$  is a fixed number such that  $0 < \frac{m-n+1}{m+1}\varepsilon + \frac{2n}{m+1} = \eta < 1$ . Let  $k_r$  be the integer given in step I with the associated assumption that  $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \varepsilon \|f_{r-1}\|$  when  $|x_p - x'_p| \leq \gamma^{-k_r}$  for  $p = 1, \dots, n$ . Let  $\mathbf{x} \in [0, 1]^n$  be an arbitrary point and let  $q_j$ ,  $j = 1, \dots, m - n + 1$ , denote the values of  $q$  such that  $\mathbf{x} \in S_k^{q_j}(\mathbf{d}_k^{q_j})$ . For the point

$\mathbf{d}_{k_r} \in S_{k_r}^{q_j}(\mathbf{d}_{k_r}^{q_j})$  we have

$$|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{d}_{k_r})| \leq \varepsilon \|f_{r-1}\| \quad (2.57)$$

and for  $\mathbf{x}$  it holds that  $\Psi_{q_j}(\mathbf{x}) \in T_{k_r}^{q_j}(\mathbf{d}_{k_r}^{q_j})$ . The support  $U_{k_r}^{q_j}(\mathbf{d}_{k_r})$  of the function  $\omega_{k_r, \mathbf{d}_{k_r}}^{q_j}(y)$  contains the interval  $T_{k_r}^{q_j}(\mathbf{d}_{k_r}^{q_j})$ . Furthermore, from definition (2.38) we see that it is constant on that interval. With (2.42) we then get

$$\begin{aligned} \phi^r \circ \Psi_{q_j}(\mathbf{x}) &= \frac{1}{m+1} \sum_{q=0}^m \sum_{\mathbf{d}_{k_r}^q} f_{r-1}(\mathbf{d}_{k_r}) \omega_{k_r, \mathbf{d}_{k_r}}^q(\Psi_{q_j}(\mathbf{x})) \\ &= \frac{1}{m+1} f_{r-1}(\mathbf{d}_{k_r}). \end{aligned}$$

Together with (2.57) this shows

$$\left| \frac{1}{m+1} f_{r-1}(\mathbf{x}) - \phi^r \circ \Psi_{q_j}(\mathbf{x}) \right| \leq \frac{\varepsilon}{m+1} \|f_{r-1}\|$$

for all  $q_j$ ,  $j = 1, \dots, m-n+1$ . Note that this estimate does not hold for the remaining values of  $q$  for which  $\mathbf{x}$  is not contained in the cube  $S_{k_r}^q(\mathbf{d}_{k_r}^{q_j})$ . Let us now denote these values by  $\bar{q}_j$ ,  $j = 1, \dots, n$ . We can apply Lemma 2.20 and with the special choice of the values  $\varepsilon$  and  $\eta$  we obtain the estimate

$$\begin{aligned} |f_r(\mathbf{x})| &= \left| f_{r-1}(\mathbf{x}) - \sum_{q=0}^m \phi^r \circ \Psi_q(\mathbf{x}) \right| \\ &= \left| \sum_{q=0}^m \frac{1}{m+1} f_{r-1}(\mathbf{x}) - \sum_{j=1}^{m-n+1} \phi^r \circ \Psi_{q_j}(\mathbf{x}) - \sum_{j=1}^n \phi^r \circ \Psi_{\bar{q}_j}(\mathbf{x}) \right| \\ &\leq \left| \frac{n}{m+1} f_{r-1}(\mathbf{x}) + \sum_{j=1}^{m-n+1} \frac{1}{m+1} f_{r-1}(\mathbf{x}) - \phi^r \circ \Psi_{q_j}(\mathbf{x}) \right| + \frac{n}{m+1} \|f_{r-1}\| \\ &\leq \left[ \frac{m-n+1}{m+1} \varepsilon + \frac{2n}{m+1} \right] \|f_{r-1}\| = \eta \|f_{r-1}\|. \end{aligned}$$

This completes the proof of Theorem 2.15.  $\square$

We now state a fact that follows directly from the previous results.

**Corollary 2.21.** For  $r = 1, 2, 3, \dots$  the following estimates hold:

$$\|\phi^r(y)\| \leq \frac{1}{m+1} \eta^{r-1} \|f\| \quad (2.58)$$

and

$$\|f_r\| = \left\| f(\mathbf{x}) - \sum_{q=0}^m \sum_{j=1}^r \phi^j \circ \Psi_q(\mathbf{x}) \right\| \leq \eta^r \|f\|. \quad (2.59)$$

**Proof.** Remember that  $f_0 \equiv f$ . The first estimate follows from Lemma 2.20 and a recursive application of Theorem 2.15. The second estimate can be derived from the definition (2.42) of  $f_r$  and again a recursive application of Theorem 2.15.  $\square$

Since  $\eta < 1$ , the estimate (2.59) shows that the partial sum

$$\Phi^r(y) := \sum_{j=1}^r \phi^j(y) \quad (2.60)$$

gives an approximative representation of the function  $f$ . We finally are in the position to prove Theorem 2.14.

**Proof.** [Proof of Theorem 2.14.] From Corollary 2.21 and the fact that  $\eta < 1$  it follows that we have

$$\left\| \sum_{j=1}^r \phi^j(y) \right\| \leq \sum_{j=1}^r \|\phi^j(y)\| \leq \frac{1}{m+1} \|f\| \sum_{j=0}^{r-1} \eta^j < \frac{1}{m+1} \|f\| \sum_{j=0}^{\infty} \eta^j < \infty.$$

The functions  $\phi^j(y)$  are continuous and therefore the series  $\sum_{j=1}^r \phi^j(y)$  converges absolutely to a continuous function

$$\Phi(y) = \lim_{r \rightarrow \infty} \sum_{j=1}^r \phi^j(y). \quad (2.61)$$

Since  $\eta < 1$  we see from the second estimate in Corollary 2.21 that  $f_r \rightarrow 0$  for  $r \rightarrow \infty$ . This finally proves Theorem 2.14.  $\square$

Next, we comment on the parameter

$$\eta := \frac{m-n+1}{m+1} \varepsilon + \frac{2n}{m+1}$$

defined in (2.39) which determines the convergence speed of Algorithm 2.1 as can be seen from (2.59). For a fixed dimension  $n$  it depends on the parameters  $\varepsilon$  and  $m$ . We first consider the case where  $\varepsilon > 0$  is fixed. Let now  $\tilde{m} > m \geq 2n$  be positive integers. Then, simple computations show that

$$\tilde{\eta} := \frac{\tilde{m}-n+1}{\tilde{m}+1} \varepsilon + \frac{2n}{\tilde{m}+1} \leq \frac{m-n+1}{m+1} \varepsilon + \frac{2n}{m+1} = \eta. \quad (2.62)$$

Therefore, we can increase the speed of convergence if  $m$  is large.

Finally, we briefly point out the influence of the parameter  $\varepsilon$  on Algorithm 2.1. To this end, remember that in step  $I$  the integers  $k_r$  have to be determined such that  $|f_{r-1}(\mathbf{x}) - f_{r-1}(\mathbf{x}')| \leq \varepsilon \|f_{r-1}\|$  for all values  $\mathbf{x}, \mathbf{x}' \in [0, 1]^n$  with  $|x_p - x'_p| \leq \gamma^{-k_r}$  for  $p = 1, \dots, n$ . Then, in the following steps the sums are built over all values  $\mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n$  resulting in  $(\gamma^{k_r} + 1)^n$  terms. Therefore, the complexity of Algorithm 2.1 depends on  $k_r$ . Clearly, the complexity can be minimized by maximizing the parameter  $\varepsilon$  since this allows for small values  $k_r$ .

Now, let  $\tilde{\varepsilon} > \varepsilon > 0$  and  $\tilde{m} \in \mathbb{N}$  such that the equation

$$\eta = \frac{\tilde{m} - n + 1}{\tilde{m} + 1} \tilde{\varepsilon} + \frac{2n}{\tilde{m} + 1} = \frac{m - n + 1}{m + 1} \varepsilon + \frac{2n}{m + 1}$$

holds. This implies  $\tilde{m} > m$ . Thus, to achieve the same speed of convergence  $\eta$  and reduce the complexity of Algorithm 2.1, the number of terms  $m$  in the outer sum has to be increased.

## 2.4 Relation to Space Filling Curves

Let us finish this section with some further remarks on the inner function  $\psi$ . In [113] it was first mentioned that there is a close relation of  $\psi$  to space-filling curves. Loosely speaking, the inner sum  $\Psi := \Psi_0$ , see (2.36), acts on the  $n$ -dimensional unit cube like a Z-curve to the base  $\gamma$ . To illustrate this and the resulting consequences in more detail we return to the proof of Theorem 2.14 again.

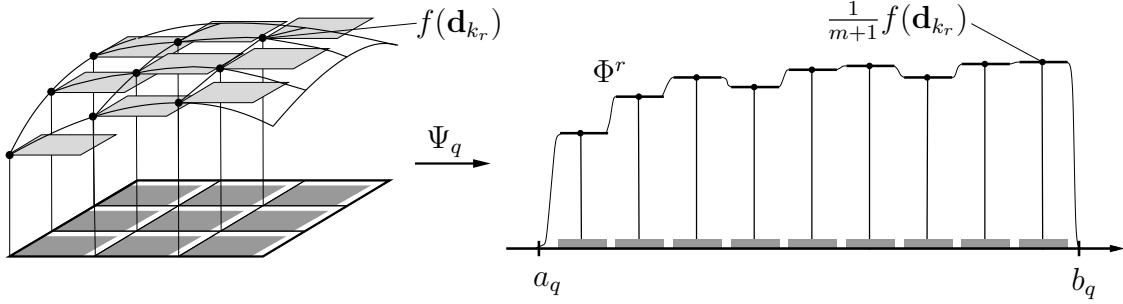
We have seen that the outer function  $\Phi$  in (2.36) is the limit of the series  $\sum_{j=1}^{\infty} \phi^j$  for which each term  $\phi^r$  is constructed by the  $r$ -th iteration step of Algorithm 2.1 in the following way.

The integer  $k_r \in \mathbb{N}$  is fixed to an appropriate value and the  $n$ -dimensional unit cube  $[0, 1]^n$  is divided into  $m + 1$  families of cubes  $\mathcal{S}_{k_r}^q := \{S_{k_r}^q(\mathbf{d}_{k_r}^q) : \mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n\}$ ,  $q = 0, \dots, m$  with edge length  $\delta_{k_r}$ . For each cube  $q = 0, \dots, m$  the point  $\mathbf{d}_{k_r}$  is contained in  $S_{k_r}^q(\mathbf{d}_{k_r}^q)$ . For fixed  $q \in \{0, \dots, m\}$  the following holds. The elements of  $\mathcal{S}_{k_r}^q$  are mutually separated by gaps and the inner sum  $\Psi_q$  uniquely maps the cube  $S_{k_r}^q(\mathbf{d}_{k_r}^q)$  to the interval  $T_{k_r}^q(\mathbf{d}_{k_r}^q)$  on the real line. It can be shown that the image-intervals are also separated by gaps which allows for the definition of a piecewise constant, continuous function  $\phi^r$  on the real line. Then, with (2.41) and (2.42) for this function it holds that

$$\phi^r|_{T_{k_r}^q(\mathbf{d}_{k_r}^q)} = \frac{1}{m+1} f_{r-1}(\mathbf{d}_{k_r}) = \frac{1}{m+1} \left( f(\mathbf{d}_{k_r}) - \sum_{q=0}^m \sum_{j=1}^{r-1} \phi^j \circ \Psi_q(\mathbf{d}_{k_r}) \right).$$

Now, for  $j = 1, \dots, r - 1$  it holds that  $\Psi_q(\mathbf{d}_{k_r}) \in T_{k_r}^q(\mathbf{d}_{k_r}^q) \subset T_{k_j}^q(\mathbf{d}_{k_j}^q)$  and therefore

$$\phi^j \circ \Psi_q(\mathbf{d}_{k_r}) = \frac{1}{m+1} f_{j-1}(\mathbf{d}_{k_r}) = \phi^j|_{T_{k_j}^q(\mathbf{d}_{k_j}^q)} = \phi^j|_{T_{k_r}^q(\mathbf{d}_{k_r}^q)}.$$



**Figure 2.8.** Construction of the  $r$ -th approximation  $\Phi^r$  on the interval  $[a_q, b_q]$  from the original function  $f$ . In the left picture the family of cubes  $S_k^q$  is dark shaded. The image-intervals  $T_{k_r}^q(\mathbf{d}_k^q)$  are marked gray in the right picture.

This shows that the restriction of the previous  $\phi^j$  to the considered interval are also constants and we finally get that

$$\phi^r|_{T_{k_r}^q(\mathbf{d}_k^q)} = \frac{1}{m+1}f(\mathbf{d}_{k_r}) - \sum_{j=1}^{r-1} \phi^j|_{T_{k_r}^q(\mathbf{d}_k^q)},$$

or equivalently with (2.60)

$$\Phi^r|_{T_{k_r}^q(\mathbf{d}_k^q)} \equiv \frac{1}{m+1}f(\mathbf{d}_{k_r}).$$

Thus, the  $r$ -th approximation  $\Phi^r$  to the outer function  $\Phi$  is a piecewise constant function on the disjoint intervals  $T_{k_r}^q(\mathbf{d}_k^q)$  where it takes the scaled function values of  $f$ . This is sketched in Figure 2.8. Now, to get an impression on the graph of  $\Phi^r$  it is necessary to investigate the ordering of these intervals on the real line.

To this end, we define an order of points in  $n$ -dimensional space by

$$\mathbf{d}_k < \mathbf{d}'_k : \Leftrightarrow \Psi_q(\mathbf{d}_k) < \Psi_q(\mathbf{d}'_k). \quad (2.63)$$

Note that due to monotonicity of  $\psi$  this definition is independent of  $q$ . This ordering gives insight into the shape of the functions  $\Phi^r$  on the intervals  $[a_q, b_q]$ , see (2.53) and (2.54). This is due to the fact that it is, for fixed  $q$ , also the ordering of the intervals  $T_{k_r}^q(\mathbf{d}_k^q)$  on the real line and

$$T_{k_r}^q(\mathbf{d}_k^q) \subset [a_q, b_q] \quad \text{for all } \mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n.$$

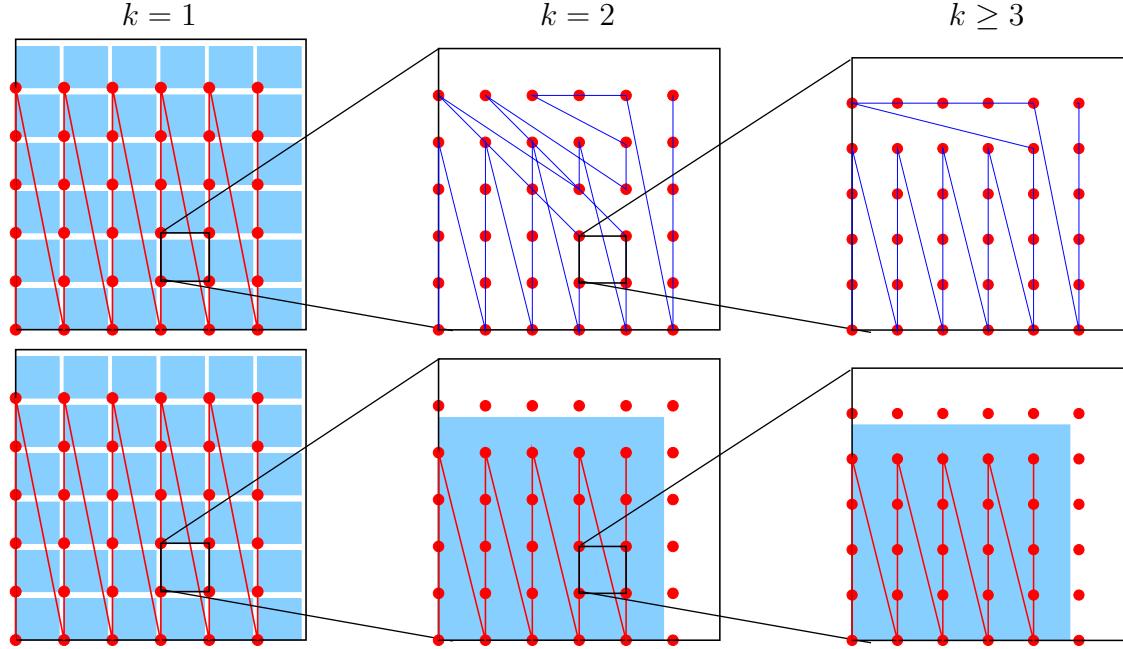
To get the full graph of  $\Phi^r$  remember that

$$a_0 < b_0 < \cdots < a_m < b_m.$$

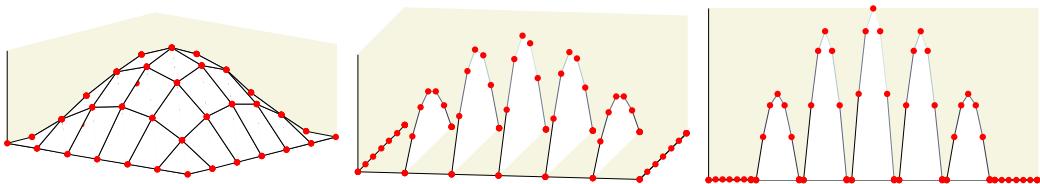
Therefore, we can fix  $q \in \{0, \dots, m\}$  in the following investigations. The top row in Figure 2.9 shows the general ordering of the points  $\mathbf{d}_k$  for  $k = 1, 2, 3$ . For  $k > 3$  it is the same as for  $k = 3$  and the pictures do not change. This is due to the recursive definition of  $\psi$  which causes self similarities in its graph. In the first iteration  $k = 1$ , the points are lexicographically ordered in a Z-curve shape. For  $k > 1$  this is not the case any more. At this point the separation of the cubes by gaps comes into play. The function  $\Phi^r$  is constructed in such a manner that  $\sum_{q=0}^m \Phi^r \circ \Psi_q$  approximates  $f$  only on the family of cubes  $\mathcal{S}_{k_r}^q$ ,  $q \in \{0, \dots, m\}$ . Now, if we restrict our investigation of the ordering of points  $\mathbf{d}_{k_r}$  to the region of approximation we see that the points are again lexicographically ordered. This is indicated in the middle column of Figure 2.9. Here, we zoomed into the marked region of the left picture. The top image shows the general ordering of the points  $\mathbf{d}_k$  for  $k = 2$  while the bottom picture shows the restriction to the cube  $S_k^q(\mathbf{d}_k^q)$ ,  $k = 1$ . The right column is a zoom into its left neighbor. Again, the general ordering of points  $\mathbf{d}_3$  is shown in the top and the restriction to  $S_2^q(\mathbf{d}_2^q)$  in the bottom image. Since the ordering of points does not change for  $k > 3$  we conclude that on the union of all cubes  $S_k^q(\mathbf{d}_k^q)$ ,  $\mathbf{d}_k \in \mathcal{D}_k$ ,  $k \geq 1$  the points are ordered lexicographically.

The relation of Sprecher's function  $\psi$  to space filling curves was first investigated in [113]. There, a space filling curve with Lebesgue measure 1 was constructed with approximating curves  $\mathcal{C}_k$  that were induced by (2.63) and  $\gamma = 10$ . See Figure 2.11 for an illustration of the approximating curve  $\mathcal{C}_k$  for  $k = 2$ . However, note that in [113] the curve was constructed with Sprecher's original non-continuous inner function  $\psi$  and a final proof of the result would require an adjustment of the arguments to Köppen's function. Figure 2.11 uses the corrected function  $\psi$ .

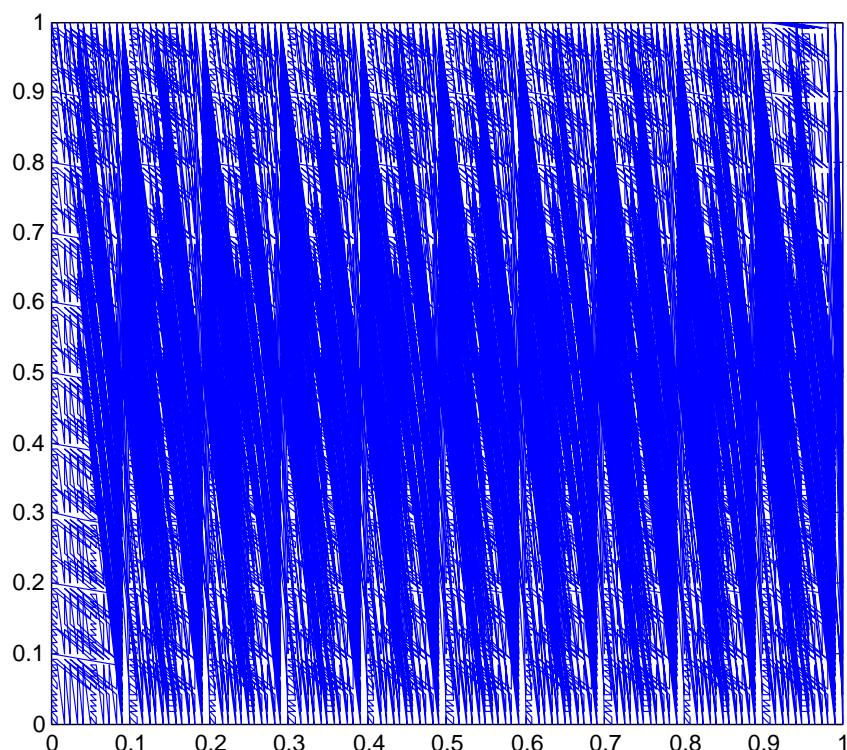
From this knowledge we now can infer the general shape of the graph of the  $r$ -th approximation  $\Phi^r$  to the outer function  $\Phi$ . Intuitively, following the Z-curve, induced by (2.63), on the image of a function  $f$  gives us, with a scaling by  $1/(m+1)$ , the outer function  $\Phi^r$  on each interval  $[a_q, b_q]$ ,  $q = 0, \dots, m$ . This can be interpreted as an “unfolding of the dimensions”, see Figure 2.10.



**Figure 2.9.** The top row shows the general ordering of the points  $\mathbf{d}_k$  for  $k = 1, 2, 3$ . The cubes  $S_k^q(\mathbf{d}_k^q)$ ,  $q = 0$  are marked blue and the lines follow the points  $\mathbf{d}_k$  according to (2.63) in increasing order. The pictures are successive zooms into the marked sub cube of their left neighbors. In the bottom row the ordering of the points in the respective cubes  $S_k^0(\mathbf{d}_k^0)$  is indicated.



**Figure 2.10.** Unfolding of dimensions for the two-dimensional function  $f(\mathbf{x}) = \sin(\pi x_1) \sin(\pi x_2)$  on the interval  $[a_q, b_q]$ ,  $q \in \{0, \dots, m\}$ .



**Figure 2.11.** General ordering of all points  $\mathbf{d}_k$ ,  $k = 2$  in  $[0, 1]^2$  ( $\gamma = 10$ ) induced by (2.63).

## Chapter 3

# Approximative versions of Kolmogorov's superposition theorem

In view of Kolmogorov's result, it seems that there are no high-dimensional functions and thus no high-dimensional problems at all. Furthermore, the criticism that the representation is not constructive could also be rejected by the proofs of Theorem 2.1 and Theorem 2.14. Despite these results it turns out that the exact representation of functions by means of superpositions of one-dimensional functions still is of no use in several applications. First, it turned out that the representing functions are quite bad [125, 126], i.e. they are at best only continuous and highly non-smooth. In particular, the inner functions cannot be chosen to be differentiable. For these reasons, their practical use for approximation and interpolation purposes is limited, like e.g., for the discretization of partial differential equations within the Galerkin approach. Additionally, (2.61) still involves in general an infinite number of iterations for the determination of the outer function. This cannot be accomplished by a computer program and a termination of the algorithm after a finite number of steps always results in an approximation of the function  $f$ .

In this section we will discuss approximation schemes that are based on the Kolmogorov superposition theorem. Note, that the constructive versions of Kolmogorov's theorem, presented in Section 2.3, are very recent results. Therefore, most of the approximation schemes that will be presented here, are based on previous versions of the theorem, see Chapter 2.

We have already seen that the inner functions in Kolmogorov's theorem are non-smooth. This is a necessary condition as long as we wish to exactly represent continuous functions. An inverse question is now how broad is the class of functions that can be approximated using superpositions of a single, arbitrary nonlinear continuous function. Gorban showed in [44] that every continuous function can be arbitrarily accurately approximated by the operations of addition, multiplication by a number, superposition, and an arbitrary (one is sufficient) continuous nonlinear function of one variable. For neural networks, see below, this means that the activation function must be nonlinear.

A promising strategy to construct an approximation scheme out of Kolmogorov's superposition theorem is to avoid non-constructive arguments in its proof, like infinite limit processes.

A first example for this is a result presented by de Figueiredo. Based on constructions in the proof of (2.3) he computed in [22] a piecewise linear approximation  $\tilde{\psi}$  of the inner function. Then, the outer function was determined from an interpolation problem. Here, for data points  $(f(\mathbf{x}_j), \mathbf{x}_j)_{j=1}^P$  the parameters  $a_r$  of an interpolant  $\sum_{r=1}^{\ell} a_r \varphi_r(t)$  were obtained by demanding for  $j = 1, \dots, P$  that

$$f(\mathbf{x}_j) = \sum_{r=1}^{\ell} \sum_{q=0}^{2n} a_r \varphi_r(t_{j,q}), \quad \text{where} \quad t_{j,q} := \sum_{p=1}^n \lambda_p \tilde{\psi}(x_{p,j} + \eta q) + q.$$

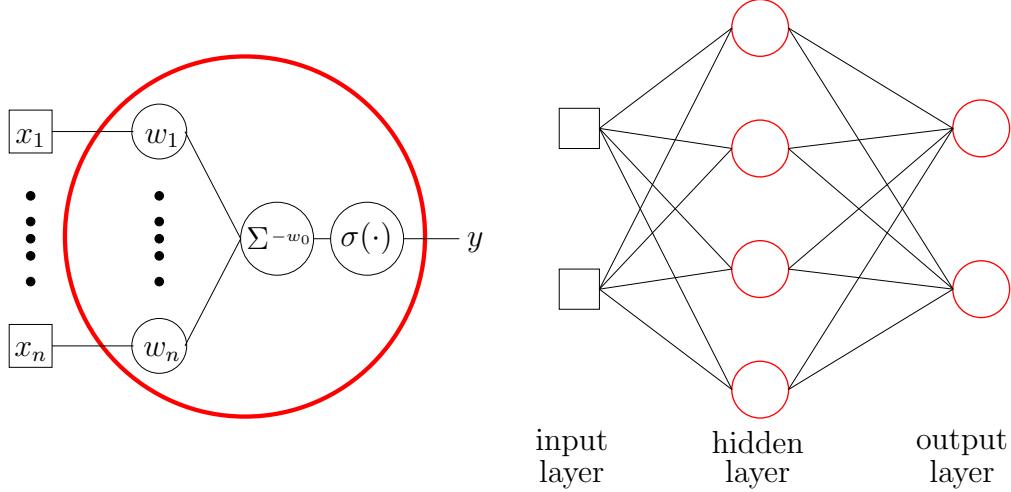
A slightly different idea was presented very briefly for a single two-dimensional numerical example in [36]. Here, applying constructions from a proof of a representation of the form (2.6) by Lorentz in [81], fixed approximations for the inner functions were used while the outer function was computed iteratively. Unfortunately, the authors do not give any further details on their implementation.

In [85] Nees approximated both, inner and outer functions, by sequences  $\{g_r\}_r$ ,  $\{\psi_q^r\}_r$ ,  $q = 1, \dots, 2n+1$  of continuous functions that were constructed explicitly. She also gave an upper bound for the point wise approximation error at the  $r$ th iterate. In a further paper [86], the algorithms were extended such that the approximating function can be constructed from discrete data sets. Also, the number of terms was increased from  $2n+1$  to  $m$  to achieve faster decay of the approximation error. In particular,  $m$  can be chosen such that one iteration suffices to attain every error bound. Then, this result was applied to a feedforward neural network with two hidden layers, see below, to give an upper bound on the number of neurons in the respective layers to achieve a desired accuracy. Furthermore, the estimates showed that for Nees' algorithms the computational costs to compute the respective functions in the sequences grow exponentially with the dimensionality  $n$ .

Note that the construction of a function sequence that approximates the outer function in Kolmogorov's representation is similar to Sprecher's constructive proof where the outer function is approximated by a function series that is computed by an algorithm.

### 3.1 Relations to neural networks and other approximation schemes

Beginning with the works by Hecht–Nielsen from 1987, [53, 54] Kolmogorov's superposition theorem found much attention in the literature on neural networks. Before this is explained in more detail, we briefly introduce what we mean by a feedforward neural network.



**Figure 3.1.** Sketch of the functionality of a neuron (left) and graphical illustration of a multilayer feedforward neural network (right).

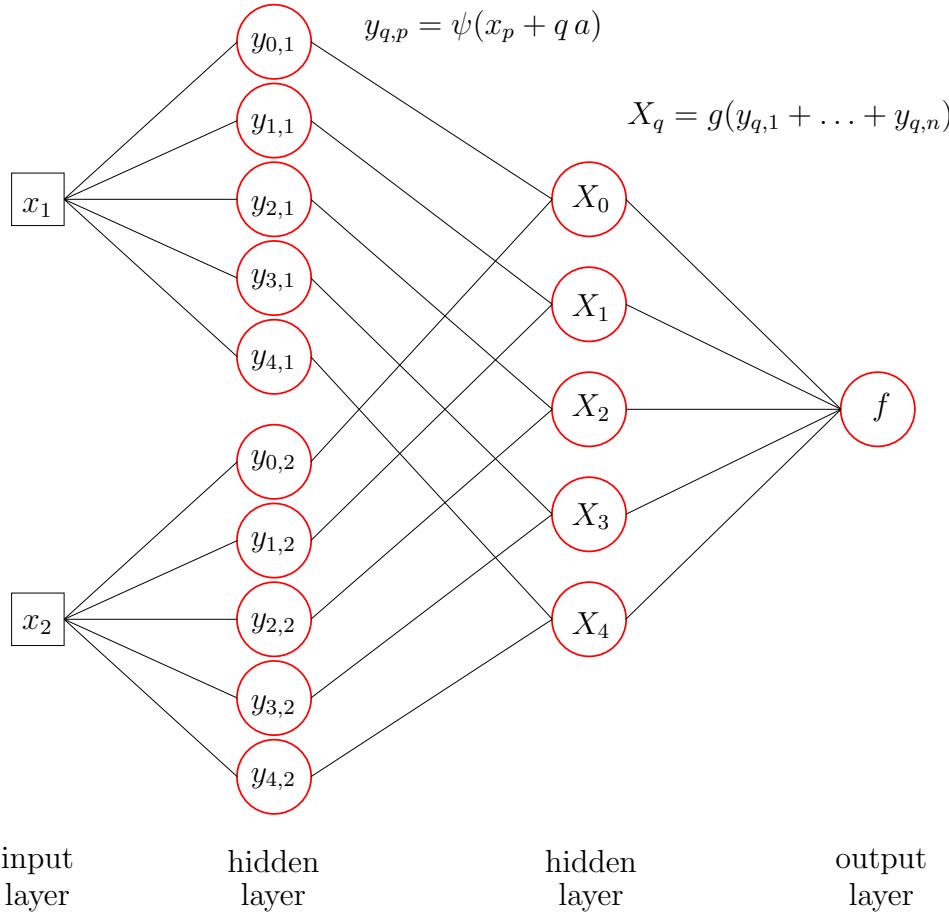
In [84], a neural network was described as a way to perform computations by networks of interconnected neurons. In general, a neuron has  $n$  real inputs  $x_1, \dots, x_n$  and produces its output by

$$y = \sigma(w_1x_1 + \dots + w_nx_n - w_0).$$

Here,  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  is called activation function and the parameters  $w_0, \dots, w_n$  are called weights. See the left picture in Figure 3.1 for an illustration of the functionality of a single neuron. A neural network sends the output of some neurons as inputs to others, see Figure 3.1 (right) and [51]. A net is trained, i.e. its parameters are determined, with input signals for which the output is known. The weights  $w_j$  of each neuron are changed in such a manner that the output of the net fits the training data in a prescribed way.

Hecht-Nielsen observed in [53, 54] the following interpretation of (2.3) as a feed-forward neural network: Any continuous function  $f : [0, 1]^n \rightarrow \mathbb{R}$  can be represented exactly by a neural network with one hidden layer and activation function  $\psi$ , and a single output layer with activation function  $g$  that produces the output  $f(x_1, \dots, x_n)$ .

We remark that our terminology on feedforward neural network slightly differs from [53], where Kolmogorov's representation is identified to have three layers. Since this disagreement is also present in the literature on neural networks that uses Hecht–Nielsen's interpretation, we adhere to our definitions and concretize the functionality of the net in more detail here. The input layer sends the inputs  $x_1, \dots, x_n$  to the first hidden layer with  $n(2n + 1)$  neurons and activation function  $\psi$ . Here, the outputs  $y_{q,p} = \psi(x_p + q a)$ ,  $q = 0, \dots, 2n$ ,  $p = 1, \dots, n$  are generated and sent to the second hidden layer that consists of  $2n + 1$  neurons, activation function  $g$ , and produces



**Figure 3.2.** Sketch of Hecht–Nielsen's interpretation of the Kolmogorov superposition theorem as neural network. The net is assembled by an input layer, two hidden layers and the output layer.

the outputs  $X_q = g(y_{q,1} + \dots + y_{q,n})$ . Finally, the output layer just sums up the inputs, i.e. its activation function is the identity, to produce  $f(x_1, \dots, x_n)$  exactly. See Figure 3.2 for an illustration of the net. Altogether this gives us a four-layered feedforward network with two hidden layers.

Due to the lack of a constructive proof of Kolmogorov's theorem at the time when Hecht–Nielsen formulated this result, he also emphasized the limited practical use of the results in [53]: The activation functions  $\psi$  and  $g$  are not known explicitly. This still holds true for the function  $g$ , even if we take under consideration the recent constructive proofs. Nonetheless, this interpretation was an important result and allowed for many further investigations on the relation of the Kolmogorov superposition theorem and neural networks. We will give some examples that fit into our context of function approximation.

The fact that the non-smoothness of inner and outer functions is problematic

in neural networks was first mentioned by Girosi and Poggio. In [43], they argued the following way: The intuition underlying Hilbert's conjecture is that not all functions with a given degree of complexity can be represented in simple way by means of functions with a lower degree of complexity. In this sense, Kolmogorov's theorem shows that the number of variables is not sufficient to characterize the complexity of a function.<sup>1</sup> In approximation theory, the choice of representation for the function is equivalent to the choice of a particular network architecture. Therefore, Kolmogorov's representation suggests that a network with two hidden layers allows for exact representation. However, for an implementation of a network that has good generalization properties and is robust against noise, the functions corresponding to the units in the network have to be smooth. This is not the case for the inner and outer functions in Kolmogorov's theorem. Additionally, since the inner functions are universal, the outer function has to be of the same complexity as the approximated function itself. Therefore, an exact representation in terms of two or more layers network is not possible and one has to ask for approximate representations. This led Girosi and Poggio to the conclusion that Kolmogorov's theorem is irrelevant in this context.

Directly referring to this criticism Kurkova pointed out in [75, 76] that the question on the relevance of Kolmogorov's theorem for approximation is different. She argued that in this context one has to ask whether Kolmogorov's construction can be modified in such a way that the one-dimensional functions were limits of sequences of smooth functions used in perceptron type networks. In a perceptron type network all one-variable functions are finite linear combinations  $\sum_r a_r \sigma(b_r x + c_r)$  of affine transformations of a single arbitrary continuous sigmoidal<sup>2</sup> activation function  $\sigma : \mathbb{R} \rightarrow [0, 1]$ . Additionally, Kurkova increased the number of terms in the outer sum from  $2n+1$  to a larger  $m$  what resulted in an approximation model of the form

$$f_{M,N}(x_1, \dots, x_N) = \sum_{q=1}^m \left( \sum_{j=1}^M \left( d_j \sigma \left( \sum_{p=1}^n \sum_{i=1}^N v_j w_{pq} a_{qi} \sigma(b_{qi} x_q + c_{qi}) + u_j \right) \right) \right),$$

where the net parameters  $d_j, v_j, u_j$ , and  $w_{pq}, a_{qi}, b_{qi}, c_{qi}$  are real numbers. Her direct approach also enabled an estimation of the number of hidden units (neurons) as a function of the desired accuracy  $\epsilon$  and the modulus of continuity of  $f$  being approximated. For general continuous functions one has to choose  $N = m + 1$  and  $M = m(m+1)^n$  to achieve an error below  $\epsilon$ . Note that  $m$  depends on  $\epsilon$  and increases when  $\epsilon \rightarrow 0$ .

This dependency has been formally removed in [84]. Here, based on the proof of (2.3) from [107], a constructive algorithm is proposed that approximates a function

---

<sup>1</sup>Vitushkin introduced a characteristic to measure the inverse complexity of a function. He could show that not all functions of a given characteristic can be represented by superpositions of functions with higher characteristic, [43].

<sup>2</sup> $\sigma : \mathbb{R} \rightarrow [0, 1]$  is called sigmoidal function if  $\lim_{t \rightarrow -\infty} \sigma(t) = 0$  and  $\lim_{t \rightarrow +\infty} \sigma(t) = 1$ .

$f$  to any desired accuracy  $\epsilon$  with one universal design. All numbers, operations and functions in (2.3) were defined in terms of constructive mathematics which means that they are implementable in a computer program. By universal design the authors of [84] mean that the number of neurons, and the weights remain the same for all  $\varepsilon > 0$ . In their algorithm, the accuracy of the approximation was completely controlled via the exactness of the computations of the outer and the inner function. This is closely related to [85, 111, 112]. However, we remark that [84] does not provide an estimation of the total complexity of the algorithm while Kurkova's result gives us insight in the total number of degrees of freedom that are needed to compute the approximation.

Based on Kurkova's idea, Sprecher and Katsura constructed in [67] by means of a fixed continuous mapping  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  a sequence  $\{\psi_k\}_k$  that converged to the inner functions from [110]. Using the same  $\sigma$ , they constructed  $2n + 1$  series of functions that converged uniformly to the outer functions.

Beside these constructive approaches, there are many theoretical results in the neural networks literature that employ Kolmogorov's theorem to generalize the property that the neural net under consideration is a universal approximator for one-dimensional functions to the  $n$ -dimensional case. Here, a neural network is called an universal approximator if for any desired accuracy  $\varepsilon > 0$ , there exists a finite number of neurons in a finite number of layers, and a network architecture  $NN : [0, 1]^n \rightarrow \mathbb{R}$ , such that  $|f(\mathbf{x}) - NN(\mathbf{x})| < \varepsilon$  for all  $\mathbf{x} \in [0, 1]^n$ . To show that this can be accomplished by the network, it is argued in the following way: First,  $f$  is represented by Kolmogorov's superpositions. Then, the fact that the net under consideration is a universal approximator for one-dimensional functions, see e.g. [20], is used to show that inner and outer functions can be approximated to any desired accuracy what shows that the respective neural network is an universal approximator for  $n$ -dimensional continuous functions. Examples for this can be found in [10, 11, 37, 71, 72, 74, 79, 134, 136]. We remark that these are theoretical results that do not take into account the regularity of the functions in Kolmogorov's representation or includ any complexity considerations like in [67, 75, 76, 84].

Another practical result is due to Igelnik. Starting from a representation of the form (2.6), he introduced in [62, 64] so-called Kolmogorov spline networks to approximate continuously differentiable functions. There,  $2n + 1$  was replaced by a general larger  $m$ . The outer function  $g$  and the inner functions  $\psi_q$  were replaced by cubic splines  $s(\cdot, \gamma_q)$  and  $s(\cdot, \gamma_{q,p})$  where the parameters  $\gamma_q$  and  $\gamma_{q,p}$  of the splines were adjusted to fit given function values of  $f$  properly. The approximation scheme was defined as

$$f_m(x_1, \dots, x_n) = \sum_{q=1}^m s \left( \sum_{p=1}^n \lambda_p s(x_p, \gamma_{q,p}), \gamma_q \right). \quad (3.1)$$

It was shown that, for any function  $f$  from the class of continuously differentiable

functions on  $[0, 1]^n$  with bounded gradient, there exists a function  $f_m$  of the form (3.1) such that  $\|f - f_m\| = \mathcal{O}(1/m)$ . The number of degrees of freedom involved in the network is of the order  $\mathcal{O}(m^{3/2})$ . This result compares favorably with the approximation order  $\mathcal{O}(1/\sqrt{m})$  for general one-hidden layer feedforward networks and this class of functions what requires  $\mathcal{O}(m^2)$  degrees of freedom to achieve the same error, compare also [5, 64] and the references cited therein. However, the class of functions had to be restricted to achieve these rates and no experimental work was shown for the suggested model.

Furthermore, an important property of the inner functions in Kolmogorov's theorem was neglected in this approach. To be precise, they were not monotone increasing. This deficiency was eliminated by Coppejans in [17] where a similar approach, using cubic B-splines to approximate inner and outer functions, was presented. Here, necessary and sufficient conditions to impose monotonicity on cubic B-splines were formulated and included into the model. Again, to achieve convergence rates that are independent of the dimension  $n$ , the class of functions  $f$  had to be restricted. The approximated functions were assumed to be of the form

$$f(x_1, \dots, x_n) = \sum_{q=0}^m g_q \left( \sum_{p=1}^n \lambda_{p,q} \psi_q(x_p) \right) \quad \text{with} \quad g_q, \psi_q \in \mathcal{C}^3([0, 1]).$$

We have seen by Vitushkin's results [125, 126] that this is a notable restriction and that this class of functions is even smaller than  $\mathcal{C}^3([0, 1]^n)$ . Coppejans gave numerical examples for this approximator and compared it to kernel regressors for moderately large sample sizes of artificial data.

We presented approximation schemes for multi-dimensional functions that are directly based on different versions of Kolmogorov's superposition theorem, and that the constructions in the respective proofs can be seen as approximative versions of the Kolmogorov superposition theorem. Note that there exist further algorithms and methods in other contexts, like e.g. boolean functions, that are based on Kolmogorov's theorem or draw inspiration from it, see e.g. [6, 7, 9, 94] which are not listed here. We rather continue with a short note on well-known approximation schemes that are closely related to Kolmogorov's theorem. One example is the projection pursuit algorithm [34, 117] that approximates a function  $f$  by

$$f(x_1, \dots, x_n) \approx \sum_{q=1}^m g_q(\lambda_{q,1}x_1 + \dots + \lambda_{q,n}x_n). \quad (3.2)$$

This model directly results from (2.1) if we approximate the inner functions by  $\psi_{q,p}(t) \approx \lambda_{q,p} \cdot t$  and increase the number of terms.

Popular approximation schemes in statistics are the so-called additive models,

see [48, 49]. They resemble the approximation

$$f(x_1, \dots, x_n) \approx \sum_{q=1}^n f_q(x_q). \quad (3.3)$$

This form can be derived from (2.1) by choosing  $n$  instead of  $2n + 1$  and replacing the inner functions  $\psi_{q,p}$  trivially by the identity if  $q = p$  and zero otherwise.

For most of these approximation schemes the parameters are obtained by some kind of (least-squares) minimization. Here, however the objective functional may not be globally convex and can have many minima which results in non-unique representations. Thus, the associated approximation rates for these schemes for increasing  $m$  are not always fully understood and, moreover, it is not clear which representation to prefer over another for a particular application. We refer to [45] for more examples of related approximation schemes.

In summary, all approximation schemes presented in this section first use Kolmogorov's theorem to represent an  $n$ -dimensional function  $f$  as superposition of one-dimensional functions. Then, either both, inner and outer functions are approximated, or an approximation for the inner functions is fixed and approximations to the outer functions are computed. However, we have already seen that the one-dimensional functions are non-smooth. Therefore, we have to expect high complexities of the approximation schemes due to the non-smoothness of these functions. Additionally, the inner function also depends on the dimension  $n$  in some versions of the Kolmogorov superposition theorem. This brings up the question, how the complexities of the algorithms depend on the dimension  $n$ . This is where the curse of dimensionality still can be present.

### 3.2 Two models based on Sprecher's theorem

The aim of this thesis is now to define a model to reconstruct  $n$ -dimensional functions for moderate values of  $n$  that benefits from the constructive versions of the Kolmogorov superposition theorem introduced in Section 2.3. Namely, the parametric knowledge of the inner function  $\psi$  in (2.36) will be exploited. To this end, we keep its exact construction on the set of rational numbers, and introduce two ways to compute an approximation to the outer function  $\Phi$ . However, before doing so, we will discuss an important aspect that has only been addressed briefly yet.

In [87–89] a numerical implementation of Sprecher's algorithm from [111, 112] to compute the  $(m + 1)$  outer functions  $\Phi_q$  in the representation (2.9),

$$f(\mathbf{x}) = \sum_{q=0}^m \Phi_q \left( \sum_{p=1}^n \lambda_p \psi(x_p + qa) \right),$$

was discussed. This included estimates on the numerical complexity, i.e. the number of arithmetic operations, and run times. These investigations clearly revealed an

exponential dependency of the numerical complexity on the dimension  $n$  in each iteration step. Since our algorithm is an adaption of Sprecher's version, this also holds true for the present result with only a single outer function,

$$f(\mathbf{x}) = \sum_{q=0}^m \Phi \left( \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q \right),$$

see (2.36). The crucial point in Algorithm 2.1 that increases the costs, is the computation of the approximations  $\Phi^r$  to  $\Phi$ . To compute these functions, the constructions involve calculations for all points  $\mathbf{d}_{k_r} \in \mathcal{D}_{k_r}^n$ , which is equivalent to the construction on a regular grid, see step *II*. It is a known fact that for regular grids the number of grid points grows exponentially with their dimensionality and therefore the curse of dimensionality is still present, even though only one-dimensional functions are involved. Anyhow, we remark that the algorithm converges very fast in terms of the iterations  $r$ , see [73, 78, 87].

Due to the previous considerations, another approach to compute an approximation of the outer function  $\Phi$  will be introduced in this work. Note that the choice of representation (2.36) with only one inner and outer function is motivated by computational aspects. Here, we will only have to compute one function rather than  $(m+1)$  functions for representation (2.9).

At this point, we will introduce our two models only briefly for motivation. A further extension and a detailed investigation will then follow in Section 4.5.1 and Section 4.5.2.

### 3.2.1 First model

The first approach is very simple but gives insight into the structure of the outer function and is therefore presented here. Furthermore it motivates the more sophisticated and better second approach. To this end, we choose a countable basis  $\varphi_0^r, \varphi_1^r, \dots$  of  $\mathcal{C}^0([a_0, b_m])$ , see (2.53), (2.54) for a definition of  $a_0, b_m$ , and expand the outer function in this basis:

$$\Phi(t) = \sum_{j=0}^{\infty} c_j \varphi_j^r(t). \quad (3.4)$$

Expansion (3.4) still involves an infinite process and corresponds to the case of exact representation. We now can use only a finite number  $\ell^r$  of basis functions in (3.4).<sup>3</sup> This corresponds to the choice of a subspace  $V_{\ell^r} \subset \mathcal{C}^0([a_0, b_m])$  for the outer function

---

<sup>3</sup>The use of the superscript at  $\ell^r$  and  $\varphi_j^r$  indicates the use of real valued basis functions and is used to avoid confusion between the numbers  $\ell$  in Section 4.5, where complex valued basis functions will be introduced.

and results in an approximation. Note that the basis functions also depend on the values  $a_0$ ,  $b_m$ , and  $m$ , however we dispense with additional indices for simplification.

This leads, in view of (2.36), to the approximand

$$f_{\ell^r}(\mathbf{x}) := \sum_{q=0}^m \sum_{j=0}^{\ell^r} c_j \varphi_j^r \left( \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q \right), \quad (3.5)$$

with the coefficient vector  $\mathbf{c}_{\ell^r}^{(1)} := (c_0, \dots, c_{\ell^r})^T \in \mathbb{R}^{(\ell^r+1)}$ .

### 3.2.2 Second model

In the second approach we presume that the outer function  $\Phi(t)$  has a certain product structure. We then expand each factor  $\phi^d(t)$  in the given basis:

$$\Phi(t) = \prod_{d=0}^n \phi^d(t) = \prod_{d=0}^n \left( \sum_{j=0}^{\infty} c_j^d \varphi_j^r(t) \right). \quad (3.6)$$

Then, for each  $d = 0, \dots, n$ , we replace the inner infinite sum by a finite one and end up with an approximand of the form

$$f_{L^r}(\mathbf{x}) := \sum_{q=0}^m \prod_{d=0}^n \left( \sum_{j=0}^{\ell_d^r} c_j^d \varphi_j^{r,d} \left( \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q \right) \right). \quad (3.7)$$

Since the transfer from the infinite basis representation of each factor  $\phi^d$  to a finite one is explicitly allowed to differ for different values of  $d$ , the subscript for  $\ell_d^r$  and the superscript for the respective basis functions  $\varphi_j^{r,d}$  was added in the notation. This corresponds to the fact that we allow for  $d = 0, \dots, n$ , different approximation spaces  $V_{\ell_d^r}^d \subset \mathcal{C}^0([a_0, b_m])$  for the respective factor  $\phi^d$ . Here, we collect the model parameters in a coefficient vector  $\mathbf{c}_{L^r}^{(n)} := (c_0^0, \dots, c_{\ell_0^r}^0, \dots, c_0^n, \dots, c_{\ell_n^r}^n)^T \in \mathbb{R}^{L^r}$ , with  $L^r := \sum_{d=0}^n (\ell_d^r + 1)$ .

At this point, the choice of the second model is not obvious. Anyhow, an investigation of the first model will give substantial motivation. For further details and examples see Section 6.1.3 and Section 6.2.

# Chapter 4

## Regularization networks

We now turn to the computation of approximands  $f_{\ell,\mathbf{r}}$  and  $f_{L,\mathbf{r}}$ , see Section 4.5 for a definition, for a given continuous function  $f : [0, 1]^n \rightarrow \mathbb{R}$ . In many practical applications the function  $f$  is in general not given in analytical form but is accessible only in a point wise sampled fashion. Therefore we focus on reconstruction of  $f$  from given data points, i.e. value

$$\mathcal{Z} := \{(\mathbf{x}_j, y_j) : j = 1, \dots, P\} \subset [0, 1]^n \times \mathbb{R}. \quad (4.1)$$

The values  $y_j$  are assumed to be sample values  $y_j = f(\mathbf{x}_j)$  on the set  $\{\mathbf{x}_j\}_{j=1}^P$ .  $\mathcal{Z}$  will also be called **training set** or **sample** of the function  $f$ .

In this section, we will present a theoretical framework for the approximation of multi-dimensional functions from samples that essentially follows [18, 28]. The problem of learning (approximation of functions) from sparse data (samples) is an ill-posed<sup>1</sup> problem which is typically solved by regularization methods as developed by Tikhonov and Arsenin, see [118, 119]. There, the regression problem is restricted to the domain  $H$  of a positive functional  $\mathcal{R}(f)$ , which is called the **stabilizer**. The regularized problem is then formulated as the variational problem of minimizing the functional

$$E(f) = \frac{1}{P} \sum_{j=1}^P V(f(\mathbf{x}_j), y_j) + \nu \mathcal{R}(f).$$

$V$  is called **loss function** and enforces closeness to the data. The regularization term  $\mathcal{R}(f)$  enforces the solution to be in the set  $\{f : \mathcal{R}(f) \leq C\}$  with a small constant  $C$ . The positive value  $\nu \in \mathbb{R}$  is called **regularization parameter** and weights between the two terms.

The typical choice for the loss function is the squared error  $V(f(\mathbf{x}_i), y_i) = (f(\mathbf{x}_i) - y_i)^2$  which will also be used here. For the regularization term, a common strategy is to choose a regularization operator  $\mathcal{S} : H \rightarrow L_2(\Omega)$ , and set

---

<sup>1</sup>See Definition A.1 in the appendix.

$\mathcal{R}(f) = \|\mathcal{S}f\|_{L_2(\Omega)}^2$ . As an example,  $\mathcal{S}$  might be the gradient<sup>2</sup>  $\mathcal{S}f := \nabla f$ , which enforces smoothness of the function  $f$ . Hence,  $\mathcal{R}(f)$  is often called the **smoothness functional**.

An important example is  $\mathcal{R}(f) = \|f\|_{H(K)}^2$ , which is the norm in a, so called, reproducing kernel Hilbert space  $H(K)$ , induced by a symmetric positive function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$ , see Section 4.1 or [18, 28, 41, 42, 102, 124]. As we will see in more detail later, these spaces play an important role in the mathematical theory of learning from samples. For now, we only mention that a reproducing kernel Hilbert space  $H(K)$  with reproducing kernel  $K$  is directly related to a regularization operator  $\mathcal{S} : H(K) \rightarrow L_2(\Omega)$ , see [38, 41, 42, 95, 101]. In fact, if the reproducing kernel  $K$  is a Green's function of the operator  $(\mathcal{S}^* \mathcal{S})$ ,  $K$  and  $S$  are interrelated by

$$K(\mathbf{x}, \mathbf{y}) = \langle (\mathcal{S} K)(\mathbf{x}, \cdot), (\mathcal{S} K)(\mathbf{y}, \cdot) \rangle_{L_2(\Omega)}. \quad (4.2)$$

Thus, the formulation  $\mathcal{R}(f) = \|f\|_{H(K)}^2$  also includes the previously mentioned case via the choice of an appropriate kernel  $K$  and we may also consider functionals of the form

$$E(f) = \frac{1}{P} \sum_{j=1}^P (f(\mathbf{x}_j) - y_j)^2 + \nu \|f\|_{H(K)}^2. \quad (4.3)$$

Now, under certain assumptions on  $K$ , it can be shown that a minimizer  $f_K$  of  $E(f)$  has the form

$$f_K(\mathbf{x}) := \sum_{j=1}^P c_j K(\mathbf{x}, \mathbf{x}_j). \quad (4.4)$$

This fact is also called the **representer theorem**. Here, the coefficient vector  $\mathbf{c} := (c_1, \dots, c_P)^T \in \mathbb{R}^P$  is the solution of the linear system

$$(\mathbf{K} + \nu \mathbf{I}) \mathbf{c} = \mathbf{y},$$

with data vector  $\mathbf{y} := (y_1, \dots, y_P)^T \in \mathbb{R}^P$ , identity matrix  $\mathbf{I} \in \mathbb{R}^{P \times P}$  and the kernel matrix

$$\mathbf{K} := (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^P \in \mathbb{R}^{P \times P}.$$

For details, see [28, 41, 42, 127] and the references therein.

Girosi and Poggio [41, 95] gave a simple interpretation of the approximand  $f_K$  in terms of a network with one hidden layer of units. Therefore, it is called **Regularization Network**.

This data based approach provides us with a minimizer of (4.3) by the solution of a linear system for which the size is determined by the number  $P$  of given data points. This number can become very large. As we will see later, in Statistical

---

<sup>2</sup>To be more precise, one chooses in this case  $\mathcal{S} : H \rightarrow L_2^n(\Omega)$  and  $\mathcal{R}(f) = \|\mathcal{S}f\|_{L_2^n(\Omega)}^2$ .

learning theory it is even assumed that  $P \rightarrow \infty$ . Thus, the numerical costs for the solution of the system grow and for large data sets  $\mathcal{Z}$ , the computation of  $f_K$  becomes impractical. Note however, that the complexity of this data based approach only depends on the dimension  $n$  by means of  $P$  evaluations of the function  $K$ .

Alternatively, one might compute a representation of the minimizer of (4.3) in a given basis or frame for  $H(K)$ . This is the point of view which is taken in this thesis. The duality between these two approaches has been investigated in [38].

**Remark 4.1.** *Since by the representer theorem a minimization of (4.3) is equivalent to the use of a Regularization Network  $f_K$ , see (4.4) and [28], we will always refer to both strategies as **Regularization Network approach** and to both approximands, i.e. to the minimizer of (4.3) and to  $f_K$ , as **Regularization Networks**.*

In the following, the previously mentioned reproducing kernel Hilbert spaces are introduced in more detail.

## 4.1 Reproducing kernel Hilbert spaces

The theory that is presented in this section is standard theory on reproducing kernel Hilbert spaces that can be found e.g. in [4, 93, 100, 102, 127, 132]. However, since the results are a compilation of different definitions and theorems that is suited for our needs, they cannot be found exactly in the form that is presented here in the literature. Therefore, and for clarity, we briefly give the proofs for the respective results.

In the following, we always assume that  $\Omega$  is a general non-empty set. The results in this section do not need any further assumption on  $\Omega$  but note that in our application, usually  $\Omega \subset \mathbb{R}^n$ . We start with the definition of the central object of the theory.

**Definition 4.2.** *Let  $H$  be a complex Hilbert space of functions  $f : \Omega \rightarrow \mathbb{C}$  with scalar product  $\langle \cdot, \cdot \rangle_H$ . A function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  is called the **reproducing kernel** for  $H$  if*

- (i)  $K(\cdot, y) \in H$  for all  $y \in \Omega$ .
- (ii)  $f(y) = \langle f, K(\cdot, y) \rangle_H$  for all  $f \in H$  and all  $y \in \Omega$ .

The second property clearly motivates the notation and guarantees that a reproducing kernel is uniquely defined. Suppose there are two reproducing kernels  $K_1$  and  $K_2$ . Then, by this property we have that  $\langle f, K_1(\cdot, y) - K_2(\cdot, y) \rangle_H = 0$  for all  $f \in H$  and all  $y \in \Omega$ . Choosing  $f = K_1(\cdot, y) - K_2(\cdot, y) \in H$  for a fixed  $y \in \Omega$  shows uniqueness.

Next, we will show that continuity of the evaluation functionals  $E_y : H \rightarrow \mathbb{C}$ , which are defined for each  $y \in \Omega$  by

$$E_y(f) := f(y), \quad f \in H, \quad (4.5)$$

is a necessary and sufficient condition for a reproducing kernel  $K$  to exist.

**Theorem 4.3.** *Suppose that  $H$  is a Hilbert space of functions  $f : \Omega \rightarrow \mathbb{C}$ . Then the following statements are equivalent:*

- (i) *the point evaluation functionals  $E_y$  defined by (4.5) are continuous for all  $y \in \Omega$ ;*
- (ii)  *$H$  has a reproducing kernel.*

**Proof.** Assume (i) holds. Then, since  $E_y$  is bounded, we know from the Riesz representation theorem that there exists for any  $y \in H$  an element  $k_y \in H$  such that

$$E_y(f) = \langle f, k_y \rangle_H = f(y), \quad \text{for all } f \in H.$$

Thus,  $K(x, y) := k_y(x)$  is the reproducing kernel of  $H$ . Now, suppose (ii) holds, i.e. for  $H$  there exist a reproducing kernel  $K$ . This directly gives the estimate

$$\begin{aligned} |E_y(f)| &= |f(y)| = |\langle f, K(\cdot, y) \rangle_H| \\ &\leq \|f\|_H \|K(\cdot, y)\|_H = \|f\|_H \langle K(\cdot, y), K(\cdot, y) \rangle_H^{1/2} = \|f\|_H K(y, y)^{1/2}, \end{aligned}$$

what shows continuity of  $E_y$ .  $\square$

For the reproducing kernel  $K$  of a Hilbert space  $H$  it holds that

$$K(x, y) = \langle K(\cdot, y), K(\cdot, x) \rangle = \overline{\langle K(\cdot, x), K(\cdot, y) \rangle} = \overline{K(y, x)}.$$

In the following, it will be convenient to use the following terminology for such functions.

**Definition 4.4.** *A function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  with the property that  $K(x, y) = \overline{K(y, x)}$  is called a **symmetric function**.  $K$  is called  $d$ -dimensional if  $\Omega \subset \mathbb{R}^d$ .*

These preliminary considerations now motivate the following definition.

**Definition 4.5.** *Let  $H$  be the Hilbert space of complex valued functions  $f : \Omega \rightarrow \mathbb{C}$  and let  $\langle \cdot, \cdot \rangle_H$  denote its scalar product with the associated norm  $\|f\|_H^2 := \langle f, f \rangle_H$ . Then,  $H$  is called a **reproducing kernel Hilbert space (RKHS)** if for every  $y \in \Omega$  the linear evaluation functional  $E_y : H \rightarrow \mathbb{C}$ , defined by (4.5), is continuous.*

So far, we have seen that to any RKHS  $H$  there exists a unique reproducing kernel  $K : \Omega \times \Omega \rightarrow \mathbb{C}$ . In the sequel we will do the converse, i.e. to a given function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  with further properties that will be specified in the following, we will associate a Hilbert space  $H$ , such that  $K$  is the reproducing kernel in that space. To concretize the conditions on the function, we first give a further characterization of a reproducing kernel  $K$  of a RKHS  $H$ .

**Definition 4.6.** *Let  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  be a symmetric function of two variables. Then  $K$  is said to be a **positive definite** function provided that for every choice of distinct points  $\{x_1, \dots, x_p\} \subset \Omega$  and complex coefficients  $c_1, \dots, c_p \in \mathbb{C}$*

$$\sum_{i,j=1}^p \bar{c}_i c_j K(x_i, x_j) \geq 0,$$

*holds. Note that this is the condition for the matrix  $\{K(x_i, x_j)\}_{i,j=1}^{p,p}$  to be positive definite.  $K$  is called **strictly positive definite** if in addition  $\sum_{i,j=1}^p \bar{c}_i c_j K(x_i, x_j) = 0$  is equivalent to  $c_1 = \dots = c_p = 0$ .*

**Theorem 4.7.** *Suppose that  $H$  is a RKHS with reproducing kernel  $K : \Omega \times \Omega \rightarrow \mathbb{C}$ . Then  $K$  is positive definite. Moreover,  $K$  is strictly positive definite if the functions  $K(\cdot, y)$ ,  $y \in \Omega$  are linearly independent in  $H$ .*

**Proof.** *Let  $\{x_1, \dots, x_p\} \subset \Omega$ , and  $c_1, \dots, c_p \in \mathbb{C}$  be arbitrary. Then*

$$\sum_{i,j=1}^p \bar{c}_i c_j K(x_i, x_j) = \left\langle \sum_{j=1}^p c_j K(\cdot, x_j), \sum_{i=1}^p c_i K(\cdot, x_i) \right\rangle_H = \left\| \sum_{i=1}^p c_i K(\cdot, x_i) \right\|_H^2 \geq 0.$$

*The last expression can become zero for nonzero coefficients  $c_1, \dots, c_p \in \mathbb{C}$  if and only if the functions  $K(\cdot, y)$ ,  $y \in \Omega$  are linearly dependent in  $H$ .  $\square$*

Thus, the reproducing kernel of a given RKHS  $H$  is a symmetric positive definite function. Now, let a symmetric positive definite function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  be given. We turn to the construction of a Hilbert space for which this function is the reproducing kernel. We start with the following

**Theorem 4.8.** *Let a symmetric positive definite function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  be given. We define a linear vector space of functions by*

$$\begin{aligned} H^o &:= \text{span } \{K(\cdot, y) : y \in \Omega\} \\ &= \left\{ \sum_{i=1}^p a_i K(\cdot, y_i) : p \in \mathbb{N}, y_i \in \Omega, a_i \in \mathbb{C} \right\}. \end{aligned} \tag{4.6}$$

Then

$$\left\langle \sum_{i=1}^p a_i K(\cdot, x_i), \sum_{j=1}^{p'} b_j K(\cdot, y_j) \right\rangle_H := \sum_{i=1}^p \sum_{j=1}^{p'} a_i \bar{b}_j K(y_j, x_i) \quad (4.7)$$

defines an inner product on  $H^\circ$  and its completion with respect to this scalar product  $H = \overline{H^\circ}$ . Furthermore,  $H$  is the unique RKHS with kernel  $K$ .

**Proof.** From the fact that  $K$  is positive definite we can directly conclude that the bilinear form  $\langle \cdot, \cdot \rangle_H$  is well-defined and defines an inner product on  $H^\circ$ . Now, we complete the space  $H^\circ$ , by taking equivalence classes of Cauchy sequences in  $H^\circ$  and denote the resulting space by  $H$ . We have to show that every element of  $H$  is actually a function. To this end, let  $f \in H$  and  $\{f_n\} \subset H^\circ$  be a Cauchy sequence with limit  $f$ . Since

$$|f_n(x) - f_m(x)| = \langle f_n - f_m, K(\cdot, x) \rangle_H \leq K(x, x)^{1/2} \|f_n - f_m\|_H ,$$

we know that the sequence converges point wise and we can uniquely define  $f(x) := \lim_{n \rightarrow \infty} f_n(x)$ . Finally, in slight abuse of notation, we denote by  $\langle \cdot, \cdot \rangle_H$  the scalar product on  $H$  and compute

$$\langle f, K(\cdot, x) \rangle_H = \lim_{n \rightarrow \infty} \langle f_n, K(\cdot, x) \rangle_H = \lim_{n \rightarrow \infty} f_n(x) = f(x) ,$$

what shows that  $H$  is a RKHS with kernel  $K$ . Now, since the completion  $H$  of  $H^\circ$  is unique this shows Theorem 4.8.  $\square$

In conclusion, we have a one to one correspondence of a RKHS  $H$  over  $\Omega$  and a symmetric positive definite function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  which is the reproducing kernel for  $H$ . In the following, this interrelation will be reflected in the notation  $H(K)$  which will always denote the RKHS with reproducing kernel  $K$ .

We continue with summarizing some well known and useful results on reproducing kernels in a RKHS.

**Theorem 4.9.** Let  $H(K)$  be a RKHS with reproducing kernel  $K$ . If  $\{\varphi_1, \dots, \varphi_\ell\}$  is an orthonormal basis for  $H(K)$ , where  $\ell$  is possibly infinite, then

$$K(x, y) = \sum_{i=1}^{\ell} \varphi_i(x) \bar{\varphi}_i(y) , \quad (4.8)$$

and the series converges point wise.

**Proof.** For any  $y \in \Omega$  we have that  $K(\cdot, y) \in H(K)$  and thus

$$K(x, y) = \sum_{i=1}^{\ell} \langle K(\cdot, y), \varphi_i \rangle_{H(K)} \varphi_i(x) = \sum_{i=1}^{\ell} \bar{\varphi}_i(y) \varphi_i(x) ,$$

where the series converges in the norm. We have already seen that this directly implies point wise convergence.  $\square$

We now give simple examples on how new reproducing kernels can be constructed from existing ones. To this end, let first  $\varphi : \tilde{\Omega} \rightarrow \Omega$  be a function that maps the set  $\tilde{\Omega}$  onto  $\Omega$ . Then the following theorem holds:

**Theorem 4.10.** *Let  $\varphi : \tilde{\Omega} \rightarrow \Omega$  and let  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  be a symmetric positive definite function. Then  $(K \circ \varphi)(\tilde{x}, \tilde{y}) := K(\varphi(\tilde{x}), \varphi(\tilde{y}))$  is a symmetric positive definite function  $K \circ \varphi : \tilde{\Omega} \times \tilde{\Omega} \rightarrow \mathbb{C}$ .*

**Proof.** Let  $\{\tilde{x}_1, \dots, \tilde{x}_p\} \subset \tilde{\Omega}$  and  $c_1, \dots, c_p \in \mathbb{C}$  be arbitrary. We set  $\{x_1, \dots, x_s\} = \{\varphi(\tilde{x}_1), \dots, \varphi(\tilde{x}_s)\}$ ,  $A_k = \{i : \varphi(\tilde{x}_i) = x_k\}$ ,  $k = 1, \dots, s$ , and  $a_k = \sum_{i \in A_k} c_i$ . Then

$$\begin{aligned} \sum_{i,j=1}^p \bar{c}_i c_j K(\varphi(\tilde{x}_i), \varphi(\tilde{x}_j)) &= \sum_{k,l=1}^s \sum_{i \in A_k} \sum_{j \in A_l} \bar{c}_i c_j K(\varphi(\tilde{x}_i), \varphi(\tilde{x}_j)) \\ &= \sum_{k,l=1}^s \bar{a}_k a_l K(x_k, x_l) \geq 0, \end{aligned}$$

what shows that  $K \circ \varphi$  is positive definite. Symmetry can be seen directly from its definition.  $\square$

Note that the one to one correspondence of symmetric positive definite functions and RKHS's now uniquely defines the RKHS  $H(K \circ \varphi)$ . This space is also called the **pull-back** of  $H(K)$ . Another construction of reproducing kernels and thus of RKHS's is given by the following results.

Let  $K_i : \Omega \times \Omega \rightarrow \mathbb{C}$  be reproducing kernels of  $H(K_i)$ ,  $i \in \{1, 2\}$ , respectively. Then

- (i)  $K(x, y) = K_1(x, y) + K_2(x, y)$  is the reproducing kernel of the completion of

$$H(K_1 + K_2)^o = \{f_1 + f_2 : f_i \in H(K_i), i \in \{1, 2\}\}. \quad (4.9)$$

- (ii) Furthermore,  $K(x, y) = K_1(x, y) \cdot K_2(x, y)$  is the reproducing kernel of the completion of

$$H(K_1 \cdot K_2)^o = \left\{ \sum_{k=1}^p f_1^k f_2^k : f_i^k \in H(K_i), i \in \{1, 2\}, p \in \mathbb{N} \right\}. \quad (4.10)$$

- (iii) The function  $K(\mathbf{x}, \mathbf{y}) = K_1(x_1, y_1) \cdot K_2(x_2, y_2)$  is the reproducing kernel of the completion of

$$(H(K_1) \otimes H(K_2))^o = \left\{ \sum_{k=1}^p f_1^k(x_1) f_2^k(x_2) : f_i^k \in H(K_i), i \in \{1, 2\}, p \in \mathbb{N} \right\}. \quad (4.11)$$

In the first two cases, it is important to see that the elements of the spaces  $H(K_1)$ ,  $H(K_2)$ ,  $H(K_1 + K_2)^o$ , and especially  $H(K_1 \cdot K_2)^o$  are always functions  $f : \Omega \rightarrow \mathbb{C}$ , in contrast to  $(H(K_1) \otimes H(K_2))^o$ . The proofs for these facts can be found in [4, 93, 102].

In this section, we briefly introduced reproducing kernel Hilbert spaces and gave their characteristic properties. Here, we restricted ourselves to the basic facts and some results that will be needed in the remainder of this thesis. For a comprehensive theory of RKHS's and their applications, e.g. in machine learning or meshless methods, we refer to [4, 93, 100, 102, 127, 132].

Until now, the usage of the Regularization Network approach has been introduced and motivated by regularization theory which even justified its name. However, the reconstruction from finite data is not considered in classical regularization theory which is based on functional analysis arguments and relies on asymptotic results, [28]. In other words, the fact that the regularized solution has good predictive capabilities needs a probabilistic treatment that is not studied in regularization theory. This issue is considered in Statistical learning theory, and we will see in the following that there our formulation (4.3) with the smoothness operator  $\mathcal{R}(f) = \|f\|_{H(K)}^2$  also plays a central role.

## 4.2 Statistical learning theory

In this section we present the mathematical theory of Statistical learning. Basically, we follow the lines of [18].

For a finite set of training samples, the basic idea of Statistical learning theory is, as in regularization theory, to restrict the search for an optimizer to an appropriately small hypothesis space  $H$ : If  $H$  is large, one can find a model that fits the data well but has poor predictive capabilities on new data. This effect is called **overfitting** of the data. To avoid this, the size of  $H$  has to be controlled. This can be done in terms of the capacity of a set of functions, using capacity control which also considers the given samples. This leads to the technique of Structural risk minimization in Section 4.3. See [28] for an overview.

Statistical learning theory can be formulated in a very general setting and we will start with two sets  $X$  and  $Y$  whose elements are related by a probabilistic relationship in the following way: An element  $x \in X$  does not uniquely determine an element  $y \in Y$ , but rather a probability distribution on  $Y$ . This can be formalized by a probability distribution  $\rho$  which is defined over the set  $Z = X \times Y$  and governs

the sampling. The unknown distribution  $\rho$  is the primary object to study, but note that the goal is not to reveal  $\rho$ . We start with a formal definition of the objects that are needed in the following.

Let  $X$  be a compact domain or a manifold in Euclidian space and  $Y := \mathbb{R}$ . Let  $\rho$  be a Borel probability measure on  $Z := X \times Y$ . For a random variable  $\xi$ , we denote its expected value by  $\mathbf{E}[\xi]$  and its variance by  $\sigma^2(\xi)$ . The values are given by

$$\mathbf{E}[\xi] := \int_Z \xi d\rho \quad \text{and} \quad \sigma^2(\xi) := \mathbf{E}[(\xi - \mathbf{E}(\xi))^2] = \mathbf{E}[\xi^2] - \mathbf{E}[\xi]^2.$$

Now, if  $f : X \rightarrow Y$  is a model for the process that produces  $y \in Y$  from  $x \in X$ , the (least squares) error of  $f$  is defined by

$$\mathcal{E}(f) := \mathcal{E}_\rho(f) := \int_Z (f(x) - y)^2 d\rho.$$

For  $x \in X$ , let  $\rho(y|x)$  be the conditional probability measure on  $Y$  and  $\rho_X$  be the marginal probability measure on  $X$ . These probabilities are related to  $\rho$  by Fubini's theorem in the following way: For any integrable function  $\varphi : X \times Y \rightarrow \mathbb{R}$  it holds

$$\int_{X \times Y} \varphi(x, y) d\rho = \int_X \left( \int_Y \varphi(x, y) d\rho(y|x) \right) d\rho_X.$$

For  $\rho_X$ , we denote by  $\mathcal{L}_{\rho_X}^2(X)$  the Hilbert space of square integrable functions  $f : X \rightarrow \mathbb{R}$  with finite norm

$$\|f\|_{\rho_X}^2 := \int_X f^2(x) d\rho_X.$$

Next, we define the **regression function**  $f_\rho : X \rightarrow Y$  of  $\rho$  by

$$f_\rho(x) := \int_Y y d\rho(y|x).$$

It gives for each  $x \in X$  the average of the  $y$ -coordinate on the fiber  $\{x\} \times Y$ . Note that this is the expected value of  $y$  for given  $x$  with respect to the conditional probability measure  $\rho(y|x)$ .

For fixed  $x \in X$ , consider the function  $f_x(y) := (y - f_\rho(x))$ . Its expected value is zero what implies that its variance is given by

$$\sigma^2(f_x) := \int_Y (y - f_\rho(x))^2 d\rho(y|x).$$

Averaging over  $X$  we set

$$\sigma_\rho^2 := \int_X \sigma^2(f_x) d\rho_X = \int_X \left( \int_Y (y - f_\rho(x))^2 d\rho(y|x) \right) d\rho_X = \mathcal{E}(f_\rho).$$

Note that in general  $\rho$  and  $f_\rho$  are unknown, while in some situations  $\rho_X$  is known. An example would be that  $\rho_X$  is the Lebesgue measure on  $X$  inherited from Euclidian space.

As a consequence of the previous computations, the error  $\mathcal{E}(f)$  can be split into two parts:

$$\mathcal{E}(f) = \int_X (f(x) - f_\rho(x))^2 d\rho_X + \sigma_\rho^2. \quad (4.12)$$

The first term in the right-hand side of (4.12) is the average of the error that will be made when using  $f$  as a model for  $f_\rho$ . The second term  $\sigma_\rho^2$  is independent of  $f$ , what implies that  $f_\rho$  has the smallest possible error among all functions  $f : X \rightarrow Y$ . Thus,  $\sigma_\rho^2$  is a lower bound on the error  $\mathcal{E}(f)$  which is completely determined by the probability measure  $\rho$ .

The goal is now to learn (i.e. find a good approximation of)  $f_\rho$  from random samples on  $Z$ . We assume that the approximand  $f$  belongs to a compact subset  $H \subset (\mathcal{C}^0(X), \|\cdot\|_\infty)$ , which will be called the **hypothesis space**. The choice of compact subsets with respect to the norm  $\|\cdot\|_\infty$  is important, e.g. to guarantee existence of  $f_H$  and  $f_z$  (see below) but also for other assumptions and results. We do not go into detail here, but refer to the literature [18] and the references therein.

Let  $f_H$  be a minimizer of  $\mathcal{E}(f)$ , i.e.

$$f_H := \arg \min_{f \in H} \int_Z (f(x) - y)^2 d\rho.$$

From (4.12) we see that for this function it also holds

$$f_H = \arg \min_{f \in H} \int_X (f(x) - f_\rho(x))^2 d\rho_X.$$

Note that this minimizer is not necessarily unique unless  $H$  is convex.

Now, assume that we are provided with a sample

$$\mathbf{z} := ((x_1, y_1), \dots, (x_P, y_P)) \subset X \times Y, \quad (4.13)$$

that is drawn according to  $\rho$ . Note that the data (4.1) fits into this setting since  $[0, 1]^n$  is a compact subset of Euclidian space and we could also take  $\mathbf{z} = \mathcal{Z}$ . However, we stick to the new notation to be consistent with the general setting in this section. The set  $\mathbf{z}$  is called **training data**. We define the **empirical error** of  $f$  to be

$$\mathcal{E}_z(f) := \frac{1}{P} \sum_{i=1}^P (f(x_i) - y_i)^2, \quad (4.14)$$

and denote by  $f_{H,z}$  the minimizer of (4.14) in  $H$ . The function

$$f_z := f_{H,z} := \arg \min_{f \in H} \mathcal{E}_z(f)$$

is called the **empirical target function**. We remark that  $f_z$  does not depend on  $\rho$  and that  $\mathcal{E}(f_z)$  and  $\mathcal{E}_z(f)$  are different objects. The same holds for  $\mathcal{E}(f_H)$  and  $\mathcal{E}_H(f)$  (see below).

For a given hypothesis space  $H$ , and  $f \in H$ , the *error in  $H$*  is defined by the normalized value

$$\mathcal{E}_H(f) := \mathcal{E}(f) - \mathcal{E}(f_H) = \int_Z (f(x) - y)^2 d\rho - \int_Z (f_H(x) - y)^2 d\rho \geq 0.$$

With this quantity we can split the error of  $f_z$  into two parts:

$$\mathcal{E}(f_z) = \int_X (f_z - f_\rho)^2 d\rho_X + \sigma_\rho^2 = \mathcal{E}_H(f_z) + \mathcal{E}(f_H). \quad (4.15)$$

Next, we consider the right hand side of this equation.

Remember that our primary goal is to estimate the predictive capabilities of  $f_z$  which is given by  $\mathcal{E}(f_z)$ , or equivalently by  $\int_X (f_z - f_\rho)^2 d\rho_X$ . Equation (4.15) now shows that this breaks down into two problems:

- (i) Find an estimate for  $\mathcal{E}(f_H)$ . This term is independent of the sample but depends on the choice of the hypothesis space  $H$ . It will be called the **approximation error**.
- (ii) Find an estimate for the term  $\mathcal{E}_H(f_z)$ . This problem is posed on  $H$  and depends on the sample.  $\mathcal{E}_H(f_z)$  is called the **sampling error** or **estimation error**.

In statistics, the following relation between (i) and (ii) is called the **bias-variance trade-off**:

- (bias) For a fixed sample size, typically, the approximation error in (i) will decrease when enlarging  $H$ , but the sample error in (ii) will increase.
- (variance) For fixed  $H$ , the sample error in (ii) decreases when the number  $P$  of samples increases.

Next, we present results from [18] that give estimates for the approximation and sampling error in probability.

**Remark 4.11.** *We stick to statements from [18] that fit into the following general setting: Let  $\mathbb{E}$  be a subset of a Banach space of functions  $f : X \rightarrow Y$  for which the embedding  $J_{\mathbb{E}} : (\mathbb{E}, \|\cdot\|_{\mathbb{E}}) \hookrightarrow (\mathcal{C}^0(X), \|\cdot\|_{\infty})$  is compact. We then define, for  $R > 0$ , the hypothesis space*

$$H = H_{\mathbb{E}, R} = \overline{J_{\mathbb{E}}(B_R)}, \quad (4.16)$$

where  $B_R$  denotes the closed ball of radius  $R$  in  $\mathbb{E}$ ,

$$B_R := \overline{\{f \in \mathbb{E} : \|f\|_{\mathbb{E}} \leq R\}}.$$

The first question is now the following: Let  $\varepsilon, \delta > 0$  be given. How many samples have to be drawn to assert, with a confidence greater than  $1-\delta$ , that  $\int_X (f_{\mathbf{z}} - f_H)^2 d\rho_X$  is smaller than  $\varepsilon$ ? To give a quantitative answer to this question we need a further definition.

**Definition 4.12.** *Let  $S$  be a metric space and  $s > 0$ . We define the **covering number**  $\mathcal{N}(S, s)$  to be the minimal  $k \in \mathbb{N}$ , such that there exist  $k$  discs in  $S$  with radius  $s$  covering  $S$ .*

Note that the discs are defined with respect to the maximum norm  $\|\cdot\|_\infty$  and that for compact sets  $S$ , the covering number is finite. In the following we will always denote the size of the sample  $\mathbf{z}$  in (4.13) by  $P \in \mathbb{N}$ .

**Theorem 4.13.** *Let  $H$  be a compact subset of  $\mathcal{C}^0(X)$ . Assume that, for all  $f \in H$ ,  $|f(x) - y| \leq M$  almost everywhere. Let*

$$\sigma^2(H) = \sup_{f \in H} \sigma^2(f_Y^2),$$

where  $\sigma^2(f_Y^2)$  is the variance of  $f_Y^2(x, y) := (f(x) - y)^2$ . Then for all  $\varepsilon > 0$ ,

$$\text{Prob}_{\mathbf{z} \in Z^P} \{ \mathcal{E}_H(f_{\mathbf{z}}) \leq \varepsilon \} \geq 1 - 2 \mathcal{N}\left(H, \frac{\varepsilon}{16M}\right) \exp\left(-\frac{P\varepsilon^2}{8(4\sigma^2(H) + \frac{1}{3}M^2\varepsilon)}\right).$$

This theorem shows that for an increasing sample size, the sample error decreases. It directly gives the following corollary which answers the question on the size of the sample to ensure a desired accuracy.

**Corollary 4.14.** *Let the assumptions of Theorem 4.13 hold and  $\varepsilon, \delta > 0$  be given. If*

$$P \geq \frac{8(4\sigma^2(H) + \frac{1}{3}M^2\varepsilon)}{\varepsilon^2} \left[ \ln\left(2\mathcal{N}\left(H, \frac{\varepsilon}{16M}\right)\right) + \ln\left(\frac{1}{\delta}\right) \right],$$

then it holds that

$$\text{Prob}_{\mathbf{z} \in Z^P} \{ \mathcal{E}_H(f_{\mathbf{z}}) \leq \varepsilon \} \geq 1 - \delta.$$

The proofs of the previous results can be found in [18]. There, also estimates on the covering numbers  $\mathcal{N}(S, s)$  for Sobolev spaces or RKHS's with reproducing kernel  $K \in \mathcal{C}^\infty$  are given. We do not state these results here, since the reproducing kernels for our approach do not fulfill this constraint.

However, for a result on the approximation error  $\mathcal{E}(f_H)$  the restriction  $K \in \mathcal{C}^\infty$  is not necessary. Therefore, we can give the following theorem from [18] directly in

terms of a RKHS:

**Theorem 4.15.** *Let  $K : X \times X \rightarrow \mathbb{R}$  be a continuous, symmetric, positive definite function,  $R > 0$ , and  $H = \overline{J_{\mathbb{E}}(B_R)}$  as defined in Remark 4.11 with  $\mathbb{E} = H(K)$ . Then, the approximation error satisfies, for  $0 < r < 1$ ,*

$$\mathcal{E}(f_H) \leq \left(\frac{1}{R}\right)^{\frac{2r}{1-r}} \|L_K^{-r/2} f_\rho\|_{\rho_X}^{\frac{2}{1-r}} + \sigma_\rho^2.$$

Here, the operator  $L_K : \mathcal{L}_{\rho_X}^2(X) \rightarrow \mathcal{C}(X)$  is defined by

$$(L_K f)(x) := \int_X K(x, \cdot) f(\cdot) d\rho_X.$$

Now, consider the bias–variance problem. To this end, remember that to each value  $R > 0$  we associated the hypothesis space  $H_{\mathbb{E}, R}$  by (4.16). Furthermore, fix the sample size  $P$  and choose a confidence  $1 - \delta$ ,  $0 < \delta < 1$ . The bias–variance problem consists of finding an optimal value  $R^* > 0$  which minimizes the error bound for  $\mathcal{E}(f_z)$  from (4.15) with confidence  $1 - \delta$ . The following result can be found in [18].

**Theorem 4.16.** *For all  $P \in \mathbb{N}$ ,  $\delta \in \mathbb{R}$ ,  $0 < \delta < 1$ , and all  $0 < r < 1$ , there exists a unique solution  $R^*$  of the bias–variance problem in the general setting, described in Remark 4.11.*

The minimizer  $R^*$  determines a hypothesis space  $H_{\mathbb{E}, R^*}$  in the family of all spaces  $H_{\mathbb{E}, R}$ ,  $R > 0$ . Of course, one may as well consider the variation of other parameters like the kernel  $K$  for  $H(K)$ . This is also called the *kernel trick*, see [102]. In learning literature the choice of the hypothesis space is called the selection of a model.

Practically, the computation of  $R^*$  is often not possible. Here, one reason might be the lack of estimates for the covering numbers  $\mathcal{N}(H_{\mathbb{E}, R}, \varepsilon)$  of the hypothesis spaces  $H_{\mathbb{E}, R}$ ,  $R > 0$ .

As an alternative approach to deal with the bias–variance trade–off, the method of Structured risk minimization was introduced in [28, 29].

## 4.3 Structural risk minimization

In this section, we briefly introduce the method of Structural risk minimization [23, 28, 29, 102, 122–124] which then will be used as a motivation for our Regularization Network approach.

To this end, let a training set be given by (4.13) that consists of  $P$  samples. Furthermore, let  $h \in \mathbb{N}$  be a measure for the complexity of a hypothesis space  $H$ , which will be called the capacity of  $H$ . For instance,  $H$  could be the set of

polynomials of degree  $h$ , or a set of splines with  $h$  nodes. Then, in the first step of Structural risk minimization (SRM), a nested sequence of hypothesis spaces  $H_1 \subset H_2 \subset \dots \subset H_{\ell(P)}$  is defined, such that the capacities of the spaces form a non-decreasing sequence  $h_1 \leq h_2 \leq \dots h_{\ell(P)} < \infty$ . Here,  $\ell(P)$  is a non-decreasing integer function of  $P$ .

In [28], two different types of capacity for SRM have been suggested: The  $VC$ -dimension and the  $V_\gamma$ -dimension. In fact, the SRM-method uses the  $VC$ -dimension as a measure of complexity, while the use of the  $V_\gamma$ -dimension gives rise to the, so called, *extended* SRM-method, see [28, 29]. Since the SRM-method is only of theoretical interest here, and will only be used as a motivation for Regularization Networks, we dispense with a detailed definition of the quantities in this section and refer to Appendix A.4. For further details see [23, 28, 29, 102, 122–124].

In the next step of SRM, compute for  $i = 1 \dots, \ell(P)$  the minimizer of the empirical error (4.14) in the space  $H_i$ , which is given by

$$f_{H_i, \mathbf{z}} := \arg \min_{f \in H_i} \frac{1}{P} \sum_{j=1}^P (f(x_j) - y_j)^2.$$

Now, if  $h_i$  is the capacity of a hypothesis space  $H_i$ , estimates in probability of the following type can be given [28]:

$$\text{Prob} \left\{ \sup_{f \in H_i} |\mathcal{E}(f) - \mathcal{E}_{\mathbf{z}}(f)| > \varepsilon \right\} \leq \mathcal{G}(\varepsilon, P, h_i),$$

where  $\mathcal{G}(\varepsilon, P, h_i)$  is an increasing function of  $h_i$ . Note that  $\mathcal{G}$  depends on the type of capacity that is used. Thus, for the sequence of hypothesis spaces  $H_i$ , this bound becomes weaker which allows for possibly worse predictive capabilities of the minimizer  $f_{H_i, \mathbf{z}}$ , while the data is better approximated. This is again the bias-variance trade-off as introduced in Section 4.2.

Finally, the last step of the SRM-method consists of finding the structure  $H_{\text{opt}}$  for which the trade-off between the empirical error  $\mathcal{E}_{\mathbf{z}}(f_{H_{\text{opt}}, \mathbf{z}})$  and  $\mathcal{G}(\varepsilon, P, h_{\text{opt}})$  is optimal. Note that  $\text{opt} = \text{opt}(\varepsilon, P)$ . To find the optimal structure, in [28] the following strategy is suggested: For each  $i = 1, \dots, \ell(P)$ , find numerically the “effective”  $\varepsilon_i$  so that  $\mathcal{G}(\varepsilon_i, P, h_i)$  is constant for all  $H_i$ , and then choose the function  $f_{H_i, \mathbf{z}}$  for which the sum of the empirical error and  $\varepsilon_i$  is minimized. Note that the strategy to handle this trade-off also depends on the respective type of capacity that is used. Under further assumptions it can then be shown that the error  $\mathcal{E}(f_{H_{\text{opt}}, \mathbf{z}})$  of the SRM-solution approaches for  $P \rightarrow \infty$  and  $\ell(P) \rightarrow \infty$  in probability to the optimal error  $\mathcal{E}(f_{H, \mathbf{z}})$  in  $H = \bigcup_{j=1}^{\infty} H_j$ .

In practice, the SRM-method is difficult to implement due to the requirement of solving of a large, in principle infinite, number of optimization problems with non-linear constraints. Additionally, it is difficult to decide the optimal trade-off

strategy between empirical error and  $\mathcal{G}$ . In the following, we rather use SRM as a motivation for the Regularization Network approach (4.3).

To this end, consider a RKHS  $H(K)$ , real numbers  $R_1 \leq R_2 \dots \leq R_{\ell(P)} < \infty$ , and the hypothesis spaces

$$H_i := \{f \in H(K) : \|f\|_{H(K)} \leq R_i\}.$$

Remember that these are the hypothesis spaces in the Regularization Network approach, see Remark 4.1. Note also that Theorem 4.16 and (4.16) theoretically guarantee that for such spaces there exists an optimal  $R^*$ , solving the bias–variance problem. But in general we cannot expect  $R^* = R_i$ , for any  $i = 1, \dots, \ell(P)$ . It can be shown [27, 28] that the capacities<sup>3</sup>  $h_i$  of the spaces  $H_i$  form a non-decreasing sequence  $h_1 \leq h_2 \leq \dots \leq h_{\ell(P)} < \infty$ , and thus can be used as hypothesis spaces in the SRM–method. Therewith, we have to solve for each  $i = 1, \dots, \ell(P)$  the constrained minimization problem

$$\frac{1}{P} \sum_{j=1}^P (f(x_j) - y_j)^2 \longrightarrow \min !, \quad \text{subject to} \quad \|f\|_{H(K)} \leq R_i.$$

This leads, for  $i = 1 \dots, \ell(P)$ , to a minimization of the functional

$$E_i(f) := \frac{1}{P} \sum_{j=1}^P (f(x_j) - y_j)^2 + \nu (\|f\|_{H(K)}^2 - R_i^2)$$

with respect to  $f$  while maximizing  $E_i$  with respect to the Lagrange multiplier  $\nu \geq 0$ . For  $i = 1 \dots, \ell(P)$ , let  $(f_i, \nu_i)$  denote the minimizer of the respective functional  $E_i$ . From these solutions, the SRM–method picks the optimal function  $f_{\text{opt}} \in H_{\text{opt}}$ ,  $\text{opt} \in \{1, \dots, \ell(P)\}$ , which is now associated with an optimal Lagrange multiplier  $\nu_{\text{opt}}$ . Clearly, a minimization of

$$\frac{1}{P} \sum_{j=1}^P (f(x_j) - y_j)^2 + \nu_{\text{opt}} \|f\|_{H(K)}^2,$$

using the optimal Lagrange multiplier  $\nu_{\text{opt}}$ , produces the same solution  $f_{\text{opt}}$  as the SRM–method. Thus, SRM leads to the error–functional (4.3) which is used in the Regularization Network approach and additionally provides us with an estimate of the optimal regularization parameter. Note again that a numerical implementation of SRM is impractical.

In the next section, a probabilistic justification of Regularization Networks will be given by a maximum a posteriori interpretation of the given learning problem.

---

<sup>3</sup>Here, essentially the  $V_\gamma$ –dimension has to be used. The VC–dimension is independent of  $R_i$ , [27, 28].

## 4.4 Maximum A Posteriori Interpretation

The variational principle of minimizing (4.3) can also be derived in a probabilistic context, see [28] and the references therein. This will be the topic of this section whereas we follow the lines of [28].

To this end, let  $\mathbf{z} \in Z^P$  be the given sample from (4.13). We denote by

- $\text{Prob}(f|\mathbf{z})$  the conditional probability of the function  $f$  given the sample  $\mathbf{z}$ .
- $\text{Prob}(\mathbf{z}|f)$  is the conditional probability that by a random sampling of the underlying function  $f$  at the points  $\{x_1, \dots, x_P\}$  we get  $\{y_1, \dots, y_P\}$ . This is a model of the noise.
- $\text{Prob}(f)$  denotes some *a priori* knowledge of the function  $f$  which can be used to impose constraints on the model. This can be done by assigning significant probability only to those functions that satisfy those constraints.

If the probability distribution for the noise  $\text{Prob}(\mathbf{z}|f)$  and the constraints  $\text{Prob}(f)$  are known, we can compute the posterior distribution  $\text{Prob}(f|\mathbf{z})$  by the Bayes rule, which is

$$\text{Prob}(f|\mathbf{z}) \propto \text{Prob}(\mathbf{z}|f) \text{Prob}(f).$$

Let us assume that  $\text{Prob}(f)$  is a multivariate Gaussian with zero mean<sup>4</sup> in the Hilbert space  $H(K)$ . Then, see [28] and Appendix A.1 for details, the prior has the form

$$\text{Prob}(f) \propto \exp(-\|f\|_{H(K)}^2).$$

Furthermore, we assume that the noise is normally distributed with variance  $\sigma$ . Then, its probability is

$$\text{Prob}(\mathbf{z}|f) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^P (f(x_j) - y_j)^2\right),$$

and using Bayes formula we get for the a posteriori probability of  $f$  with the given sample  $\mathbf{z}$  that

$$\text{Prob}(f|\mathbf{z}) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^P (f(x_j) - y_j)^2 - \|f\|_{H(K)}^2\right).$$

Now, amongst others, the *Maximum A Posteriori* (MAP) estimate provides an opportunity to estimate  $f$  from the given distribution. To maximize the a posteriori

---

<sup>4</sup> Let  $X(t)$ ,  $t \in \Omega$  be Gaussian random variables with zero mean and covariance  $\mathbf{E}[X(s)X(t)] = K(s,t)$ . Then, there exists a one to one correspondence between the RKHS  $H(K)$  and the Hilbert space  $\mathcal{X}$  which is spanned the  $X(t)$ ,  $t \in \Omega$ , see [28, 127] and Appendix A.1.

probability  $\text{Prob}(f|\mathbf{z})$ , one can equivalently minimize

$$\frac{1}{P} \sum_{j=1}^P (f(x_j) - y_j)^2 + \nu_* \|f\|_{H(K)}^2, \quad \text{where } \nu_* = 2\sigma^2/P,$$

with respect to  $f$ . Thus, the MAP interpretation of learning from samples leads, under certain assumptions, to a minimization of the error–functional (4.3) for a given regularization parameter  $\nu$ .

To sum up, we presented two different motivations, namely Structural risk minimization and the Maximum A Posteriori interpretation, for the use of the Regularization Networks approach in the form of (4.3). They both provided us with an estimate of the regularization parameter  $\nu$ . However, for the SRM–method a computation of  $\nu_{\text{opt}}$  is inefficient and for  $\nu_*$  from the MAP interpretation it was argued in [28] that it does not provide a good estimate for the optimal regularization parameter. To compute  $\nu_{\text{opt}}$  and  $\nu_*$ , further assumptions were made during SRM and the MAP interpretation. Therefore it is important to note that they can only be seen as estimates for the optimal regularization parameter in the Regularization Network approach. In [19], further strategies for the choice of  $\nu$  and estimates on the approximation errors are given which are based on statistical computations. More practical techniques to estimate  $\nu$  are cross-validation, [50, 127], or the L–curve, [46, 47].

To summarize, we substantiated the Regularization Network approach, i.e. a minimization of (4.3), in the RKHS setting, and observed that special care has to be taken to the choice of the regularization parameter.

The next step is now to concretize the RKHS setting of the Regularization Network approach for extensions of the models  $f_{\ell r}$  and  $f_{Lr}$  which have been introduced in (3.5) and (3.7), respectively. This will be the topic of Section 4.5.

## 4.5 Approximation spaces and norms

In Section 3.2.1 and Section 3.2.2 we briefly introduced two models  $f_{\ell r}$  and  $f_{Lr}$  for the approximation of an  $n$ –dimensional function  $f : [0, 1] \rightarrow \mathbb{R}$ . We will now identify the underlying approximation spaces for an extension of these models to the complex valued case as reproducing kernel Hilbert spaces and compute the according norms that are used for regularization in the Regularization Network approach. To do this, we first define a reproducing kernel  $K : [0, 1]^n \times [0, 1]^n \rightarrow \mathbb{C}$  and thus a RKHS  $\mathcal{H}(K)$  of  $n$ –dimensional functions which contains both extended models. In a second step two subsets  $\mathcal{H}_L^{(1)}$  and  $\mathcal{K}_L$  of this space are identified which describe the extended models  $f_\ell$  and  $f_L$ , respectively.

This finally enables us to formulate the Regularization Networks which are based on the extended models in Section 4.5.1 and Section 4.5.2. To do this, the error functionals  $E^{(1)}$  and  $E^{(n)}$  will be derived from (4.3) and necessary conditions for their minimizers will be given.

The following computations are based on the statements in Section 4.1. In this notation, the general set  $\Omega$  will now be the real interval  $\Omega := [a_0, b_m] \subset \mathbb{R}$ , where the values  $a_0$  and  $b_m$  are defined by (2.53), and (2.54) respectively. For any Hilbert space with scalar product  $\langle \cdot, \cdot \rangle$  we can define a norm by  $\|\phi\| := \langle \phi, \phi \rangle^{1/2}$ . Furthermore, we denote by  $\mathcal{C}^0(\Omega, \mathbb{C})$  the set of continuous functions  $\phi : \Omega \rightarrow \mathbb{C}$ , for which we will use the usual definition of the maximum norm

$$\|\phi\|_\infty := \sup_{s \in \Omega} |\phi(s)|.$$

Note that in the remainder of this thesis it is important to distinguish between one-dimensional and  $n$ -dimensional functions. To make this easier, we will denote Hilbert spaces which contain functions  $\phi : \mathbb{R} \rightarrow \mathbb{C}$  with  $\mathcal{H}$ , while Hilbert spaces containing functions  $f : \mathbb{R}^n \rightarrow \mathbb{C}$  are denoted with  $\mathcal{H}$ . For the respective one-dimensional and  $n$ -dimensional kernels we use lower and upper case letters.

Now, let a family of Hilbert spaces with associated scalar products and norms

$$\mathcal{H}^d \subset \mathcal{C}^0(\Omega, \mathbb{C}), \quad \langle \cdot, \cdot \rangle_{\mathcal{H}^d}, \quad \|\cdot\|_{\mathcal{H}^d}, \quad d = 0, \dots, n \quad (4.17)$$

be given. Furthermore, let for  $d = 0, \dots, n$ ,

$$\mathcal{F}^d := \{\varphi_0^d, \varphi_1^d, \dots\} \subset \mathcal{H}^d \quad (4.18)$$

be a basis of  $\mathcal{H}^d$ , and assume that for  $d = 0, \dots, n$  the following properties hold:

- (P1) The embeddings  $(\mathcal{H}^d, \|\cdot\|_{\mathcal{H}^d}) \hookrightarrow (\mathcal{C}^0(\Omega, \mathbb{C}), \|\cdot\|_\infty)$  are continuous.
- (P2) The elements of  $\mathcal{F}^d$  are orthogonal with respect to the inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}^d}$ , i.e. it holds

$$\langle \varphi_i^d, \varphi_j^d \rangle_{\mathcal{H}^d} = \gamma_j^d \delta_{i,j}, \quad \text{with} \quad \gamma_j^d \in \mathbb{R}. \quad (4.19)$$

Furthermore, we assume that without loss of generality  $\gamma_0^d := 1$ . Here,  $\delta_{i,j}$  denotes Kronecker's delta function, which is defined by  $\delta_{i,j} := 1$ , if  $i = j$ , and  $\delta_{i,j} := 0$  else. Note that we do not claim that the bases  $\mathcal{F}^0, \dots, \mathcal{F}^n$  are finite.

- (P3) For any  $\phi \in \mathcal{H}^d$  it holds that  $\bar{\phi} \in \mathcal{H}^d$ .
- (P4) The constant function  $\mathbf{1} : \Omega \rightarrow \mathbb{C}$ , defined by  $\mathbf{1}(s) := 1$ , is included in the space, i.e.  $\mathbf{1} \in \mathcal{H}^d$ .

To show that each  $\mathcal{H}^d$  is a reproducing kernel Hilbert space (RKHS), we have to establish continuity of the evaluation functionals  $E_t : \mathcal{H}^d \rightarrow \mathbb{C}$  with respect to the norm  $\|\cdot\|_{\mathcal{H}^d}$ , see (4.5) and Definition 4.5. Indeed, since the embedding  $(\mathcal{H}^d, \|\cdot\|_{\mathcal{H}^d}) \hookrightarrow (\mathcal{C}^0(\Omega, \mathbb{C}), \|\cdot\|_\infty)$  is continuous we can estimate for  $0 \neq \phi \in \mathcal{H}^d \subset \mathcal{C}^0(\Omega, \mathbb{C})$ :

$$\frac{|E_t(\phi)|}{\|\phi\|_{\mathcal{H}^d}} = \frac{|\phi(t)|}{\|\phi\|_{\mathcal{H}^d}} \lesssim \frac{|\phi(t)|}{\|\phi\|_\infty} = \frac{|E_t(\phi)|}{\|\phi\|_\infty} \leq \sup_{\substack{\phi \in \mathcal{C}^0(\Omega, \mathbb{C}) \\ \phi \neq 0}} \frac{|E_t(\phi)|}{\|\phi\|_\infty} \leq C_\infty.$$

The last estimate is simply the fact that the operator norm of  $E_t$  with respect to the maximum norm is bounded on  $\mathcal{C}^0(\Omega, \mathbb{C})$  by some constant  $C_\infty < \infty$ . This shows that the point evaluation functional  $E_t : \mathcal{H}^d \rightarrow \mathbb{C}$  is also continuous with respect to the norm  $\|\cdot\|_{\mathcal{H}^d}$ , and therefore that  $\mathcal{H}^d$  is a RKHS. Let two elements  $\phi_1, \phi_2 \in \mathcal{H}^d$  be given by

$$\phi_1(s) := \sum_{i=0}^{\infty} a_i \varphi_i^d(s), \quad \text{and} \quad \phi_2(s) := \sum_{i=0}^{\infty} b_i \varphi_i^d(s), \quad a_i, b_i \in \mathbb{C}.$$

With (4.19), it is easy to see that the scalar product in  $\mathcal{H}^d$  has the form

$$\langle \phi_1, \phi_2 \rangle_{\mathcal{H}^d} = \sum_{i=0}^{\infty} \gamma_i^d a_i \bar{b}_i, \quad (4.20)$$

and Theorem 4.9 then shows that the reproducing kernel  $k^d : \Omega \times \Omega \rightarrow \mathbb{C}$  for  $\mathcal{H}^d$  has the representation

$$k^d(s, t) = \sum_{i=0}^{\infty} \frac{1}{\gamma_i^d} \varphi_i^d(s) \overline{\varphi_i^d(t)}. \quad (4.21)$$

Thus,  $\mathcal{H}^0(k^0), \dots, \mathcal{H}^n(k^n)$  are RKHS's over the real interval  $[a_0, b_m]$ , i.e. their elements are functions  $\phi : [a_0, b_m] \rightarrow \mathbb{C}$ , and the sets  $\mathcal{F}^d$  form their bases.

Multiplying the reproducing kernels (4.21) for  $d = 0, \dots, n$ , now gives the product function

$$k^{(n)}(s, t) := k^0(s, t) \cdots k^n(s, t), \quad (4.22)$$

which is known to be the reproducing kernel of some RKHS  $\mathcal{H}(k^{(n)})$ , see (4.10). It is important to note that this is not a tensor product construction, but still a function  $k^{(n)} : \Omega \times \Omega \rightarrow \mathbb{C}$ . With (4.21) and (4.22) the product function has the expansion

$$k^{(n)}(s, t) = \sum_{i_0, \dots, i_n=0}^{\infty} \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \left( \prod_{d=0}^n \varphi_{i_d}^d(s) \right) \left( \prod_{d=0}^n \overline{\varphi}_{i_d}^d(t) \right). \quad (4.23)$$

To identify the RKHS for which  $k^{(n)}$  is the reproducing kernel, we next show that the linear space

$$\begin{aligned} \mathcal{H}(k^{(n)})^o &:= \text{span} \left\{ k^{(n)}(\cdot, t) : t \in \Omega \right\} \\ &= \left\{ \sum_{i=1}^p a_i k^{(n)}(s, t_i) : p \in \mathbb{N}, t_i \in \Omega, a_i \in \mathbb{C} \right\} \end{aligned}$$

is also spanned by the products  $\prod_{d=0}^n \varphi_{i_d}^d(s)$ ,  $i_d \in \mathbb{N}_0$ . Indeed, for an element  $\phi \in \mathcal{H}(k^{(n)})$  it holds with (4.23) that

$$\begin{aligned}\phi(s) &= \sum_{i=1}^p a_i k^{(n)}(s, t_i) \\ &= \sum_{i=1}^p a_i \sum_{i_0, \dots, i_n=0}^{\infty} \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \left( \prod_{d=0}^n \varphi_{i_d}^d(s) \right) \left( \prod_{d=0}^n \overline{\varphi}_{i_d}^d(t_i) \right) \\ &= \sum_{i_0, \dots, i_n=0}^{\infty} \mathbf{a}_{i_0, \dots, i_n} \left( \prod_{d=0}^n \varphi_{i_d}^d(s) \right),\end{aligned}\quad (4.24)$$

with the complex coefficients

$$\mathbf{a}_{i_0, \dots, i_n} := \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \sum_{i=1}^p a_i \left( \prod_{d=0}^n \overline{\varphi}_{i_d}^d(t_i) \right), \quad i_d \in \mathbb{N}_0. \quad (4.25)$$

Due to this fact, and with Theorem 4.8, the set

$$\left\{ f(s) = \sum_{i_0, \dots, i_n=0}^{\infty} \mathbf{a}_{i_0, \dots, i_n} \left( \prod_{d=0}^n \varphi_{i_d}^d(s) \right) : \mathbf{a}_{i_0, \dots, i_n} \in \mathbb{C}, \|f\|_{\mathcal{H}(k^{(n)})}^2 < \infty \right\} \quad (4.26)$$

contains  $\mathcal{H}(k^{(n)})^\circ$  and thus it is also dense in  $\mathcal{H}(k^{(n)})$ . Furthermore, the assumption that  $\mathbf{1} \in \mathcal{H}^d$ , for all  $d = 0, \dots, n$ , directly implies

$$\mathcal{H}^0(k^0) \cup \dots \cup \mathcal{H}^n(k^n) \subset \mathcal{H}(k^{(n)}) \subset \mathcal{C}^0([a_0, b_m], \mathbb{C}).$$

From this multiplicative extension to a RKHS of one-dimensional functions  $\phi : [a_0, b_m] \rightarrow \mathbb{C}$ , we next construct a RKHS of  $n$ -dimensional functions. Typically, this is done by a tensor product construction of one-dimensional reproducing kernels to construct an  $n$ -dimensional reproducing kernel, see (4.11). Here, we pursue another strategy and use Kolmogorov's superposition theorem instead.

Returning to Kolmogorov's representation (2.36) of  $n$ -dimensional continuous functions by

$$f(\mathbf{x}) = \sum_{q=0}^m \phi_f \circ \Psi_q(\mathbf{x}), \quad \text{with fixed} \quad \Psi_q(\mathbf{x}) = \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q,$$

we see that it provides us for any  $f : [0, 1]^n \rightarrow \mathbb{C}$  with a function  $\phi_f : [a_0, b_m] \rightarrow \mathbb{C}$ , which depends on  $f$ . Note that for a complex valued function, one can simply apply Kolmogorov's theorem separately to its real and imaginary part. Conversely,

it directly gives us, for any continuous one-dimensional function  $\phi : [a_0, b_m] \rightarrow \mathbb{C}$ , a continuous  $n$ -dimensional function  $f_\phi : [0, 1]^n \rightarrow \mathbb{C}$  by

$$f_\phi(\mathbf{x}) := \sum_{q=0}^m \phi \circ \Psi_q(\mathbf{x}),$$

which now depends on  $\phi$ . This fact will be used to construct an  $n$ -dimensional function  $K : [0, 1]^n \times [0, 1]^n \rightarrow \mathbb{C}$  and a RKHS  $\mathcal{H}(K)$  for which  $K$  is the reproducing kernel.

To this end, define for  $d = 0, \dots, n$  and  $i_d \in \mathbb{N}_0$  the functions

$$\Phi_{i_0, \dots, i_n}(\mathbf{x}) := \sum_{q=0}^m \varphi_{i_0, \dots, i_n} \circ \Psi_q(\mathbf{x}), \quad \text{where} \quad \varphi_{i_0, \dots, i_n}(s) := \prod_{d=0}^n \varphi_{i_d}^d(s) \quad (4.27)$$

are the products spanning the set (4.26), and the  $n$ -dimensional symmetric function

$$K(\mathbf{x}, \mathbf{y}) := \sum_{i_0, \dots, i_n=0}^{\infty} \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \Phi_{i_0, \dots, i_n}(\mathbf{x}) \overline{\Phi}_{i_0, \dots, i_n}(\mathbf{y}). \quad (4.28)$$

Here, the values  $\gamma_{i_d}^d \in \mathbb{R}$ ,  $d = 0, \dots, n$ ,  $i_d \in \mathbb{N}_0$ , are still given by (4.19). The fact that this series converges follows from the relation between  $K(\mathbf{x}, \mathbf{y})$  and  $k^{(n)}(s, t)$  which is given by

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}) &= \sum_{i_0, \dots, i_n=0}^{\infty} \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \left( \sum_{q=0}^m \prod_{d=0}^n \varphi_{i_d}^d(\Psi_q(\mathbf{x})) \right) \left( \sum_{q=0}^m \prod_{d=0}^n \overline{\varphi}_{i_d}^d(\Psi_q(\mathbf{y})) \right) \\ &= \sum_{i_0, \dots, i_n=0}^{\infty} \sum_{q_1, q_2=0}^m \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \left( \prod_{d=0}^n \varphi_{i_d}^d(\Psi_{q_1}(\mathbf{x})) \right) \left( \prod_{d=0}^n \overline{\varphi}_{i_d}^d(\Psi_{q_2}(\mathbf{y})) \right) \\ &= \sum_{q_1, q_2=0}^m \sum_{i_0, \dots, i_n=0}^{\infty} \prod_{d=0}^n \frac{1}{\gamma_{i_d}^d} \varphi_{i_d}^d(\Psi_{q_1}(\mathbf{x})) \overline{\varphi}_{i_d}^d(\Psi_{q_2}(\mathbf{y})) \\ &= \sum_{q_1, q_2=0}^m \prod_{d=0}^n \left( \sum_{i_d=0}^{\infty} \frac{1}{\gamma_{i_d}^d} \varphi_{i_d}^d(\Psi_{q_1}(\mathbf{x})) \overline{\varphi}_{i_d}^d(\Psi_{q_2}(\mathbf{y})) \right) \\ &= \sum_{q_1, q_2=0}^m \prod_{d=0}^n k^d(\Psi_{q_1}(\mathbf{x}), \Psi_{q_2}(\mathbf{y})) = \sum_{q_1, q_2=0}^m k^{(n)}(\Psi_{q_1}(\mathbf{x}), \Psi_{q_2}(\mathbf{y})). \end{aligned} \quad (4.29)$$

This representation will also enable us to show that  $K(\mathbf{x}, \mathbf{y})$  is a symmetric positive definite function, see Definition 4.6. To prove this, we have to show that for an arbitrary set of points  $\mathcal{X} := \{\mathbf{x}_1, \dots, \mathbf{x}_p\} \subset [0, 1]^n$ , the matrix  $\mathbf{K}_{\mathcal{X}} := \{K(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^p$

is positive definite. Let  $\mathcal{X}$  be given as above. We define the points in a further set

$$\mathcal{Y} := \{y_{0,1}, \dots, y_{0,p}, \dots, y_{m,1}, \dots, y_{m,p}\} \subset \mathbb{R}, \quad \text{by} \quad y_{q,i} := \Psi_q(\mathbf{x}_i).$$

With this set we define the block-matrix

$$\mathbf{K}_{\mathcal{Y}} := \left\{ \mathbf{K}_{q_1, q_2} \right\}_{q_1, q_2=0}^m \quad \text{with the blocks} \quad \mathbf{K}_{q_1, q_2} := \left\{ k^{(n)}(y_{q_1, i}, y_{q_2, j}) \right\}_{i,j=1}^p.$$

Now, let  $\mathbf{c} \in \mathbb{R}^p$  be an arbitrary vector. Since  $k^{(n)}$  is the reproducing kernel of  $\mathcal{H}(k^{(n)})$ , Theorem 4.7 guarantees that  $\mathbf{K}_{\mathcal{Y}}$  is positive definite. For this reason, it holds for  $\mathbf{d} := (\mathbf{c}^T, \dots, \mathbf{c}^T)^T \in \mathbb{R}^{(m+1)p}$  that

$$\mathbf{c}^T \mathbf{K}_{\mathcal{X}} \mathbf{c} = \mathbf{c}^T \left( \sum_{q_1, q_2=0}^m \mathbf{K}_{q_1, q_2} \right) \mathbf{c} = \mathbf{d}^T \mathbf{K}_{\mathcal{Y}} \mathbf{d} \geq 0.$$

This shows that  $K$  is positive definite.

Furthermore, Theorem 4.8 gives the existence of a unique RKHS, denoted by  $\mathcal{H}(K)$ , for which  $K$  is the reproducing kernel.

Note the similarity of this construction to the result from Theorem 4.10 and the remark following it. This motivates an interpretation of  $\mathcal{H}(K)$  as the *generalized pull-back* of  $\mathcal{H}(k^{(n)})$ . However, due to the sums in (4.29), Theorem 4.10 cannot be applied directly to  $K$ .

Starting from the linear space

$$\mathcal{H}(K)^o := \left\{ \sum_{i=1}^p a_i K(\cdot, \mathbf{y}_i) : p \in \mathbb{N}, \mathbf{y}_i \in [0, 1]^n, a_i \in \mathbb{C} \right\},$$

we now investigate the RKHS  $\mathcal{H}(K)$  and its scalar product  $\langle \cdot, \cdot \rangle_{\mathcal{H}(K)}$ .

First, analogue computations to the one-dimensional case show that any  $f \in \mathcal{H}(K)^o$  can be expanded in terms of  $\Phi_{i_0, \dots, i_n}$ :

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^p a_i K(\mathbf{x}, \mathbf{y}_i) = \sum_{i=1}^p a_i \sum_{i_0, \dots, i_n=0}^{\infty} \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \Phi_{i_0, \dots, i_n}(\mathbf{x}) \overline{\Phi}_{i_0, \dots, i_n}(\mathbf{y}_i) \\ &= \sum_{i_0, \dots, i_n=0}^{\infty} \mathbf{a}_{i_0, \dots, i_n} \Phi_{i_0, \dots, i_n}(\mathbf{x}), \end{aligned} \tag{4.30}$$

where

$$\mathbf{a}_{i_0, \dots, i_n} := \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \sum_{i=1}^p a_i \overline{\Phi}_{i_0, \dots, i_n}(\mathbf{y}_i) \in \mathbb{C}. \tag{4.31}$$

The similar argument as for (4.24) shows, together with Theorem 4.8, that  $\mathcal{H}(K)$  is the completion of

$$\left\{ f(\mathbf{x}) = \sum_{i_0, \dots, i_n=0}^{\infty} \mathbf{a}_{i_0, \dots, i_n} \Phi_{i_0, \dots, i_n}(\mathbf{x}) : \mathbf{a}_{i_0, \dots, i_n} \in \mathbb{C}, \|f\|_{\mathcal{H}(K)}^2 < \infty \right\}. \quad (4.32)$$

Now, let  $f$  from this set be given. We compute

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i_0, \dots, i_n=0}^{\infty} \mathbf{a}_{i_0, \dots, i_n} \Phi_{i_0, \dots, i_n}(\mathbf{x}) = \sum_{i_0, \dots, i_n=0}^{\infty} \mathbf{a}_{i_0, \dots, i_n} \sum_{q=0}^m \varphi_{i_0, \dots, i_n} \circ \Psi_q(\mathbf{x}) \\ &= \sum_{q=0}^m \left( \sum_{i_0, \dots, i_n=0}^{\infty} \mathbf{a}_{i_0, \dots, i_n} \varphi_{i_0, \dots, i_n} \right) \circ \Psi_q(\mathbf{x}) =: \sum_{q=0}^m \phi_f \circ \Psi_q(\mathbf{x}), \end{aligned} \quad (4.33)$$

and find that  $\phi_f$  is contained in the set (4.26) which is dense in  $\mathcal{H}(k^{(n)})$ , see (4.27). Clearly, the converse also holds true, i.e. for  $\phi$  from (4.26) we have that  $f_\phi := \sum_{q=0}^m \phi \circ \Psi_q$  lies in (4.32). Consequently, this also holds for the respective closures, i.e. we can replace (4.26) with  $\mathcal{H}(k^{(n)})$  and (4.32) with  $\mathcal{H}(K)$  in this argumentation.

Now, let  $f, g \in \mathcal{H}(K)^\circ$  be given by

$$f(\mathbf{x}) := \sum_{j=1}^p a_j K(\mathbf{x}, \mathbf{y}_j) \quad \text{and} \quad g(\mathbf{x}) := \sum_{j=1}^s b_j K(\mathbf{x}, \tilde{\mathbf{y}}_j),$$

where  $p, s \in \mathbb{N}$ , and  $a_j, b_j \in \mathbb{C}$ ,  $\mathbf{y}_j, \tilde{\mathbf{y}}_j \in [0, 1]^n$ . Then, the coefficients  $\mathbf{a}_{i_0, \dots, i_n}$ ,  $\mathbf{b}_{i_0, \dots, i_n} \in \mathbb{C}$ ,  $d = 0, \dots, n$ ,  $i_d \in \mathbb{N}_0$ , are given by (4.31) respectively, and  $\phi_f, \phi_g \in \mathcal{H}(k^{(n)})$  can be derived from  $f, g$  as in (4.33). From (4.7) and (4.33), we get that the scalar product of  $f$  and  $g$  is given by

$$\begin{aligned} \langle f, g \rangle_{\mathcal{H}(K)} &= \sum_{j_1=1}^p \sum_{j_2=1}^s a_{j_1} \bar{b}_{j_2} K(\tilde{\mathbf{y}}_{j_2}, \mathbf{y}_{j_1}) \\ &= \sum_{j_1=1}^p \sum_{j_2=1}^s a_{j_1} \bar{b}_{j_2} \sum_{i_0, \dots, i_n=0}^{\infty} \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \Phi_{i_0, \dots, i_n}(\tilde{\mathbf{y}}_{j_2}) \overline{\Phi}_{i_0, \dots, i_n}(\mathbf{y}_{j_1}) \\ &= \sum_{i_0, \dots, i_n=0}^{\infty} \frac{1}{\gamma_{i_0}^0 \cdots \gamma_{i_n}^n} \left( \sum_{j_1=1}^p a_{j_1} \overline{\Phi}_{i_0, \dots, i_n}(\mathbf{y}_{j_1}) \right) \overline{\left( \sum_{j_2=1}^s b_{j_2} \overline{\Phi}_{i_0, \dots, i_n}(\tilde{\mathbf{y}}_{j_2}) \right)} \\ &= \sum_{i_0, \dots, i_n=0}^{\infty} \gamma_{i_0}^0 \cdots \gamma_{i_n}^n \mathbf{a}_{i_0, \dots, i_n} \overline{\mathbf{b}}_{i_0, \dots, i_n} = \langle \phi_f, \phi_g \rangle_{\mathcal{H}(k^{(n)})}. \end{aligned} \quad (4.34)$$

Here, the last equation can be realized in analogue manner, using (4.7), (4.23), and (4.25) for  $\langle \cdot, \cdot \rangle_{\mathcal{H}(k^{(n)})}$ . Now, since the scalar product is continuous, (4.34) also

holds for functions in  $\mathcal{H}(K)$ . Thus, the scalar product of  $n$ -dimensional functions  $f, g \in \mathcal{H}(K)$  is simply the scalar product of the one-dimensional functions  $\phi_f, \phi_g \in \mathcal{H}(k^{(n)})$ , resulting from (4.33). Note that  $\|f\|_{\mathcal{H}(K)}^2 < \infty$  imposes a decay condition on the values  $|\mathbf{a}_{i_0, \dots, i_n}|^2$ .

So far, we have defined a RKHS  $\mathcal{H}(K)$  of  $n$ -dimensional functions by means of a reproducing kernel  $K : [0, 1]^n \times [0, 1]^n \rightarrow \mathbb{C}$ , and identified its associated scalar product.

Next, we turn back to the models for approximation that have been introduced in Section 3.2.1 and Section 3.2.2. Remember that there we used the superscript  $r$  to indicate the use of real valued basis functions. In fact, we defined

$$f_\ell(\mathbf{x}) = \sum_{q=0}^m \phi_\ell \circ \Psi_q(\mathbf{x}), \quad \text{with} \quad \phi_\ell := \sum_{j=0}^\ell c_j \varphi_j, \quad (4.35)$$

$\ell = \ell^r < \infty$ ,  $\varphi_j = \varphi_j^r$ , and  $c_j \in \mathbb{R}$ , for  $j = 0, \dots, \ell^r$ , see (3.5). The second model was given by

$$f_L(\mathbf{x}) = \sum_{q=0}^m \left( \prod_{d=0}^n \phi_{\ell_d}^d \right) \circ \Psi_q(\mathbf{x}), \quad \text{with} \quad \phi_{\ell_d}^d = \sum_{j=0}^{\ell_d} c_j^d \varphi_j^d, \quad (4.36)$$

$\ell_d = \ell_d^r < \infty$ ,  $L := \sum_{d=0}^n (\ell_d + 1) = L^r$ ,  $\varphi_j^d = \varphi_j^{r,d}$ , and  $c_j^d \in \mathbb{R}$ , for  $d = 0, \dots, n$ ,  $j = 0, \dots, \ell_d^r$ , see (3.7).

**Remark 4.17.** Note that these models have been defined, using coefficients  $c_j, c_j^d \in \mathbb{R}$ , and basis functions  $\varphi_j^r, \varphi_j^{r,d} \in \mathcal{C}^0([a_0, b_m])$ , which is the set of continuous real valued functions, while the investigations in this section deal with complex valued functions from  $\mathcal{C}^0([a_0, b_m], \mathbb{C})$ . However, for our computations it is advantageous to first extend the models  $f_{\ell^r}$  and  $f_{L^r}$  to the complex valued case. Then, from property (P3) of the spaces  $\mathcal{H}^0, \dots, \mathcal{H}^n$ , we can conclude that for any  $f \in \mathcal{H}(K)$ , it holds that its complex conjugate  $\bar{f} \in \mathcal{H}(K)$ , and thus that the restriction of  $f$  to its real part

$$f_r(\mathbf{x}) := \operatorname{Re}(f(\mathbf{x})) := \frac{f(\mathbf{x}) + \bar{f}(\mathbf{x})}{2} \quad (4.37)$$

also belongs to the space  $\mathcal{H}(K)$ . The same holds for the imaginary part  $f_i := \operatorname{Im}(f) := (f - i\bar{f})/2 \in \mathcal{H}(K)$  of the function  $f$ . Note that  $f_r$  and  $f_i$  are real valued functions and  $f = f_r + i f_i$ .

To extend the models (3.5), (3.7) and define finite dimensional approximation spaces, we assume in the following that in (4.36) the basis functions are given by finite subsets

$$\mathcal{F}_{\ell_d}^d = \{\varphi_0^d, \dots, \varphi_{\ell_d}^d\} \subset \mathcal{F}^d, \quad d = 0, \dots, n, \quad (4.38)$$

of (4.18), and in (4.35) we set  $\varphi_i := \varphi_i^0$ , for  $i = 0, \dots, \ell := \ell_0$ . Furthermore, all coefficients are now complex numbers  $c_j, c_j^d \in \mathbb{C}$ . This directly gives the subspaces

$$\mathcal{H}_{\ell_d}^d := \text{span}\{\varphi_0^d, \dots, \varphi_{\ell_d}^d\} \subset \mathcal{H}^d(k^d), \quad d = 0, \dots, n. \quad (4.39)$$

Additionally, we claim that the subsets  $\mathcal{F}_{\ell_d}^0, \dots, \mathcal{F}_{\ell_n}^n$  in (4.38) are chosen such that the properties (P1)–(P4) still hold for the spaces  $\mathcal{H}_{\ell_0}^0, \dots, \mathcal{H}_{\ell_n}^n$ . Now, since all previous calculations for the respective RKHS's are also valid for finite sums<sup>5</sup>, this leads to subspaces

$$\mathcal{H}_L^{(n)} := \text{span}\left\{\prod_{d=0}^n \varphi_{i_d}^d : i_d \in \{0, \dots, \ell_d\}\right\} \subset \mathcal{H}(k^{(n)}), \quad (4.40)$$

and

$$\mathcal{H}_L := \text{span}\left\{\Phi_{i_0, \dots, i_n} : i_d \in \{0, \dots, \ell_d\}\right\} \subset \mathcal{H}(K). \quad (4.41)$$

Obviously, for the outer functions in the extended models it holds

$$\phi_\ell \in \mathcal{H}_L^{(n)}, \quad \prod_{d=0}^n \phi_{\ell_d}^d \in \mathcal{H}_L^{(n)}, \quad \text{and thus} \quad f_\ell, f_L \in \mathcal{H}_L \subset \mathcal{H}(K).$$

However, the space  $\mathcal{H}_L$  is much richer in the sense that it contains functions that do not coincide with our extended models. To be more concrete in our description of the underlying approximation sets, we will next construct two subsets of  $\mathcal{H}_L$  which only contain (complex valued) functions of the form  $f_\ell$  or  $f_L$ , respectively.

### 4.5.1 Regularization Network approach for the first model

For the first (extended) model  $f_\ell$ , which is given by (4.36), this construction is straightforward. We simply take  $\ell_1 = \dots = \ell_n = 0$ , and since we always claim that  $\mathbf{1} \in \mathcal{H}_{\ell_d}^d$ , this implies for  $d = 1, \dots, n$ , that the subsets in (4.38) have to be chosen as  $\mathcal{F}^d = \{\varphi_0^d := C^d\}$ , where  $C^d \in \mathbb{C}$  is a constant such that  $\langle \varphi_0^d, \varphi_0^d \rangle = 1$ . The subspace of  $\mathcal{H}(K)$  which is defined by (4.41) for these sets will be denoted by  $\mathcal{H}_\ell^{(1)}$ . It exactly describes the first model, i.e.  $f_\ell \in \mathcal{H}_\ell^{(1)}$ , and conversely all functions  $f \in \mathcal{H}_\ell^{(1)}$ , given by

$$f(\mathbf{x}) = \sum_{i=0}^{\ell} c_i \Phi_{i,0,\dots,0}(\mathbf{x}) = \sum_{q=0}^m \left( \sum_{i=0}^{\ell} c_i \varphi_i \right) \circ \Psi_q(\mathbf{x}) =: \sum_{q=0}^m \phi_\ell \circ \Psi_q(\mathbf{x}),$$

---

<sup>5</sup>Replace in Section 4.5 all sums  $\sum_{i_d=0}^\infty$  with  $\sum_{i_d=0}^{\ell_d}$  and all  $\sum_{i_0, \dots, i_n=0}^\infty$  with  $\sum_{i_0, \dots, i_n=0}^{\ell_0, \dots, \ell_n}$ .

have the same form as  $f_\ell$ . Consequently, with  $\langle \varphi_0^d, \varphi_0^d \rangle_{\mathcal{H}^d(k^d)} = \gamma_0^d = 1$ ,  $d = 1, \dots, n$ , the norm of  $f_\ell \in \mathcal{H}_\ell^{(1)}$  can be computed from

$$\|f_\ell\|_{\mathcal{H}(K)}^2 = \sum_{i=0}^{\ell} \gamma_i^0 c_i^2 = \left\| \sum_{i=0}^{\ell} c_i \varphi_i \right\|_{\mathcal{H}^0(k^0)}^2. \quad (4.42)$$

Now, remember that the task is to approximate functions  $f : [0, 1]^n \rightarrow \mathbb{R}$  from their function values  $f(\mathbf{x}_j) = y_j \in \mathbb{R}$ , at locations  $\mathbf{x}_j \in [0, 1]^n$ ,  $j = 1, \dots, P$ . Finally, following Remark 4.17, we replace  $f_\ell$  with the real valued function  $f_{\ell,\mathfrak{r}}$ , which is defined by (4.37), and observe that the Regularization Network approach for the first model consists of the following

**Problem 4.18.** *For a given data set  $\mathcal{Z} \subset [0, 1]^n \times \mathbb{R}$ , see (4.1), and a given regularization parameter  $0 < \nu \in \mathbb{R}$ , find a minimum  $f_{\ell,\mathfrak{r}} : [0, 1]^n \rightarrow \mathbb{R}$  of*

$$E^{(1)}(f_{\ell,\mathfrak{r}}) := \frac{1}{P} \sum_{j=1}^P (f_{\ell,\mathfrak{r}}(\mathbf{x}_j) - y_j)^2 + \nu \|f_{\ell,\mathfrak{r}}\|_{\mathcal{H}(K)}^2, \quad (4.43)$$

with  $f_{\ell,\mathfrak{r}} = (f_\ell + \bar{f}_\ell^{(1)})/2$  and  $f_\ell \in \mathcal{H}_\ell^{(1)}$ , see (4.37).

Note that we are looking for a minimizer  $f_{\ell,\mathfrak{r}} \in \mathcal{H}(K)$ , which is enforced by the norm, and claim that  $f_\ell \in \mathcal{H}_\ell^{(1)}$ , which defines our model. Next, we will show that this complex valued formulation can be directly reinterpreted as a convex minimization problem in  $\mathbb{R}^{2(\ell+1)}$ . To this end, let  $\mathbf{c}_{\ell,\mathbb{C}}^{(1)} := (c_0, \dots, c_\ell)^T \in \mathbb{C}^{(\ell+1)}$  denote the complex coefficient vector of  $f_\ell$ . We split up all complex values  $c_j$  and basis functions  $\varphi_j$ ,  $j = 0, \dots, \ell$ , into their real and imaginary parts by

$$c_j = c_{j,\mathfrak{r}} + i c_{j,\mathfrak{i}}, \quad \text{and} \quad \varphi_j(s) = \varphi_{j,\mathfrak{r}}(s) + i \varphi_{j,\mathfrak{i}}(s).$$

Furthermore, let

$$\tilde{\varphi}_{j,\mathfrak{r}}(s) := \varphi_{j,\mathfrak{r}}(s), \quad \tilde{\varphi}_{j,\mathfrak{i}}(s) := -\varphi_{j,\mathfrak{i}}(s), \quad j = 0, \dots, \ell.$$

With these real valued basis functions,  $f_{\ell,\mathfrak{r}} : [0, 1]^n \rightarrow \mathbb{R}$  can be represented by

$$f_{\ell,\mathfrak{r}}(\mathbf{x}) = \sum_{q=0}^m \phi_{\ell,\mathfrak{r}} \circ \Psi_q(\mathbf{x}) = \sum_{q=0}^m \left( \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{j=0}^{\ell} c_{j,e} \tilde{\varphi}_{j,e} \right) \circ \Psi_q(\mathbf{x}). \quad (4.44)$$

For more details on the calculations in this section we refer to Appendix A.2.

We remark that with  $\ell^r := 2\ell$ , the function  $f_{\ell,\mathfrak{r}}$  has a similar structure as in (3.5). This is due to the fact that  $\mathcal{H}_\ell^{(1)}$  is a linear space what implies  $f_{\ell,\mathfrak{r}} \in \mathcal{H}_\ell^{(1)}$ .

Next, we define for  $j, k = 0, \dots, \ell$ ,  $e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}$  the values

$$S_{j,k}^{e_1,e_2} := \frac{1}{2} \begin{cases} \delta_{j,k} \gamma_k + \operatorname{Re}(\langle \varphi_j, \bar{\varphi}_k \rangle_{\mathcal{H}^0(k^0)}), & e_1 = e_2 = \mathfrak{r}, \\ \delta_{j,k} \gamma_k - \operatorname{Re}(\langle \varphi_j, \bar{\varphi}_k \rangle_{\mathcal{H}^0(k^0)}), & e_1 = e_2 = \mathfrak{i}, \\ -\operatorname{Im}(\langle \varphi_j, \bar{\varphi}_k \rangle_{\mathcal{H}^0(k^0)}), & e_1 \neq e_2. \end{cases}$$

Note that  $S_{j,k}^{e_1,e_2} = S_{k,j}^{e_2,e_1}$ , and the regularization term is given by

$$\|f_{\ell,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 = \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k,l=0}^{\ell} c_{k,e_1} c_{l,e_2} S_{k,l}^{e_1,e_2}.$$

Thus, minimizing (4.43) is equivalent to finding a minimum of

$$\begin{aligned} E^{(1)}(f_{\ell,\mathfrak{r}}) &= E^{(1)}(\mathbf{c}_{\ell}^{(1)}) \\ &= \frac{1}{P} \sum_{j=1}^P \left( \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k=0}^{\ell} c_{k,e} \sum_{q=0}^m \tilde{\varphi}_{k,e}(\Psi_q(\mathbf{x}_j)) - y_j \right)^2 + \nu \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k,l=0}^{\ell} c_{k,e_1} c_{l,e_2} S_{k,l}^{e_1,e_2} \end{aligned} \quad (4.45)$$

over all real vectors  $\mathbf{c}_{\ell}^{(1)} := (c_{0,\mathfrak{r}}, c_{0,\mathfrak{i}}, \dots, c_{\ell,\mathfrak{r}}, c_{\ell,\mathfrak{i}})^T \in \mathbb{R}^{2(\ell+1)}$ . This reinterpretation as real valued problem shows that  $E^{(1)}$  is also convex with respect to the real coefficient vector  $\mathbf{c}_{\ell}^{(1)} \in \mathbb{R}^{2(\ell+1)}$ .

Therefore, a minimizer of  $E^{(1)}(\mathbf{c}_{\ell}^{(1)})$  is uniquely characterized by the necessary and sufficient conditions

$$\frac{\partial}{\partial c_{\mu,\theta}} E^{(1)}(\mathbf{c}_{\ell}^{(1)}) = 0, \quad \text{for all } \mu = 0, \dots, \ell, \quad \theta \in \{\mathfrak{r}, \mathfrak{i}\}.$$

Further computations then show that the minimizer is given by the solution of the following system of linear equations:

$$(\mathbf{B}^T \mathbf{B} + \nu P \mathbf{S}) \mathbf{c}_{\ell}^{(1)} = \mathbf{B}^T \mathbf{y}. \quad (4.46)$$

Here, the data matrix  $\mathbf{B} \in \mathbb{R}^{P \times 2(\ell+1)}$  is given by an evaluation of  $\sum_{q=0}^m \tilde{\varphi}_{k,e} \circ \Psi_q$  at the given data points, i.e.

$$\mathbf{B} := \sum_{q=0}^m \begin{pmatrix} \tilde{\varphi}_{0,\mathfrak{r}}(\Psi_q(\mathbf{x}_1)) & \tilde{\varphi}_{0,\mathfrak{i}}(\Psi_q(\mathbf{x}_1)) & \dots & \tilde{\varphi}_{\ell,\mathfrak{r}}(\Psi_q(\mathbf{x}_1)) & \tilde{\varphi}_{\ell,\mathfrak{i}}(\Psi_q(\mathbf{x}_1)) \\ \vdots & & & & \vdots \\ \tilde{\varphi}_{0,\mathfrak{r}}(\Psi_q(\mathbf{x}_P)) & \tilde{\varphi}_{0,\mathfrak{i}}(\Psi_q(\mathbf{x}_P)) & \dots & \tilde{\varphi}_{\ell,\mathfrak{r}}(\Psi_q(\mathbf{x}_P)) & \tilde{\varphi}_{\ell,\mathfrak{i}}(\Psi_q(\mathbf{x}_P)) \end{pmatrix}, \quad (4.47)$$

while the symmetric regularization matrix is given by

$$\mathbf{S} := \begin{pmatrix} S_{0,0}^{\text{r,r}} & S_{0,0}^{\text{r,i}} & \dots & S_{0,\ell}^{\text{r,r}} & S_{0,\ell}^{\text{r,i}} \\ S_{0,0}^{\text{i,r}} & S_{0,0}^{\text{i,i}} & \dots & S_{0,\ell}^{\text{i,r}} & S_{0,\ell}^{\text{i,i}} \\ \vdots & & & & \vdots \\ S_{\ell,0}^{\text{r,r}} & S_{\ell,0}^{\text{r,i}} & \dots & S_{\ell,\ell}^{\text{r,r}} & S_{\ell,\ell}^{\text{r,i}} \\ S_{\ell,0}^{\text{i,r}} & S_{\ell,0}^{\text{i,i}} & \dots & S_{\ell,\ell}^{\text{i,r}} & S_{\ell,\ell}^{\text{i,i}} \end{pmatrix} \in \mathbb{R}^{2(\ell+1) \times 2(\ell+1)}.$$

The data vector  $\mathbf{y} = (y_1, \dots, y_P)^T \in \mathbb{R}^P$  collects all known function values  $y_j = f(\mathbf{x}_j)$ ,  $j = 1, \dots, P$ .

Thus, the numerical costs to compute a minimizer of (4.45) are at most of the order  $\mathcal{O}(P 2^2(\ell+1)^2)$  to compute  $\mathbf{B}^T \mathbf{B}$ , and  $\mathcal{O}(8(\ell+1)^3)$  to solve (4.46), e.g. when using a direct solver.

In this section, we could identify the underlying approximation space for the extended model  $f_\ell$  as defined in (4.35), to be the space  $\mathcal{H}_\ell^{(1)}$ . To be more precise, to approximate a given function  $f \in \mathcal{C}^0([0, 1]^n)$  we fixed the first model by claiming  $f_\ell \in \mathcal{H}_\ell^{(1)}$  and approximate  $f$  by  $f_{\ell,\text{r}} = (f_\ell + \bar{f}_\ell^{(1)})/2 \in \mathcal{H}_\ell^{(1)}$ . Additionally, the corresponding norm has been computed in (4.42) such that we could formulate the Regularization Network approach for the first model in Problem 4.18. Finally, the necessary and sufficient conditions for an approximand were given by means of a system of linear equations (4.46).

### 4.5.2 Regularization Network approach for the second model

To compute the set of functions that describes the (extended) second model  $f_L$  from (4.36), we take the finite sets  $\mathcal{F}_{\ell_0}^0, \dots, \mathcal{F}_{\ell_n}^n$  in their general form, as introduced in (4.38). Thus, our starting point is the space  $\mathcal{H}_L \subset \mathcal{H}(K)$  from (4.41). We define a subset of this space by

$$\mathcal{K}_L := \left\{ \sum_{i_0, \dots, i_n=0}^{\ell_0, \dots, \ell_n} c_{i_0, \dots, i_n} \Phi_{i_0, \dots, i_n}(\mathbf{x}) : c_{i_0, \dots, i_n} = c_{i_0}^0 \cdots c_{i_n}^n, c_{i_d}^d \in \mathbb{C} \right\} \subset \mathcal{H}_L. \quad (4.48)$$

One can easily compute that an element  $f \in \mathcal{K}_L$  is of the form

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i_0, \dots, i_n=0}^{\ell_0, \dots, \ell_n} c_{i_0, \dots, i_n} \Phi_{i_0, \dots, i_n}(\mathbf{x}) = \sum_{q=0}^m \left( \sum_{i_0, \dots, i_n=0}^{\ell_0, \dots, \ell_n} c_{i_0}^0 \cdots c_{i_n}^n \varphi_{i_0}^0 \cdots \varphi_{i_n}^n \right) \circ \Psi_q(\mathbf{x}) \\ &= \sum_{q=0}^m \prod_{d=0}^n \left( \sum_{i=0}^{\ell_d} c_i^d \varphi_i^d \right) \circ \Psi_q(\mathbf{x}) =: \sum_{q=0}^m \left( \prod_{d=0}^n \phi_{\ell_d}^d \right) \circ \Psi_q(\mathbf{x}). \end{aligned}$$

This shows that  $f_L \in \mathcal{K}_L$ , and conversely that  $\mathcal{K}_L$  only consists of elements that have the same structure as the extended second model. Note that  $\mathcal{H}_\ell^{(1)} \subset \mathcal{K}_L$ , and

that  $\mathcal{K}_L$  is not a linear space: For elements  $f_1, f_2 \in \mathcal{K}_L$  and  $\alpha, \beta \in \mathbb{C}$ , in general it holds that  $\alpha f_1 + \beta f_2 \notin \mathcal{K}_L$ .

For the  $n$ -dimensional norm  $\|\cdot\|_{\mathcal{H}(K)}$  of elements in  $\mathcal{K}_L$ , in particular of  $f_L$ , we derive with (4.34)

$$\begin{aligned} \|f_L\|_{\mathcal{H}(K)}^2 &= \sum_{i_0, \dots, i_n=0}^{\ell_0, \dots, \ell_n} \gamma_{i_0}^0 \cdots \gamma_{i_n}^n (c_{i_0}^0 \cdots c_{i_n}^n)^2 = \prod_{d=0}^n \left( \sum_{i=0}^{\ell_d} \gamma_i^d (c_i^d)^2 \right) \\ &= \prod_{d=0}^n \left\| \sum_{i=0}^{\ell_d} c_i^d \varphi_i^d \right\|_{\mathcal{H}^d(k^d)}^2. \end{aligned} \quad (4.49)$$

Thus, it can be computed by a product of one-dimensional norms in  $\mathcal{H}^d(k^d)$ .

Now, for the second model the Regularization Network approach consists of the following

**Problem 4.19.** *Let the data set  $\mathcal{Z} \subset [0, 1]^n \times \mathbb{R}$  be given by (4.1) and let  $0 < \nu \in \mathbb{R}$ . Then, find a minimum  $f_{L,\mathbf{r}} : [0, 1]^n \rightarrow \mathbb{R}$  of*

$$E^{(n)}(f_{L,\mathbf{r}}) := \frac{1}{P} \sum_{j=1}^P (f_{L,\mathbf{r}}(\mathbf{x}_j) - y_j)^2 + \nu \|f_{L,\mathbf{r}}\|_{\mathcal{H}(K)}^2, \quad (4.50)$$

with  $f_L \in \mathcal{K}_L \subset \mathcal{H}_L$ . Again,  $f_{L,\mathbf{r}}$  and  $f_L$  are interrelated by (4.37).

Analogue to the first model, we replaced  $f_L$  with the real valued function  $f_{L,\mathbf{r}}$  in the minimization. This means that we define the model by claiming  $f_L \in \mathcal{K}_L$  and search for a minimizer  $f_{L,\mathbf{r}} \in \mathcal{H}(K)$  which is enforced by the norm.

Let  $\mathbf{c}_{L,\mathbb{C}}^{(n)} := (c_0^0, \dots, c_{\ell_0}^0, \dots, c_0^n, \dots, c_{\ell_n}^n) \in \mathbb{C}^L$ ,  $L := \sum_{d=0}^n (\ell_d + 1)$  denote the complex coefficient vector of  $f_L$ . In general, similar to the computations in Section 4.5.1, the real valued formulation follows from the splitting of all complex coefficients  $c_j^d$  and basis functions  $\varphi_j^d$ ,  $d = 0, \dots, n$ ,  $j = 0, \dots, \ell_d$  into their real and imaginary part:

$$c_j^d = c_{j,\mathbf{r}}^d + i c_{j,\mathbf{i}}^d, \quad \text{and} \quad \varphi_j^d(s) = \varphi_{j,\mathbf{r}}^d(s) + i \varphi_{j,\mathbf{i}}^d(s).$$

However, an expansion of  $f_{L,\mathbf{r}}$  and  $\|f_{L,\mathbf{r}}\|_{\mathcal{H}(K)}^2$  in terms of  $(\varphi_{j,\mathbf{r}}^d \circ \Psi_q)$ , and  $(\varphi_{j,\mathbf{i}}^d \circ \Psi_q)$  similar to (4.44) becomes very tedious, see (A.6), and will be omitted here. We rather give an expansion which results from a splitting of each factor  $\phi_{\ell_d}^d = \phi^d = \phi_{\mathbf{r}}^d + i \phi_{\mathbf{i}}^d$ ,  $d = 0, \dots, n$ , into its real and imaginary part. Note that they have the

representations

$$\begin{aligned}\phi^d(s) &= \sum_{j=0}^{\ell_d} \left( c_{j,\mathfrak{r}}^d + i c_{j,\mathfrak{i}}^d \right) \left( \varphi_{j,\mathfrak{r}}^d(s) + i \varphi_{j,\mathfrak{i}}^d(s) \right) \\ &= \sum_{j=0}^{\ell_d} \left( c_{j,\mathfrak{r}}^d \varphi_{j,\mathfrak{r}}^d(s) - c_{j,\mathfrak{i}}^d \varphi_{j,\mathfrak{i}}^d(s) \right) + i \sum_{j=0}^{\ell_d} \left( c_{j,\mathfrak{i}}^d \varphi_{j,\mathfrak{r}}^d(s) + c_{j,\mathfrak{r}}^d \varphi_{j,\mathfrak{i}}^d(s) \right) \\ &=: \phi_{\mathfrak{r}}^d(s) + i \phi_{\mathfrak{i}}^d(s), \quad d = 0, \dots, n.\end{aligned}$$

For more details on the computations in this section we always refer to Appendix A.3.

With these splittings, it can be shown that  $f_{L,\mathfrak{r}}$  has the nonlinear expansion

$$f_{L,\mathfrak{r}}(\mathbf{x}) = \sum_{\substack{j_0, \dots, j_n=0 \\ (\sum_{d=0}^n j_d) \text{ even}}}^1 (-1)^{\frac{1}{2}(j_0+\dots+j_n)} \sum_{q=0}^m \left( \prod_{d=0}^n \phi_{e(j_d)}^d \left( \Psi_q(\mathbf{x}) \right) \right), \quad (4.51)$$

where we defined  $e(0) := \mathfrak{r}$ , and  $e(1) := \mathfrak{i}$ . Note that we always define 0 to be an even number since a division by 2 produces no remainder. Furthermore, defining for  $d = 0, \dots, n$ ,  $j \in \{0, 1\}$  the real values

$$S_j^d := \begin{cases} \langle \phi_{\mathfrak{r}}^d, \phi_{\mathfrak{r}}^d \rangle_{\mathcal{H}^d(k^d)} - \langle \phi_{\mathfrak{i}}^d, \phi_{\mathfrak{i}}^d \rangle_{\mathcal{H}^d(k^d)}, & j = 0, \\ \langle \phi_{\mathfrak{r}}^d, \phi_{\mathfrak{i}}^d \rangle_{\mathcal{H}^d(k^d)} + \langle \phi_{\mathfrak{i}}^d, \phi_{\mathfrak{r}}^d \rangle_{\mathcal{H}^d(k^d)}, & j = 1, \end{cases}$$

the regularization term can be computed by

$$\|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 = \frac{1}{2} \prod_{d=0}^n \left( \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \|\phi_e^d\|_{\mathcal{H}^d(k^d)}^2 \right) + \frac{1}{2} \sum_{\substack{j_0, \dots, j_n=0 \\ (\sum_{d=0}^n j_d) \text{ even}}}^1 (-1)^{\frac{1}{2}(j_0+\dots+j_n)} \prod_{d=0}^n S_{j_d}^d.$$

**Remark 4.20.** Note that another possibility to derive a real valued model  $\tilde{f}_L^{(n)} : [0, 1]^n \rightarrow \mathbb{R}$  from  $f_L$  is to define

$$\tilde{f}_L^{(n)}(\mathbf{x}) := \sum_{q=0}^m \left( \prod_{d=0}^n \operatorname{Re}(\phi^d) \right) \circ \Psi_q(\mathbf{x}) \neq \sum_{q=0}^m \operatorname{Re} \left( \prod_{d=0}^n (\phi^d) \right) \circ \Psi_q(\mathbf{x}) = f_{L,\mathfrak{r}}(\mathbf{x}).$$

Since by property (P3) it holds that  $\phi^d, \bar{\phi}^d \in \mathcal{H}_{\ell_d}^d$ ,  $d = 0, \dots, n$  this implies  $\tilde{f}_L^{(n)} \in \mathcal{K}_L$ , while in general  $f_{L,\mathfrak{r}} \notin \mathcal{K}_L$ . Remember that  $\mathcal{K}_L$  is not a linear space. However, as we will see in Section 6.2.1, this definition is more restrictive, and for this reason we stick to (4.51).

It is important to note that a direct implementation of (4.51) is impractical, since it involves the computation of  $2^{n-1}$  terms. Therefore, it is essential to use complex arithmetic and first derive  $f_L$  to get  $f_{L,\mathfrak{r}} = \operatorname{Re}(f_L)$ . The same holds for  $\|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 = \|f_L + \bar{f}_L^{(n)}\|_{\mathcal{H}(K)}^2/4$ .

From the nonlinear expansion of  $f_{L,\mathfrak{r}}$  in terms of the real functions  $(\phi_{\mathfrak{r}}^d \circ \Psi_q)$  and  $(\phi_{\mathfrak{i}}^d \circ \Psi_q)$ , we can conclude that their expansion in terms of  $(\varphi_{j,\mathfrak{r}}^d \circ \Psi_q)$  and  $(\varphi_{j,\mathfrak{i}}^d \circ \Psi_q)$  is not linear, see also Appendix A.3. This implies that the error functional

$$E^{(n)}(f_{L,\mathfrak{r}}) = E^{(n)}(\mathbf{c}_L^{(n)})$$

is not convex with respect to the real coefficient vector

$$\mathbf{c}_L^{(n)} := (c_{0,\mathfrak{r}}^0, c_{0,\mathfrak{i}}^0, \dots, c_{\ell_0,\mathfrak{r}}^0, c_{\ell_0,\mathfrak{i}}^0, \dots, c_{0,\mathfrak{r}}^n, c_{0,\mathfrak{i}}^n, \dots, c_{\ell_n,\mathfrak{r}}^n, c_{\ell_n,\mathfrak{i}}^n)^T \in \mathbb{R}^{2L}.$$

Therefore, the conditions on a potential minimizer

$$\frac{\partial}{\partial c_{\mu,\theta}^\delta} E^{(n)}(\mathbf{c}_L^{(n)}) = 0, \quad \text{for all } \delta = 0, \dots, n, \quad \mu = 0, \dots, \ell, \quad \theta \in \{\mathfrak{r}, \mathfrak{i}\},$$

are necessary but not sufficient. For a computation of the derivatives we refer to Appendix A.3.

Thus, the minimization Problem 4.19 now results in solving the system of nonlinear equations

$$\nabla E^{(n)}(\mathbf{c}_L^{(n)}) = 0 \in \mathbb{R}^{2L},$$

for which we cannot guarantee a unique solution to exist. Standard techniques to minimize nonlinear functions  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  will be briefly introduced in Section 5.1.

In this section, we defined the approximation set  $\mathcal{K}_L \subset \mathcal{H}_L \subset \mathcal{H}(K)$  which is the basis for the extended second model  $f_L$  as defined in (4.36). Then, we defined our model by claiming that  $f_L \in \mathcal{K}_L$  and approximate a given function  $f \in C^0([0, 1]^n)$  with  $f_{L,\mathfrak{r}} = (f_L + \bar{f}_L^{(n)})/2 \in \mathcal{H}_L$ . With the norm  $\|\cdot\|_{\mathcal{H}(K)}$  that is induced by (4.34) we formulated the Regularization Network approach in Problem 4.19. It turned out that the corresponding error functional is not convex and that a potential minimizer has to fulfill a system of nonlinear equations.

**Remark 4.21.** We finish with the remark that the constructions in this section and Section 4.5.1 are also valid for a real valued construction, i.e. when starting from (4.18) with basis functions  $\varphi_j^d : \Omega \rightarrow \mathbb{R}$  in the sets  $\mathcal{F}^0, \dots, \mathcal{F}^n$ . Then, the space  $\mathcal{H}(K)$  consists of real valued functions and it trivially holds  $f_{\ell,\mathfrak{r}} = f_\ell$ , and  $f_{L,\mathfrak{r}} = f_L$ . The main results remain the same: To compute the minimizer in Problem 4.18 a system of linear equations has to be solved for a coefficient vector  $\mathbf{c}_\ell^{(1)} \in \mathbb{R}^{\ell+1}$ , whereas a minimizer  $\mathbf{c}_L^{(n)} \in \mathbb{R}^L$  in Problem 4.19 is still characterized by a system of nonlinear equations.



# Chapter 5

## Implementational details

The primary goal that led us to the the Regularization Network approaches in form of Problem 4.18 and Problem 4.19 was to benefit from the constructive version of Kolmogorov's superposition Theorem 2.14. A main point in our argumentation to define this new approach was in Section 3.2 that Sprecher's Algorithm 2.1 to compute the outer function  $\Phi$  requires exponential numerical costs with respect to the dimensionality  $n$  of the function  $f$  that is approximated. To account for this point in the new method, it is important to consider details on the implementation of our Regularization Network approaches which influence their numerical complexity.

To this end, we first investigate nonlinear minimization techniques to minimize  $E^{(n)}$  in Section 5.1. The important point will be an estimation of the numerical costs to evaluate the error functional and its gradient. Then, in Section 5.2, we deal with the evaluation of the inner sums  $\Psi_q$ , and consider the influence of boundary values of  $f$  in Section 5.3. Finally, we end with some remarks on a possible parallelization of the algorithm in Section 5.4.

### 5.1 Nonlinear minimizers

In Section 4.5.1 we have seen that the error functional  $E^{(n)}(\mathbf{c}_L^{(n)})$  is not convex. Therefore, nonlinear minimization procedures have to be used to compute a solution of Problem 4.19. Here, different gradient based standard iteration schemes are tested to compute a minimum of (4.50). Namely, we use

- (a) the steepest descent method,
- (b) the Fletcher–Reeves conjugate gradient method,
- (c) the Polak–Ribi  re conjugate gradient method,
- (d) and the Broyden–Fletcher–Goldfarb–Shanno algorithm.

In the following we briefly introduce the methods to gain insight into their computational complexities. A numerical analysis will be given in Section 6.2.1. For more details on the theory we refer to [91]. In this section, we consider the general problem of minimizing a function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ .

All minimizers belong to the class of so-called line search methods. Here, for a start value  $\mathbf{x}_0 \in \mathbb{R}^d$ , in each iteration step  $k = 0, 1, \dots$ , a search direction  $\mathbf{p}_k$  is computed and for some step length  $\alpha_k$  the next iterate is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k. \quad (5.1)$$

Line search methods differ in the choice of  $\mathbf{p}_k$  and  $\alpha_k$ .

Methods (a)–(c) are gradient based minimization methods, and the simplest choice for a search direction is to use the steepest descent at  $\mathbf{x}_k$  which is given by the negative gradient  $\mathbf{p}_k = -\nabla g(\mathbf{x}_k)$ . Anyhow, the **steepest descent (SD) method** can lead to very inefficient convergence even for exact line search, i.e. when  $\alpha_k$  minimizes the one-dimensional function  $\ell(\alpha) := g(\mathbf{x}_k + \alpha \mathbf{p}_k)$  with respect to  $\alpha$  exactly.

To introduce more efficient methods we first consider the minimization of a quadratic function of the special form  $g_{\mathbf{Q}, \mathbf{b}}(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} - \mathbf{b}^T \mathbf{x}$  for some symmetric positive definite matrix  $\mathbf{Q}$  and some vector  $\mathbf{b}$ . Then, to compute the  $k$ -th iterate  $\mathbf{x}_k$  and the residual  $\mathbf{r}_k := \mathbf{Q} \mathbf{x}_k - \mathbf{b} = \nabla g_{\mathbf{Q}, \mathbf{b}}(\mathbf{x}_k)$ , the conjugate gradient method updates the last search direction by

$$\mathbf{p}_k = \begin{cases} -\mathbf{r}_0 & \text{if } k = 0, \\ -\mathbf{r}_k + \beta_k \mathbf{p}_{k-1} & \text{else,} \end{cases} \quad \text{where} \quad \beta_k = \left( \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1}} \right). \quad (5.2)$$

Then, the step size parameter  $\alpha_k$  is computed by an exact line search. For the (temporarily) considered quadratic function this value is given by

$$\alpha_k = -\frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{Q} \mathbf{p}_k}.$$

It can be shown that for any  $\mathbf{x}_0$ , the value  $\mathbf{x}_n$  is the minimizer of  $g_{\mathbf{Q}, \mathbf{b}}(\mathbf{x})$ . The basic property for this result is that the directions  $\mathbf{p}_k$  are mutually conjugate with respect to  $\mathbf{Q}$ , i.e. we have  $\mathbf{p}_j^T \mathbf{Q} \mathbf{p}_k = 0$ , if  $j \neq k$ . Note that the conjugate gradient method is defined for a minimization of quadratic functions and cannot be applied to solve Problem 4.19. For that purpose, it has to be generalized for non-quadratic functions  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ .

This is done by the **Fletcher–Reeves conjugate gradient (FRcg) method**. Here, the residual  $\mathbf{r}_k$  is replaced by the gradient  $\nabla g(\mathbf{x}_k)$  in (5.2), and the step length  $\alpha_k$  is now determined by an inexact line search. Namely it is chosen such that it

satisfies for  $0 < c_1 < c_2 < 1/2$  the so called strong Wolfe conditions

$$\begin{aligned}\ell(\alpha_k) &= g(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq g(\mathbf{x}_k) + c_1 \alpha_k \mathbf{p}_k^T \nabla g(\mathbf{x}_k) \\ \frac{\partial}{\partial \alpha_k} \ell(\alpha_k) &= \mathbf{p}_k^T \nabla g(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq c_2 |\mathbf{p}_k^T \nabla g(\mathbf{x}_k)|.\end{aligned}\quad (5.3)$$

These conditions ensure that the next direction  $\mathbf{p}_{k+1}$  is indeed a descent direction. This even holds for any line search method that yields an  $\alpha_k$  satisfying (5.3).

There are many variants of the FRcg–method that differ mainly in the choice of the parameter  $\beta_k$  in (5.2). For the **Polak–Ribi  re conjugate gradient (PRcg) method** the update parameter is defined by

$$\beta_k := \frac{\nabla g(\mathbf{x}_k)^T [\nabla g(\mathbf{x}_k) - \nabla g(\mathbf{x}_{k-1})]}{\|\nabla g(\mathbf{x}_k)\|^2}.$$

For strongly convex quadratic functions and exact line search the PRcg–method coincides with the FRcg–method but numerical experience indicates that for general nonlinear functions the PRcd–method tends to be more robust and efficient [91].

Method (d) belongs to the class of so–called Quasi–Newton methods. Solvers from this class are derived from the Newton method, where the search directions  $\mathbf{p}_k$  are computed from the gradient  $\nabla g(\mathbf{x}_k)$  and the Hessian  $\nabla^2 g(\mathbf{x}_k)$  by solving the linear system

$$\nabla^2 g(\mathbf{x}_k) \mathbf{p}_k = -\nabla g(\mathbf{x}_k).$$

Now, for Quasi–Newton methods the Hessian is approximated by a matrix  $\mathbf{H}_k$  to avoid a computation of  $\nabla^2 g(\mathbf{x}_k)$ . Practically,  $\mathbf{H}_k$  is given by means of a low–rank formula. Note that with  $\mathbf{H}_k = \mathbf{I}$  the steepest descent method belongs to this class.

In the **Broyden–Fletcher–Goldfarb–Shanno (BFGS) method** the matrix  $\mathbf{H}_{k-1}$  is updated in each iteration step by informations gained from the update of the gradients  $\nabla g(\mathbf{x}_k)$  and  $\nabla g(\mathbf{x}_{k-1})$  to compute the current approximation  $\mathbf{H}_k$  to the Hessian. Practically, rather than updating  $\mathbf{H}_{k-1}$  and inverting it, the inverse  $\mathbf{G}_{k-1} = \mathbf{H}_{k-1}^{-1}$  is updated directly in each step by

$$\mathbf{G}_k = (\mathbf{I} - \rho_{k-1} \mathbf{s}_{k-1} \mathbf{y}_{k-1}^T) \mathbf{G}_{k-1} (\mathbf{I} - \rho_{k-1} \mathbf{y}_{k-1} \mathbf{s}_{k-1}^T) + \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}_{k-1}^T,$$

where

$$\mathbf{s}_{k-1} := \mathbf{x}_k - \mathbf{x}_{k-1}, \quad \mathbf{y}_{k-1} := \nabla g(\mathbf{x}_k) - \nabla g(\mathbf{x}_{k-1}), \quad \text{and} \quad \rho_{k-1} := (\mathbf{y}_{k-1}^T \mathbf{s}_{k-1})^{-1}.$$

An initial approximation to the inverse Hessian  $\mathbf{G}_0 \approx \mathbf{H}_0^{-1}$  has to be given. The search direction is given by

$$\mathbf{p}_k = -\mathbf{G}_k \nabla g(\mathbf{x}_k).$$

method	SD	FRcg	PRcg	BFGS
complexity	$\mathcal{O}(d) + C_{g,\nabla g}$	$\mathcal{O}(d) + C_{g,\nabla g}$	$\mathcal{O}(d) + C_{g,\nabla g}$	$\mathcal{O}(d^2) + C_{g,\nabla g}$

**Table 5.1.** Numerical complexities for one iteration step of the respective nonlinear minimizers (a)–(d) to minimize a non-quadratic function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ .  $C_{g,\nabla g}$  denotes the complexity that is required to evaluate the function  $g$  and its gradient  $\nabla g$ .

For the derivation of these formulae and a detailed description of the BFGS–method we refer to [91].

Practically, the *GNU Scientific Library* (GSL) provides efficient implementations of methods (a)–(d). The GSL is freely published under the *GNU public license*, see [138]. We remark that there, rather than using the strong Wolfe conditions (5.3), the line search stepsize parameter  $\alpha_k$  is chosen such that for some tolerance  $\delta \geq 0$  it holds

$$\frac{\partial \ell(\alpha_k)}{\partial \alpha} = \mathbf{p}_k^T \nabla g(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq \delta |\mathbf{p}_k| |\nabla g(\mathbf{x}_k + \alpha_k \mathbf{p}_k)| . \quad (5.4)$$

For  $\delta = 0$  this corresponds to an exact line search. We use the methods (a)–(d) from the GSL in our implementations.

All nonlinear minimizers that are presented here only require an evaluation of the function  $g$  and its gradient  $\nabla g$ . However, the numerical complexities for these operations depend on the function that is considered. We denote the overall complexity for one evaluation of  $g$  and  $\nabla g$  by  $C_{g,\nabla g}$ . Table 5.1 shows the numerical complexities for one iteration step of the minimizers (a)–(d). There, due to the lack of a detailed information on the internal GSL–implementation of the line search minimization, we assume that the stopping condition (5.4) can be used in combination with a small  $\delta > 0$  such that  $\mathcal{O}(1)$  steps are required to approximate the minimizer of  $\ell(\alpha)$ . Numerical experiences showed that the latter assumption is reasonable.

Now, the function which has to be minimized in our context, is the error functional  $E^{(n)}(f_{L,r}) = E^{(n)}(\mathbf{c}_L^{(n)})$  from (4.50) with respect to the vector  $\mathbf{c}_L^{(n)} \in \mathbb{R}^{2L}$ . This means that in our considerations we have  $g := E^{(n)}$  and  $d := 2L$ . The entries of the gradient  $\nabla E^{(n)}(\mathbf{c}_L^{(n)})$  are given by (A.9). In Section 5.1.1 we investigate the numerical complexities for an evaluation of these functions, i.e. we compute  $C_{g,\nabla g} = C_{E^{(n)}, \nabla E^{(n)}}$ .

However, prior to this we introduce a **Nested Iteration scheme** which is additionally applied for all iterative solvers. To explain this in more detail, consider in (4.38) for each  $d = 0, \dots, n$ , and increasing integers  $\ell_{d,1} < \dots < \ell_{d,J} = \ell_d$ , the

following sequence of nested sets of basis functions:

$$\mathcal{F}_{\ell_{d,1}}^d \subset \mathcal{F}_{\ell_{d,2}}^d \subset \dots \subset \mathcal{F}_{\ell_{d,J}}^d = \mathcal{F}_{\ell_d}^d \subset \mathcal{F}^d.$$

With the constructions from Section 4.5 and (4.41) this defines the nested sequence of subspaces

$$\mathcal{H}_{L_1} \subset \mathcal{H}_{L_2} \subset \dots \subset \mathcal{H}_{L_J} = \mathcal{H}_L \subset \mathcal{H}(K),$$

where  $L_j = \sum_{d=0}^n (\ell_{d,j} + 1)$ .

Now, to compute a minimizer of  $E^{(n)}$  in  $\mathcal{H}_L$ , first select a method  $\text{ITER}(\cdot) \in \{(a), (b), (c), (d)\}$  which performs one iteration step of the respective nonlinear minimizer. Furthermore, define for each  $j = 1, \dots, J$  a maximum number of iterations  $k_{\max}^j \in \mathbb{N}$ , and choose an accuracy  $\varepsilon > 0$ . Then, we use the following

**ALGORITHM 5.1 (Nested Iteration scheme).** Let the start value  $\mathbf{c}_0^1 \in \mathbb{R}^{2L_1}$  be given. Set  $j \leftarrow 1$  and compute

(1) for  $k = 0, \dots, k_{\max}^j$ , the next iterate

$$\mathbf{c}_{k+1}^j = \text{ITER}(\mathbf{c}_k^j).$$

If  $\frac{|E^{(n)}(\mathbf{c}_{k+1}^j) - E^{(n)}(\mathbf{c}_k^j)|}{|E^{(n)}(\mathbf{c}_{k+1}^j)|} < \varepsilon$ , then set  $\mathbf{c}_{L_j}^{(n)} = \mathbf{c}_{k+1}^j$  and go to (2).

(2) Let  $f_{L_j,\mathbf{r}} \in \mathcal{H}_{L_j}$  be given by (A.6) with the coefficient vector  $\mathbf{c}_{L_j}^{(n)}$ .

If  $j = J$ , then stop the algorithm.

For  $j < J$ , set  $\mathbf{c}_0^{j+1} \in \mathbb{R}^{2L_j}$  to be the vector of expansion coefficients of  $f_{L_j,\mathbf{r}} \in \mathcal{H}_{L_j}$ . and increase  $j \leftarrow j + 1$ . Then go to (1).

This algorithm produces a function  $f_{L,\mathbf{r}} \in \mathcal{H}_L$ . It recursively precomputes the start values for the respective iterative nonlinear minimizer on level  $j$  as the solution  $f_{L_{j-1},\mathbf{r}}$  of the same problem at the lower resolution  $j - 1$ .

Our motivation to use Algorithm 5.1 is simple: We have seen in Section 4.5.2 and Appendix A.3 that  $E^{(n)}$  is highly non-convex. Therefore, we start with a simple form of Problem 4.19 by choosing a low resolution to get a first approximation of the desired solution. This approximation can be easily computed and might serve as a good initial guess at a higher resolution. Thus, one advantage is to reduce the dependency of the method from the choice of the start value  $\mathbf{c}_0^1$ . A numerical analysis of Algorithm 5.1 will be given in Section 6.2.

### 5.1.1 Numerical complexities for evaluations of $E^{(n)}, \nabla E^{(n)}$

To get a final estimate on the numerical costs for one iteration step of the nonlinear minimizers we have to substantiate the complexities for an evaluation of the error

functional  $E^{(n)} : \mathbb{R}^{2L} \rightarrow \mathbb{R}$  and the gradient  $\nabla E^{(n)} : \mathbb{R}^{2L} \rightarrow \mathbb{R}^{2L}$ . We already noted in Remark 4.20 that it is essential to use complex arithmetic for an evaluation of these terms to avoid exponential costs. Remember that we have to compute

$$E^{(n)}(\mathbf{c}_L^{(n)}) = \frac{1}{P} \sum_{j=1}^P (f_{L,\mathfrak{r}}(\mathbf{x}_j) - y_j)^2 + \nu \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2,$$

and  $\nabla E^{(n)}(\mathbf{c}_L^{(n)})$  which consists of the coordinate functions

$$\frac{\partial}{\partial c_{\mu,\theta}^\delta} E^{(n)}(\mathbf{c}_L^{(n)}) = \nu \left( \frac{\partial}{\partial c_{\mu,\theta}^\delta} \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 \right) + \frac{2}{P} \sum_{j=1}^P (f_{L,\mathfrak{r}}(\mathbf{x}_j) - y_j) \left( \frac{\partial}{\partial c_{\mu,\theta}^\delta} f_{L,\mathfrak{r}}(\mathbf{x}_j) \right),$$

$$\delta = 0, \dots, n, \mu = 0, \dots, \ell_\delta, \theta \in \{\mathfrak{r}, \mathfrak{i}\}.$$

Now, before we consider the evaluation of these terms in more detail we further concretize the finite sets of basis functions  $\mathcal{F}_{\ell_0}^0, \dots, \mathcal{F}_{\ell_n}^n$  from (4.38). Namely, we assume for each  $d = 0, \dots, n$  that  $\ell_d = 2N_d$ ,  $N_d \in \mathbb{N}$  and

$$\varphi_{N_d-j}^d = \bar{\varphi}_{N_d+j}^d \in \mathcal{F}_{\ell_d}^d, \quad j = 0, \dots, N_d.$$

Shifting the index set  $\{0, \dots, 2N_d\} \mapsto \{-N_d, \dots, N_d\}$  it holds

$$\mathcal{F}_{\ell_d}^d = \mathcal{F}_{N_d}^d := \{\varphi_{-N_d}^d, \dots, \varphi_0^d, \dots, \varphi_{N_d}^d\},$$

and

$$\varphi_{-k}^d = \bar{\varphi}_k^d, \quad k \in \{-N_d, \dots, N_d\}.$$

Note that we are aiming at the choice  $\varphi_{\pm j}^d(s) := \exp(\pm 2\pi i k_j^d s / \omega_m)$ ,  $\omega_m \in \mathbb{R}$ ,  $k_j^d \in \mathbb{N}$  in Section 6.2. With this definition and (4.19) we obtain for  $j, k \in \{-N_d, \dots, N_d\}$  that

$$\langle \varphi_j^d, \bar{\varphi}_k^d \rangle_{\mathcal{H}(K)} = \langle \varphi_j^d, \varphi_{-k}^d \rangle_{\mathcal{H}(K)} = \gamma_k^d \delta_{-k,j}.$$

Next, for an evaluation of  $E^{(n)}(\mathbf{c}_L^{(n)})$  and  $\nabla E^{(n)}(\mathbf{c}_L^{(n)})$  we observe that the costly terms are the following: For given  $\mathbf{c}_L^{(n)} \in \mathbb{R}^{2L}$ , remember that  $c_k^d = c_{k,\mathfrak{r}}^d + i c_{k,\mathfrak{i}}^d \in \mathbb{C}$ . Then, we have to compute the function values

$$f_{L,\mathfrak{r}}(\mathbf{x}_j) = \sum_{q=0}^m \operatorname{Re} \left( \prod_{d=0}^n \left( \sum_{k=-N_d}^{N_d} c_k^d \varphi_k^d(\Psi_q(\mathbf{x}_j)) \right) \right), \quad (\text{A})$$

and the gradient, i.e. for all  $\delta = 0, \dots, n$ ,  $\mu = -N_\delta, \dots, N_\delta$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$ , calculate

$$\begin{aligned} \frac{\partial}{\partial c_{\mu,\theta}^\delta} f_{L,\mathfrak{r}}(\mathbf{x}_j) &= \\ \sum_{q=0}^m \left[ \begin{array}{l} \left\{ \begin{array}{ll} \varphi_{\mu,\mathfrak{r}}^\delta(\Psi_q(\mathbf{x}_j)), & \theta = \mathfrak{r} \\ -\varphi_{\mu,\mathfrak{i}}^\delta(\Psi_q(\mathbf{x}_j)), & \theta = \mathfrak{i} \end{array} \right\} \operatorname{Re} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \left( \sum_{k=-N_d}^{N_d} c_k^d \varphi_k^d(\Psi_q(\mathbf{x}_j)) \right) \right) \\ - \left\{ \begin{array}{ll} \varphi_{\mu,\mathfrak{i}}^\delta(\Psi_q(\mathbf{x}_j)), & \theta = \mathfrak{r} \\ \varphi_{\mu,\mathfrak{r}}^\delta(\Psi_q(\mathbf{x}_j)), & \theta = \mathfrak{i} \end{array} \right\} \operatorname{Im} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \left( \sum_{k=-N_d}^{N_d} c_k^d \varphi_k^d(\Psi_q(\mathbf{x}_j)) \right) \right) \end{array} \right] \quad (\text{B}) \end{aligned}$$

at the positions  $\mathbf{x}_j \in [0, 1]^n$ ,  $j = 1, \dots, P$ . Furthermore, we have to compute the norm

$$\|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 = \frac{1}{2} \prod_{d=0}^n \left( \sum_{k=-N_d}^{N_d} (c_k^d)^2 \gamma_k^d \right) + \frac{1}{2} \operatorname{Re} \left( \prod_{d=0}^n \left( \sum_{k=-N_d}^{N_d} c_k^d c_{-k}^d \gamma_k^d \right) \right), \quad (\text{C})$$

and for all  $\delta = 0, \dots, n$ ,  $\mu = -N_\delta, \dots, N_\delta$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$ , the partial derivatives

$$\begin{aligned} \frac{\partial}{\partial c_{\mu,\theta}^\delta} \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 &= \gamma_\mu^\delta c_{\mu,\theta}^\delta \prod_{\substack{d=0 \\ d \neq \delta}}^n \left( \sum_{k=-N_d}^{N_d} (c_k^d)^2 \gamma_k^d \right) \\ &+ \left[ \begin{array}{l} \left\{ \begin{array}{ll} \gamma_\mu^\delta c_{-\mu,\mathfrak{r}}^\delta, & \theta = \mathfrak{r} \\ -\gamma_\mu^\delta c_{-\mu,\mathfrak{i}}^\delta, & \theta = \mathfrak{i} \end{array} \right\} \operatorname{Re} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \left( \sum_{k=-N_d}^{N_d} c_k^d c_{-k}^d \gamma_k^d \right) \right) \\ - \left\{ \begin{array}{ll} \gamma_\mu^\delta c_{-\mu,\mathfrak{i}}^\delta, & \theta = \mathfrak{r} \\ \gamma_\mu^\delta c_{-\mu,\mathfrak{r}}^\delta, & \theta = \mathfrak{i} \end{array} \right\} \operatorname{Im} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \left( \sum_{k=-N_d}^{N_d} c_k^d c_{-k}^d \gamma_k^d \right) \right) \end{array} \right] . \quad (\text{D}) \end{aligned}$$

Since (B) has to be computed for all  $\delta = 0, \dots, n$ , (A) can be calculated simultaneously from partial results of (B). The same argument holds for (D) and (C). To explain this in more detail, for  $j = 1, \dots, P$ , and  $q = 0, \dots, m$  we proceed as follows: First, compute for each  $d = 0, \dots, n$  the sums

$$\sum_{k=-N_d}^{N_d} c_k^d \varphi_k^d(\Psi_q(\mathbf{x}_j)), \quad \sum_{k=-N_d}^{N_d} (c_k^d)^2 \gamma_k^d, \quad \text{and} \quad \sum_{k=-N_d}^{N_d} c_k^d c_{-k}^d \gamma_k^d,$$

storing the intermediate values  $\varphi_{\mu,\theta}^\delta(\Psi_q(\mathbf{x}_j))$ ,  $\gamma_\mu^\delta c_{\mu,\theta}^\delta$ , and  $\gamma_\mu^\delta c_{-\mu,\theta}^\delta$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$  in vectors. We remark that all sums are complex valued. From these sums, the respective

complex valued products in (A)–(D) can be computed. Separating these products into real and imaginary part allows for a calculation of the desired terms (A), (B), (C), and (D). Note that to obtain (A) and (B) these terms have to be added up for all  $q = 0, \dots, m$ , and finally, to get  $E^{(n)}$  and  $\nabla E^{(n)}$ , an additional sum has to be built over all  $j = 1, \dots, P$ .

Now, to estimate the total complexity of these operations we observe that a complex valued addition is equivalent to two real valued additions, while a complex valued multiplication requires four real multiplications and two real additions. For a division of complex numbers we have to perform seven multiplications and three additions of real numbers. Thus, the total complexity for a calculation of the data terms

$$\frac{1}{P} \sum_{j=1}^P (f_{L,\mathfrak{r}}(\mathbf{x}_j) - y_j)^2, \quad \text{and} \quad \frac{2}{P} \sum_{j=1}^P \left( f_{L,\mathfrak{r}}(\mathbf{x}_j) - y_j \right) \left( \frac{\partial}{\partial c_{\mu,\theta}^\delta} f_{L,\mathfrak{r}}(\mathbf{x}_j) \right) \quad (5.5)$$

for all  $\delta = 0, \dots, n$ ,  $\mu = -N_\delta, \dots, N_\delta$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$ , is  $\mathcal{O}(P(m(L+n)))$ . To compute

$$\|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2, \quad \text{and} \quad \frac{\partial}{\partial c_{\mu,\theta}^\delta} \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2$$

for all  $\delta = 0, \dots, n$ ,  $\mu = -N_\delta, \dots, N_\delta$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$ , it takes  $\mathcal{O}(L+n)$  real operations.

Finally, assume for simplicity that  $N = N_0 = \dots = N_n$  which implies  $L = (n+1)(2N+1) = \mathcal{O}(nN)$  and remember that in Theorem 2.14 it is assumed that  $m \geq 2n$ . Thus, we can presume  $m = \mathcal{O}(n)$  what results in the total complexity

$$C_{E^{(n)}, \nabla E^{(n)}} = \mathcal{O}(Pn^2N) + \mathcal{O}(nN) = \mathcal{O}(Pn^2N)$$

for an evaluation of  $E^{(n)}(\mathbf{c}_L^{(n)})$  and  $\nabla E^{(n)}(\mathbf{c}_L^{(n)})$  at a given point  $\mathbf{c}_L^{(n)} \in \mathbb{R}^{2(n+1)(2N+1)}$ .

From Table 5.1 one can directly see that this term dominates the complexities  $\mathcal{O}(nN) + \mathcal{O}(Pn^2N)$  for an iteration of the methods (a)–(c). For the BFGS method this depends on the number of points  $P$ . Here, the numerical complexity is given by  $\mathcal{O}(n^2N^2) + \mathcal{O}(Pn^2N)$  where the second term becomes dominant if  $P \gg N$  which is typically the case.

Thus, for all iterative solvers, the evaluation of the error functional and its gradient are the costly parts. A further necessity that amplifies this fact will be explained in the next section.

## 5.2 Multiple precision arithmetic

We investigate a numerical aspect of the inner sums  $\Psi_q : [0, 1]^n \rightarrow \mathbb{R}$ ,  $q \in \{0, \dots, m\}$ , see (2.36), which is important in an implementation of the Regularization Network

approaches that were introduced in Section 4.5. Namely, this section will show that it is essential to use a multiple precision arithmetic to compute  $\Psi_q(\mathbf{x})$ . To explain this, remember that in Definition 2.4 the function  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  was given on the set  $\mathcal{D} = \bigcup_{k \in \mathbb{N}} \mathcal{D}_k(\gamma)$ , where

$$\mathcal{D}_k(\gamma) = \left\{ d_k \in \mathbb{Q} : d_k = \sum_{r=1}^k i_r \gamma^{-r}, i_r \in \{0, \dots, \gamma - 1\} \right\}$$

denoted the set of terminating rational numbers at resolution  $k$  in the basis  $\gamma \in \mathbb{N}$ ,  $\gamma > 1$ . It was extended to the real line by (2.24), but note that for a numerical analysis we only need to consider the sets  $\mathcal{D}_k(\gamma)$ . Then, in Lemma 2.16, the minimal distance of the images of any two points  $\mathbf{d}_k^1 \neq \mathbf{d}_k^2 \in \mathcal{D}_k(\gamma)$  under the mapping  $\Psi_q$  was estimated for any  $q \in \{0, \dots, m\}$  by

$$|\Psi_q(\mathbf{d}_k^1) - \Psi_q(\mathbf{d}_k^2)| > \Lambda_\delta^{-1} \gamma^{-n\beta(k)} \quad \text{with } \Lambda_\delta > 1,$$

see also the proof of Lemma 2.19.

Note that in the basis  $\gamma$ , and for a fixed resolution  $k$ , the minimal distance of points  $\mathbf{d}_k^1, \mathbf{d}_k^2 \in \mathcal{D}_k(\gamma)$  is  $|\mathbf{d}_k^1 - \mathbf{d}_k^2| = \gamma^{-k}$ . Now, the previous estimate shows that, to distinguish the function values  $\Psi_q(\mathbf{d}_k^1), \Psi_q(\mathbf{d}_k^2)$ , one has to numerically resolve these one-dimensional points at least with resolution  $n\beta(k)$  which is larger than  $k$ . Otherwise we would not be able to differ the images of the  $n$ -dimensional points  $\mathbf{d}_k^1, \mathbf{d}_k^2$ . However, this is essential for an approximation of  $f$ , see Section 2.3.7. In conclusion, one can see that a higher numerical resolution  $n\beta(k)$  is needed for the one-dimensional values than for the coordinates of the  $n$ -dimensional points for which the resolution is  $k$ . This discrepancy will be investigated in the following for varying dimension  $n$ .

To this end, first remember from Theorem 2.1 and Theorem 2.14 that

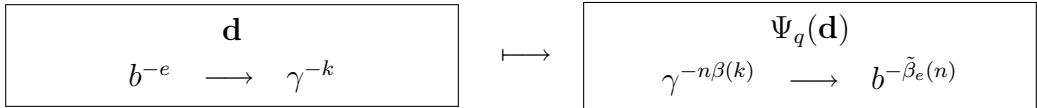
$$\beta(k) = \frac{n^k - 1}{n - 1}, \quad \text{and} \quad \gamma \geq m + 2 \geq 2n + 2.$$

Thus, for a fixed resolution  $k \in \mathbb{N}$  with respect to the basis  $\gamma$ , both values, namely the minimal distance of points  $\mathbf{d}_k^1, \mathbf{d}_k^2 \in \mathcal{D}_k(\gamma)$  which is given by  $\gamma^{-k}$ , and the minimal one-dimensional resolution  $\gamma^{-n\beta(k)}$  depend on the dimension  $n$  via the exponent and the basis for which we choose in the following  $\gamma = 2n + 2$ .

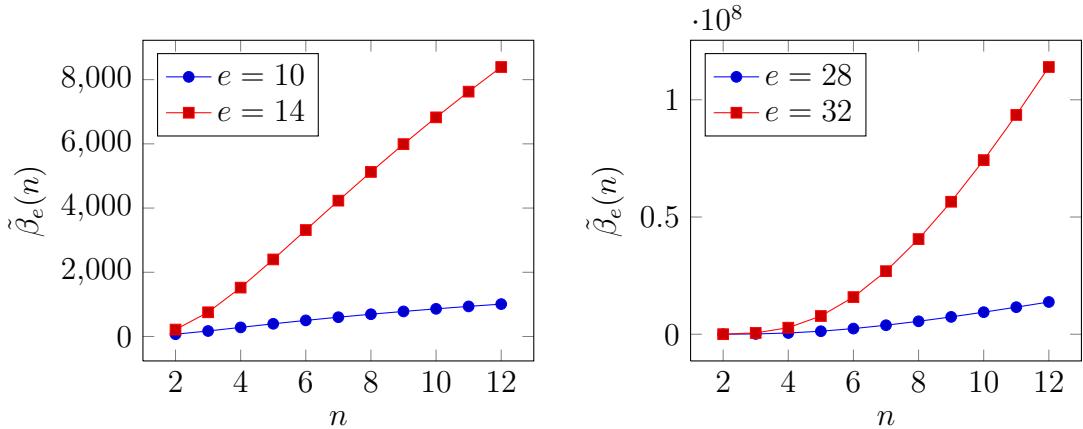
To investigate the dependency on  $n$ , we first have to transform both quantities in terms of a fixed basis  $b \in \mathbb{N}$  by

$$\begin{aligned} b^e = \gamma^k &\Leftrightarrow k = e \cdot (\log_\gamma b), \\ \gamma^{n\beta(k)} = b^{\tilde{\beta}_e(n)} &\Leftrightarrow \tilde{\beta}_e(n) = n \beta(k) (\log_b \gamma), \end{aligned} \tag{5.6}$$

and analyze for a fixed resolution  $e \in \mathbb{N}$  the growth of  $\tilde{\beta}_e(n)$ , see also Figure 5.1.



**Figure 5.1.** The values  $\mathbf{d}$  and  $\Psi_q(\mathbf{d})$  can be represented with respect to the bases  $b$  and  $\gamma$ . The resolutions  $e$ ,  $k$ ,  $n\beta(k)$ ,  $\tilde{\beta}_e(n)$  depend on the respective bases.



**Figure 5.2.** The graphs show estimates on the minimal number  $\tilde{\beta}_e(n)$  of bits that are required to resolve the function values  $\Psi_q(\mathbf{d})$  if the coordinates of the  $n$ -dimensional point  $\mathbf{d}$  are represented with  $e$  bits.  $\tilde{\beta}_e(n)$  does not grow exponentially with  $n$ , but for large values of  $e$  many bits are required.

For the basis  $b = 2$  this has a very simple interpretation: We start with a data type that provides  $e$  bits to represent the coordinates of  $\mathbf{d}$ . Then, we have to transform<sup>1</sup> this value into the basis  $\gamma$  which determines the resolution  $k$  by (5.6). Thus, to resolve the  $\Psi_q(\mathbf{d})$  we need at least resolution  $n\beta(k)$ . Finally, from (5.6) it follows that at least  $\tilde{\beta}_e(n)$  bits are needed to represent the function value  $\Psi_q(\mathbf{d})$ .

In other words,  $\tilde{\beta}_e(n)$  is a lower bound on the minimal number of bits that is required for the function value  $\Psi_q(\mathbf{d})$ , if the coordinates of the  $n$ -dimensional point  $\mathbf{d}$  are represented with  $e$  bits. It is given by

$$\tilde{\beta}_e(n) = n \cdot \beta \left( \frac{e}{\log_2 \gamma(n)} \right) (\log_2 \gamma(n)), \quad \text{where } \beta(k) = \frac{n^k - 1}{n - 1}.$$

Note that we did not take into account any floating point arithmetic here. Figure 5.2 shows plots of  $\tilde{\beta}_e(n)$  for different values of  $e$ .

The graphs show that the number of bits to resolve the inner sums  $\Psi_q$ ,  $q \in$

<sup>1</sup>The function  $\psi(x)$  is defined on  $\mathcal{D}_k(\gamma)$ .

$\{0, \dots, m\}$  become very large, even for moderate dimensions and values of  $e$ . However,  $\beta_e(n)$  does not grow exponentially with the dimension  $n$ . Therefore, special data types have to be used in an implementation that provide more memory than standard data types, like, e.g., `double` or `long double`. In our implementations, we used the *GNU Multiple Precision* arithmetic library [137] which is freely available and published under the *GNU public license*. Clearly, this also has to be applied for the evaluation of the basis functions  $\varphi_j(\Psi_q(\mathbf{x}))$ ,  $\varphi_j^d(\Psi_q(\mathbf{x}))$ .

### 5.3 Boundary values

In this section we introduce a transformation of the approximation problem from the unit cube  $[0, 1]^n$  to a sub cube  $\tilde{\Omega} \subsetneq [0, 1]^n$ , see below. In fact, we will see in Section 6.1.1 that an approximation of a function  $f : [0, 1]^n \rightarrow \mathbb{R}$  which does not vanish on the boundary  $\partial[0, 1]^n$  of the unit cube leads to large residuals near the boundary, see figures 6.4–6.7. The reason for this is the structure of the inner functions  $\Psi_q$ ,  $q = 0, \dots, n$  which transform the  $n$ -dimensional problem of approximating  $f$  to a one-dimensional approximation of the outer function  $\Phi$ . Therefore, our arguments are based on the considerations from Section 2.4.

There, we investigated the structure of the outer function  $\Phi$  by means of the ordering of points  $\mathbf{x} \in [0, 1]^n$  which is defined by (2.63). Because of the Z-curve structure of the inner sums  $\Psi_q$  there exist pairs of cubes  $S_{k,1}^q(\mathbf{d}_{k,1}^q)$ ,  $S_{k,2}^q(\mathbf{d}_{k,2}^q)$  for which the corresponding image intervals  $T_{k,1}^q(\mathbf{d}_{k,1}^q)$ ,  $T_{k,2}^q(\mathbf{d}_{k,2}^q)$  are neighboured on the real line, while the cubes themselves have large distance. From Figure 2.9 it becomes clear that the respective cubes are located at different boundaries of the hypercube  $[0, 1]^n$ . Since the piecewise constant function  $\Phi^r$  takes the function values  $f(\mathbf{d}_{k,e})/(m+1)$  on the intervals  $T_{k,e}^q(\mathbf{d}_{k,e}^q)$ ,  $e \in \{1, 2\}$ , this induces large gradients for  $\Phi^r$ , and thus for  $\Phi$ , between the intervals, if the difference  $|f(\mathbf{d}_{k,1}) - f(\mathbf{d}_{k,2})|$  is large, see Figure 2.8. It is important to note that in general a large gradient is numerically disadvantageous since the variation has to be resolved by the discretization of the outer function  $\Phi$ . However, the difference  $|f(\mathbf{d}_{k,1}) - f(\mathbf{d}_{k,2})|$  is small if the function  $f : [0, 1]^n \rightarrow \mathbb{R}$  is smooth near the boundary  $\partial[0, 1]^n$  of the hypercube and vanishes on the boundary.

Thus, one might restrict the class of functions that are approximated to those functions which have zero boundary values. However, this assumption is very restrictive and we rather use the following strategy: Let  $f : [0, 1]^n \rightarrow \mathbb{R}$  be a continuous function with arbitrary boundary values. Then, we can avoid variations of the outer function which are due to the boundary values of  $f$  by transforming the approximation problem onto a sub cube  $\tilde{\Omega} := [\alpha, 1 - \alpha]^n \subsetneq [0, 1]^n$ ,  $0 < \alpha < 1$ , and enforcing zero boundary conditions on  $\partial[0, 1]^n$  on the transformed problem.

To explain this, let  $T : [0, 1]^n \rightarrow \tilde{\Omega}$  be a bijective mapping. We define  $\tilde{f} :$

$[0, 1]^n \rightarrow \mathbb{R}$  to be a continuous function such that

$$\tilde{f}|_{\tilde{\Omega}} := f \circ T^{-1}, \quad \tilde{f}|_{\partial[0,1]^n} := 0,$$

and  $\tilde{f}$  is smooth on  $[0, 1]^n \setminus \tilde{\Omega}$ . Practically, we use a linear continuation of  $f \circ T^{-1}$  onto  $[0, 1]^n$ .

Then we can approximate  $\tilde{f}$  from the data set  $\tilde{\mathcal{Z}} := \{(T\mathbf{x}_j, y_j) : (\mathbf{x}_j, y_j) \in \mathcal{Z}\}$  with the Regularization Network approach. Let  $f_{k,\mathbf{r}} \approx \tilde{f}$ ,  $k \in \{\ell, L\}$  denote the solution of Problem 4.18 or Problem 4.19, respectively. Note that  $T\mathbf{x} \in \tilde{\Omega}$  for all  $\mathbf{x} \in [0, 1]^n$ . Clearly, it holds:

$$f_{k,\mathbf{r}} \circ T \approx \tilde{f} \circ T = (f \circ T^{-1}) \circ T = f, \quad k \in \{\ell, L\},$$

and thus  $f_{k,\mathbf{r}} \circ T$ ,  $k \in \{\ell, L\}$  approximates  $f$ .

This transformation of the approximation problem to a sub cube  $\tilde{\Omega} \subsetneq [0, 1]^n$  will always be performed in our implementation unless it is a priori known that the function  $f : [0, 1]^n \rightarrow \mathbb{R}$  vanishes on the boundary of  $[0, 1]^n$ .

## 5.4 Further remarks

We finish the section with a remark on the parallelization of Problem 4.18 and Problem 4.19.

In Section 5.1 we pointed out that for the iterative nonlinear minimizers the evaluation of the error functional  $E^{(n)}$  and its gradient  $\nabla E^{(n)}$  dominates the numerical costs. This is amplified by the necessity to use multiple precision arithmetic for an evaluation of the inner sums  $\Psi_q$  and the basis functions  $\varphi_j(\Psi_q(\mathbf{x}))$ ,  $\varphi_j^d(\Psi_q(\mathbf{x}))$ , see Section 5.2. Thus, in both cases, i.e. in Problem 4.18 and Problem 4.19, a costly part is the computation of the respective data term. To be more precise, this is for the first problem the computation of  $\mathbf{B}^T \mathbf{B}$  in (4.46), and the calculation of the sums (5.5) in the second problem. All data terms require the evaluation of the basis functions and the inner sums for all points in  $\mathcal{X} := \{\mathbf{x}_1, \dots, \mathbf{x}_P\}$ .

Now, it is important to note that the data terms are in both cases sums which are built over all data points. This allows for a trivial parallelization of the problem: First, we split the data set into disjoint parts  $\mathcal{X} = \mathcal{X}_1 \cup \dots \cup \mathcal{X}_p$ , and then compute for each subset  $\mathcal{X}_j$ ,  $j = 1, \dots, p$ , the respective data term on a separate slave processor. Finally, the results are sent back to the master processor which adds them up. Additionally, one could think of a subdivision of the sums over all  $q = 0, \dots, m$  in a similar manner. However, this is not done in our implementation.

Finally, we remark that we use a direct solver from the *PETSc library*, see [139], to solve the system of linear equations (4.46) that arises in Problem 4.18. The library also provides an interface to a parallel solver for this task.

# Chapter 6

## Numerical results

In this section we show the results of numerical experiments for the Regularization Network approaches that have been introduced in Problem 4.18 and Problem 4.19.

To investigate the general properties of the respective model  $f_{\ell,r}$ ,  $f_{L,r}$  like convergence, dependency on model parameters, and regularization, it is advantageous to setup a problem for which the exact solution is known explicitly. To this end, we choose a continuous function  $f : [0, 1]^n \rightarrow \mathbb{R}$ , generate the set of learning points

$$\{\mathbf{x}_1, \dots, \mathbf{x}_P\}$$

by the quasi-random Halton sequence, see [90], and sample the known function  $f$  on this set. This choice is motivated by a statement from [30] which says that for lower space dimensions, this sequence quickly “fills up” the  $n$ -dimensional unit cube and that for higher dimensions (e.g.  $n = 40$ ) the point set has to be large. Note that in our examples we either consider low dimensions or increase the number of learning points if  $n$  becomes larger, and thus the Halton sequence is appropriate. Now, this gives the associated values  $y_j = f(\mathbf{x}_j)$ ,  $j = 1, \dots, P$ , and thus the training set

$$\mathcal{Z} := \{(\mathbf{x}_j, y_j) : j = 1, \dots, P\} \subset [0, 1]^n \times \mathbb{R},$$

see (4.1). For the so defined data set, we will investigate our Regularization Network approaches in the following sections. There, to evaluate the quality of the computed approximation we both, either directly evaluate the respective error functional  $E^{(1)}$ , (4.43), or  $E^{(n)}$ , (4.50), and then calculate the mean-square-error (MSE)

$$e_{\mathcal{Z}_e}(f_{k,r}) := \frac{1}{P_e} \sum_{j=1}^{P_e} \left( \tilde{y}_j - f_{k,r}(\tilde{\mathbf{x}}_j) \right)^2, \quad k \in \{\ell, L\}, \quad (6.1)$$

on the test set

$$\mathcal{Z}_e := \{(\tilde{\mathbf{x}}_j, \tilde{y}_j) : j = 1, \dots, P_e\} \subset [0, 1]^n \times \mathbb{R}.$$

Here, the set of new points  $\{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{P_e}\}$  is also generated by the Halton sequence and  $\tilde{y}_j = f(\tilde{\mathbf{x}}_j)$ ,  $j = 1, \dots, P_e$ . See Table A.4 in the appendix for a list of the values  $P_e$  depending on the dimension  $n$ . The choice of  $e_{\mathcal{Z}_e}$  is motivated by the loss function  $V(f(\mathbf{x}_i), y_i) = (f(\mathbf{x}_i) - y_i)^2$  from Chapter 4 which is used to measure closeness to the data in the Regularization Network approach. (6.1) is defined on a discrete data set and thus, to compute the error of the reconstruction, an explicit knowledge of the function  $f$  is not necessary if a test set  $\mathcal{Z}_e$  is given. Note that this is a random quadratic error.

In some examples it will be more instructive to consider the maximal error on the test set  $\mathcal{Z}_e$  which is defined for  $k \in \{\ell, L\}$  by

$$e_{\mathcal{Z}_e}^\infty(f_{k,r}) := \max_{(\mathbf{x}, y) \in \mathcal{Z}_e} |y - f_{k,r}(\mathbf{x})|. \quad (6.2)$$

In the following, we will investigate the respective model for different problem parameters, i.e. we choose different functions  $f$ , dimensions  $n$ , and cardinalities  $P$  of the data set. Additionally, we analyze the dependency on the regularization parameters  $\nu$  and further model parameters.

However, note that for all examples throughout this section we choose the minimal basis  $\gamma = m + 2$  in Theorem 2.14 to define the inner function  $\psi$  if not stated otherwise. See the remarks at the end of Section 2.3.5 for this choice.

## 6.1 First model

This section will show numerical results for the solution of Problem 4.18. They will provide important insights into the structure of the outer function  $\Phi$  in (2.36) which will then serve as a motivation for the definition of the second model.

To analyze  $f_{\ell,r}$ , we will first investigate the choice of the basis functions which span the approximation space  $\mathcal{H}_\ell^0$  in (4.39) by means of two-dimensional functions  $f : [0, 1]^2 \rightarrow \mathbb{R}$ .

### 6.1.1 B-splines

For the first numerical calculations we will use B-splines of the order  $k$  to construct the approximation space  $\mathcal{H}_\ell^0$  in (4.39). Note that this is closely related to the ideas of Coppejans, Igelnik and Parikh, [17, 62, 64] who used splines to approximate both, outer and inner functions in Kolmogorov's representation.

We introduce the necessary definitions very briefly and refer to e.g. [21] for details. A B-spline of order  $k \in \mathbb{N}$  is a function

$$\mathcal{S}_k(x) := \sum_{j=0}^{\ell} c_j \mathcal{B}_{k,j}(x).$$

Here, the elementary B-splines  $\mathcal{B}_{k,j}$ ,  $j = 0, \dots, \ell$  of order  $k$  are defined for  $I + 1$  given equidistant<sup>1</sup> supporting points  $a_0 = t_0 < \dots < t_I = b_m$ ,  $I = \ell - k + 2$  on the nodes

$$\tau_j := \begin{cases} t_0, & j = 0, \dots, k-1, \\ t_{j-k+1}, & j = k, \dots, \ell, \\ t_I, & j = \ell+1, \dots, \ell+k. \end{cases}$$

Namely, they are defined by the recursive formula

$$\mathcal{B}_{k,j}(x) := \frac{x - \tau_j}{\tau_{j+k-1} - \tau_j} \mathcal{B}_{k-1,j}(x) + \frac{\tau_{j+k} - x}{\tau_{j+k} - \tau_{j+1}} \mathcal{B}_{k-1,j+1}(x),$$

where the initial elementary B-splines of order 1 are the piecewise constants

$$\mathcal{B}_{1,j}(x) := \begin{cases} 1, & x \in [\tau_j, \tau_{j+1}), \\ 0, & \text{else,} \end{cases} \quad j = 0, \dots, \ell.$$

The function  $\mathcal{S}_k$  is a piecewise polynomial of degree  $k-1$ , and for  $k \geq 2$  it holds that  $\mathcal{S}_k \in H^1([a_0, b_m])$ . Thus, we can define the finite dimensional approximation space  $\mathcal{H}_\ell^0$  in (4.39) to be given by

$$\mathcal{H}_{\ell,k}^0 := \text{span} \{ \mathcal{B}_{k,j} : j = 0, \dots, \ell \},$$

for fixed  $k \geq 2$ , and supporting points  $a_0 = t_0 < \dots < t_I = b_m$ . To make this a Hilbert space we endow it with the scalar product

$$\langle \phi_1, \phi_2 \rangle_{\mathcal{H}^0} := \int_{a_0}^{b_m} \phi_1(t) \phi_2(t) dt + \int_{a_0}^{b_m} \phi'_1(t) \phi'_2(t) dt, \quad (6.3)$$

which is simply the scalar product in the Sobolev space  $H^1([a_0, b_m])$ . The space  $(\mathcal{H}_{\ell,k}^0, \|\cdot\|_{\mathcal{H}^0})$  can be continuously embedded into  $(\mathcal{C}^0([a_0, b_m]), \|\cdot\|_\infty)$ , which shows that it is a RKHS and the constructions from Section 4.5 are applicable to construct the  $n$ -dimensional approximation space  $\mathcal{H}_{\ell,k}^{(1)} = \mathcal{H}_\ell^{(1)}$  as defined in (4.41). Here, we use the additional subscript  $k$  to indicate the dependency on the order of the splines. Now, to guarantee a continuous embedding, it would suffice to take the  $H^s$ -norm with some  $s > 1/2$ . But we also remark that, although the elementary B-splines  $\mathcal{B}_{k,j}$  are not orthogonal with respect to  $\langle \cdot, \cdot \rangle_{\mathcal{H}^0}$ , the norm of a function  $f_\ell \in \mathcal{H}_{\ell,k}^{(1)}$  can be computed with (4.42) by

$$\| f_\ell \|_{\mathcal{H}(K)}^2 = \left\| \sum_{j=0}^{\ell} c_j \mathcal{B}_{k,j} \right\|_{\mathcal{H}^0}^2.$$

---

<sup>1</sup>Note that for this definition the supporting points  $t_0, \dots, t_I$  in general do not have to be equidistant.

Thus, it is more practicable to use the  $H^1$ -norm in combination with the B-splines for computational reasons. Note that this is a real valued construction, see Remark 4.21.

Now, a minimizer in Problem 4.18 is given by the solution  $\mathbf{c}_\ell^{(1)} := (c_0, \dots, c_\ell)^T \in \mathbb{R}^{\ell+1}$  of the system of linear equations (4.46) with data matrix

$$\mathbf{B} := \sum_{q=0}^m \begin{pmatrix} \mathcal{B}_{k,0}(\Psi_q(\mathbf{x}_1)) & \dots & \mathcal{B}_{k,\ell}(\Psi_q(\mathbf{x}_1)) \\ \vdots & & \vdots \\ \mathcal{B}_{k,0}(\Psi_q(\mathbf{x}_P)) & \dots & \mathcal{B}_{k,\ell}(\Psi_q(\mathbf{x}_P)) \end{pmatrix} \in \mathbb{R}^{P \times (\ell+1)},$$

regularization matrix

$$\mathbf{S} := \begin{pmatrix} \langle \mathcal{B}_{k,0}, \mathcal{B}_{k,0} \rangle_{\mathcal{H}^0} & \dots & \langle \mathcal{B}_{k,0}, \mathcal{B}_{k,\ell} \rangle_{\mathcal{H}^0} \\ \vdots & & \vdots \\ \langle \mathcal{B}_{k,\ell}, \mathcal{B}_{k,0} \rangle_{\mathcal{H}^0} & \dots & \langle \mathcal{B}_{k,\ell}, \mathcal{B}_{k,\ell} \rangle_{\mathcal{H}^0} \end{pmatrix} \in \mathbb{R}^{(\ell+1) \times (\ell+1)},$$

and data vector  $\mathbf{y} = (y_1, \dots, y_P)^T \in \mathbb{R}^P$ . We remark that, due to the fact that  $\text{supp}(\mathcal{B}_{k,j}) = [\tau_j, \tau_{j+k}]$ , these matrices have limited bandwidth and  $\mathcal{O}(\ell k)$  non-zero entries.

### Simple example

We start with a simple two-dimensional numerical example. Similar to the constructions in Section 4.5 we first choose an outer function  $\phi$  in Kolmogorov's representation (2.36) to define the function  $f_\phi$ . To be precise, we select a polynomial  $\tilde{p}$  of degree four<sup>2</sup>

$$\tilde{p}(t) := \frac{1337}{60} t - \frac{6481}{60} t^2 + \frac{4891}{30} t^3 - \frac{386}{5} t^4,$$

and consider for  $\mathbf{x} \in [0, 1]^2$  the function

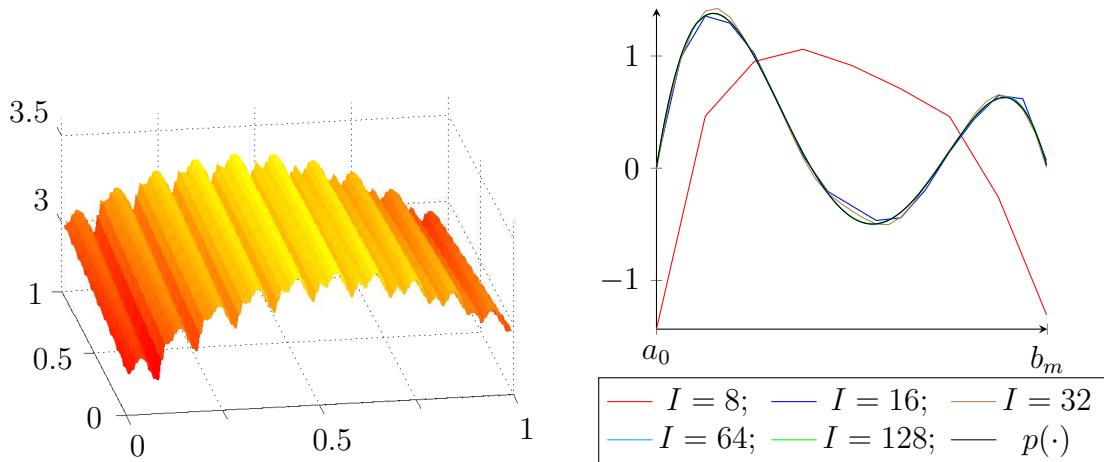
$$f_p(\mathbf{x}) := \sum_{q=0}^8 p(\Psi_q(\mathbf{x})), \quad \text{with} \quad p(t) := \tilde{p}(a_0 + t(b_m - a_0)). \quad (6.4)$$

This way we obtain a somewhat artificial, wiggly function  $f_p$  that however has the advantage of a smooth outer function  $p$ . The graph of  $f_p$  is shown in Figure 6.1 (left).

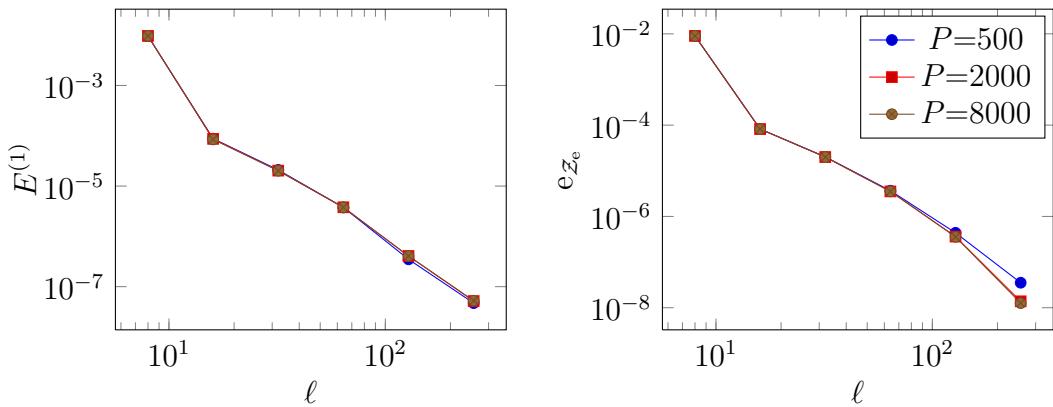
We reconstructed  $f_p$  from  $P = 500, 2000$  and  $8000$  data points with B-splines of order  $k = 2$ , i.e. piecewise linear functions, for increasing resolutions  $I =$

---

<sup>2</sup> $\tilde{p}$  is the solution of the interpolation problem:  $\tilde{p}(0) = \tilde{p}(1) = 0$ ,  $\tilde{p}(1/3) = 1/2$ ,  $\tilde{p}(1/2) = -1/3$ , and  $\tilde{p}(3/4) = 1/4$ .



**Figure 6.1.** Graph of  $f_p$  from (6.4) (left) and approximations to  $p$  from (6.4) (right) for different values of  $I$ .



**Figure 6.2.** Result for the reconstruction of  $f_p$  from (6.4) with  $P = 500, 2000$ , and  $8000$  data points and the simple model (3.5). The plots show the value of the energy functional  $E^{(1)}$  with  $\nu = 10^{-9}$  and the MSE depending on  $\ell = I + k - 2$ .

8, 16, ..., 256. Clearly, we chose  $m = 8$  in (4.35), while the regularization parameter in (4.43) was  $\nu = 10^{-9}$ . The right plot in Figure 6.1 shows the outer functions that were computed from  $P = 2000$  points for the resolutions  $I = 8, 16, 32, 64, 128$  (dotted lines) and the outer function  $p$  from (6.4) (solid line). One can see convergence towards the true function  $p$ , i.e. for an increasing value of  $I$  the deviation of the approximation decreases. For resolutions larger than  $I = 128$  no difference between the reconstruction and  $p$  is visible.

Figure 6.2 shows the convergence plots for this example. The value of the error

	$P = 500$		$P = 2000$		$P = 8000$	
$\ell$	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope
8	$8.99 \cdot 10^{-3}$	–	$9.01 \cdot 10^{-3}$	–	$9.01 \cdot 10^{-3}$	–
16	$8.23 \cdot 10^{-5}$	-6.77	$8.20 \cdot 10^{-5}$	-6.78	$8.22 \cdot 10^{-5}$	-6.78
32	$2.02 \cdot 10^{-5}$	-2.03	$2.00 \cdot 10^{-5}$	-2.04	$2.00 \cdot 10^{-5}$	-2.04
64	$3.66 \cdot 10^{-6}$	-2.46	$3.53 \cdot 10^{-6}$	-2.50	$3.54 \cdot 10^{-6}$	-2.50
128	$4.38 \cdot 10^{-7}$	-3.07	$3.62 \cdot 10^{-7}$	-3.29	$3.57 \cdot 10^{-7}$	-3.31
256	$3.53 \cdot 10^{-8}$	-3.63	$1.38 \cdot 10^{-8}$	-4.71	$1.26 \cdot 10^{-8}$	-4.82

**Table 6.1.** Values of the MSE  $e_{\mathcal{Z}_e}$  from (6.1) for the reconstruction of  $f_p$  with B-splines of the order  $k = 2$  from  $P = 500, 2000, 8000$  data points and regularization parameter  $\nu = 10^{-9}$ . Here, the slope is computed between two successive points in the right graph of Figure 6.2.

functional (4.43) (left) and the mean-square-error  $e_{\mathcal{Z}_e}$  from (6.1) (right) are plotted against  $\ell = I + k - 2$ . Additionally, in Table 6.1 the values of  $e_{\mathcal{Z}_e}$ , and the slopes between two successive points in the MSE graph are given.

In all cases  $P = 500, 2000, 8000$  we observe a rapid decay of  $E^{(1)}$  and  $e_{\mathcal{Z}_e}$  for increasing resolution with similar rates. This is due to the fact that the outer function  $p$  is smooth. We remark that the poor approximation of  $p$  in Figure 6.1 for  $I = 8$  is also visible from the values of  $E^{(1)}$  and  $e_{\mathcal{Z}_e}$ , which measure the error of  $f_p$ .

### General example

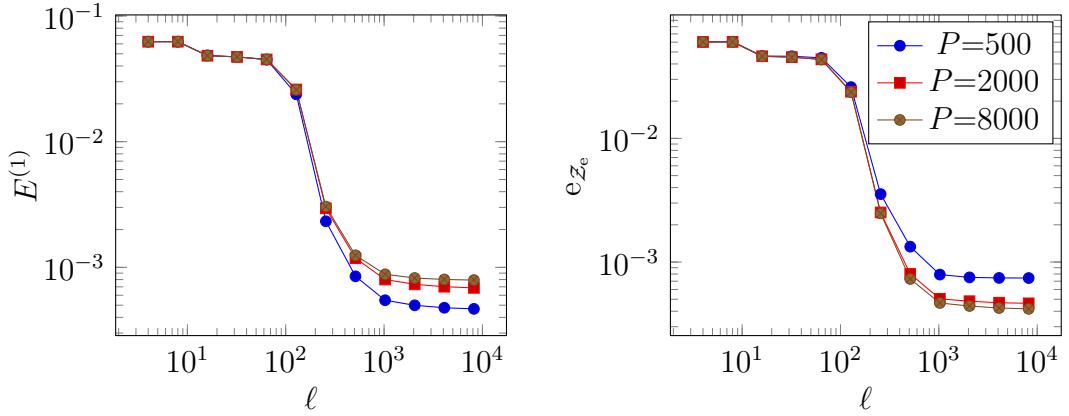
The next example deals with the original problem of reconstructing an unknown function  $f : [0, 1]^2 \rightarrow \mathbb{R}$  from discrete data points without knowledge of the outer function  $\Phi : [a_0, b_m] \rightarrow \mathbb{R}$ . We choose

$$f(\mathbf{x}) := \sin(\pi x_1) \sin(\pi x_2), \quad \mathbf{x} \in [0, 1]^2$$

in our setup to generate the data sets  $\mathcal{Z}$  and  $\mathcal{Z}_e$ .

Again, we reconstructed the function from  $P = 500, 2000, 8000$  data points with B-splines of order  $k = 2$ . For each data set and  $j = 2, \dots, 13$  we tested the spline resolution  $I = 2^j$ , and chose  $\nu = 10^{-6}$  to weight the regularization term in  $E^{(1)}$ . Similar to the previous examples we used  $m = 8$  terms in (4.35). Table 6.2 shows the values of the MSE of the reconstructions. In Figure 6.3,  $E^{(1)}$  and  $e_{\mathcal{Z}_e}$  are plotted against the value  $\ell = I + k - 2 = I$ .

In this example the decay of  $E^{(1)}$  and  $e_{\mathcal{Z}_e}$  substantially differs from the previous artificial example where the errors considerably decreased with any increase of  $\ell$ . Here, we observe a reduction in two steps for both, error functional and MSE: First,



**Figure 6.3.** The error functional  $E^{(1)}$  (left) and the MSE  $e_{\mathcal{Z}_e}$  (right) are plotted against the value  $\ell = I$ . The total number of DOF is  $\ell + 1$ .

$\ell$	$P = 500$		$P = 2000$		$P = 8000$	
	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope
4	$6.03 \cdot 10^{-2}$	–	$6.02 \cdot 10^{-2}$	–	$6.02 \cdot 10^{-2}$	–
8	$6.06 \cdot 10^{-2}$	+0.01	$6.03 \cdot 10^{-2}$	+0.00	$6.02 \cdot 10^{-2}$	+0.00
16	$4.63 \cdot 10^{-2}$	-0.39	$4.62 \cdot 10^{-2}$	-0.39	$4.61 \cdot 10^{-2}$	-0.39
32	$4.62 \cdot 10^{-2}$	+0.00	$4.54 \cdot 10^{-2}$	-0.02	$4.53 \cdot 10^{-2}$	-0.02
64	$4.48 \cdot 10^{-2}$	-0.04	$4.36 \cdot 10^{-2}$	-0.06	$4.34 \cdot 10^{-2}$	-0.06
128	$2.60 \cdot 10^{-2}$	-0.79	$2.38 \cdot 10^{-2}$	-0.87	$2.36 \cdot 10^{-2}$	-0.87
256	$3.54 \cdot 10^{-3}$	-2.87	$2.52 \cdot 10^{-3}$	-3.24	$2.47 \cdot 10^{-3}$	-3.26
512	$1.33 \cdot 10^{-3}$	-1.41	$8.02 \cdot 10^{-4}$	-1.65	$7.30 \cdot 10^{-4}$	-1.76
1 024	$7.92 \cdot 10^{-4}$	-0.75	$5.04 \cdot 10^{-4}$	-0.67	$4.67 \cdot 10^{-4}$	-0.65
2 048	$7.52 \cdot 10^{-4}$	-0.08	$4.82 \cdot 10^{-4}$	-0.06	$4.41 \cdot 10^{-4}$	-0.08
4 096	$7.44 \cdot 10^{-4}$	-0.01	$4.69 \cdot 10^{-4}$	-0.04	$4.25 \cdot 10^{-4}$	-0.05
8 192	$7.42 \cdot 10^{-4}$	+0.00	$4.64 \cdot 10^{-4}$	-0.02	$4.18 \cdot 10^{-4}$	-0.02

**Table 6.2.** MSE values depending on the value  $\ell$ . Note that  $\ell = I$  is also the resolution of the B-splines of order  $k = 2$ . Again, the slopes have been computed for successive points in the right graph of Figure 6.3.

for the values  $I = 4, \dots, 64$  the errors are diminished only by a small amount. Then, they are reduced with increasing resolutions  $I = 128, \dots, 1024$  until they reach their minimal values. Finally, error functional and MSE cannot be further improved by refinements of the spline resolutions. This saturation is due to the limited information on  $f$  which is given in the form of the finite data set  $\mathcal{Z}$ : For

larger values of  $P$  the minimal errors decrease. Altogether, we conclude that a minimal resolution  $I$  is needed to approximate the outer function  $\Phi$ .

### Boundary values

In the next tests we numerically discuss the treatment of the boundary values of the function  $f$  as introduced in Section 5.3. To this end, we compare reconstructions of several two-dimensional functions which do not vanish at the boundary of  $[0, 1]^2$  with and without a previous transformation of the problem to a sub cube. Here, the sub cube  $\tilde{\Omega}$  is chosen individually for each function  $f$  that has to be approximated. The test functions and their associated sub cubes are listed in Table 6.3. There, to evaluate the results, the maximal error on the test set (6.2) is given. This choice is more instructive since  $e_{\mathcal{Z}_e}^\infty$  detects point wise errors at the boundary.

For all reconstructions we used  $m = 8$  terms in (4.35), while the regularization parameter  $\nu$  differs for each problem, see Table 6.3. The training sets  $\mathcal{Z}$  were generated from  $P = 8000$  point samples of the respective function  $f$ . Additionally, to ensure a sufficient resolution of the B-splines of order  $k = 2$  in all computations we only considered the case  $I = 8192$ . For each function, Table 6.3 lists the maximal error without and with transformation, the sub cube  $\tilde{\Omega}$ , and the regularization parameter  $\nu$ . Furthermore, for all cases the graphs of the reconstructions  $f_\ell$  and the residuals  $f - f_\ell$  are plotted in figures 6.4–6.7. Figure 6.8 shows the corresponding outer functions  $\Phi$ .

In all tests, the error  $e_{\mathcal{Z}_e}^\infty$  could be reduced by the transformation of the problem to a sub cube  $\tilde{\Omega}$ . Furthermore, from figures 6.4–6.7 it becomes obvious that this effect is indeed due to the values of  $f$  at the boundaries: Without transformation, all residuals take their maximum or minimum near the boundary of  $[0, 1]^2$  while the error is small inside the domain. With the transformation to  $\tilde{\Omega}$  this domination could be eliminated what finally resulted in smaller errors  $e_{\mathcal{Z}_e}^\infty$ .

For the respective outer functions which are shown in Figure 6.8 we observe that the range of maximum and minimum becomes smaller if the problem is transformed to the sub cube (red graphs). Hence, the largest gradients of the functions  $\Phi$  from the non-transformed problems (blue graphs) could be diminished, see the first and the last image in Figure 6.8.

However, an important property which becomes visible here is that all outer functions strongly oscillate. Thus, to get an approximation of the outer function, the basis has to resolve very high frequencies, and for this reason, a minimal resolution of the B-splines is necessary to reduce the error as it was observed in a previous example.

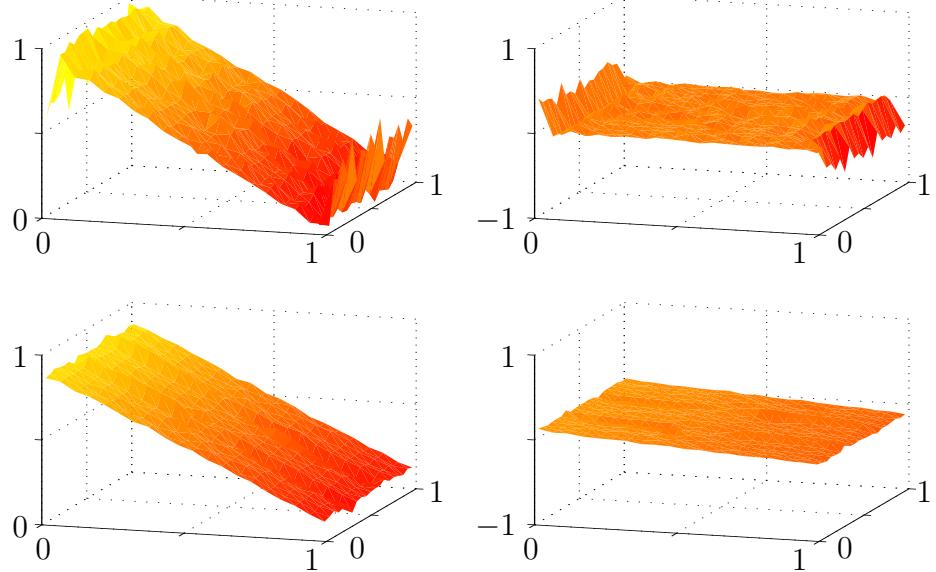
Now, for oscillating functions it is obvious to consider their Fourier transform. To this end, we computed the discrete Fourier transforms<sup>3</sup>  $\widehat{\Phi}_d(k)$  for all outer functions

---

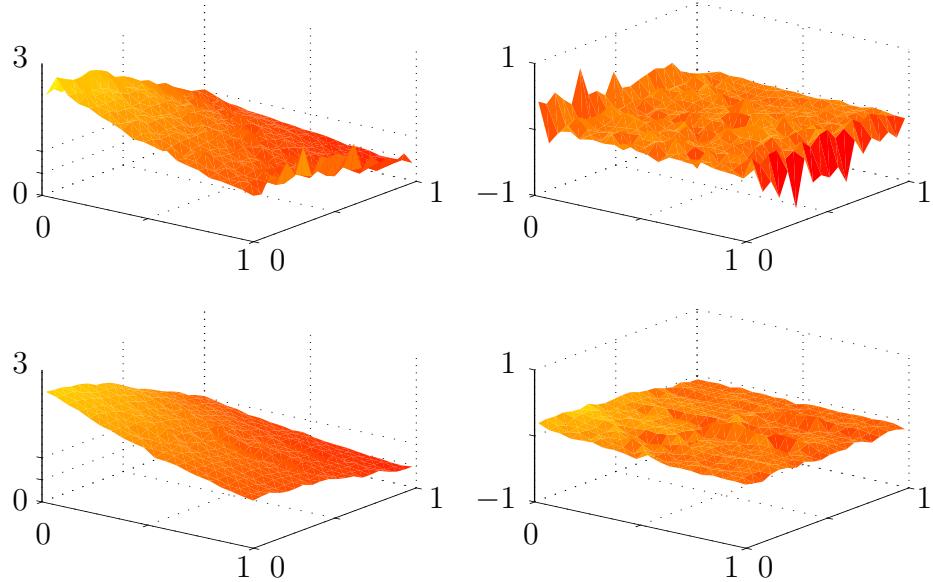
<sup>3</sup>For equidistant nodes  $t_k := k\omega/M$ ,  $M \in \mathbb{N}$ ,  $\omega \in \mathbb{R}$ , and complex function values

$f(x_1, x_2)$	$e_{Z_e}^\infty$ without transformation	sub cube $\tilde{\Omega}$	$\nu$	$e_{Z_e}^\infty$ with transformation
$x_2$	$4.97 \cdot 10^{-1}$	$[0.1, 0.9]^2$	$10^{-7}$	$1.53 \cdot 10^{-1}$
$\exp(-x_1^2 + x_2)$	$8.17 \cdot 10^{-1}$	$[0.05, 0.95]^2$	$10^{-8}$	$2.2 \cdot 10^{-1}$
$\tan\left(1 - \frac{1}{0.2x_1 + 0.3x_2}\right)$	$1.77 \cdot 10^{-1}$	$[0.1, 0.9]^2$	$10^{-7}$	$8.32 \cdot 10^{-2}$
$\prod_{d=1}^2 (x_d - \frac{1}{2}) \log( x_d - \frac{1}{2} )$	$1.05 \cdot 10^{-1}$	$[0.1, 0.9]^2$	$10^{-7}$	$2.68 \cdot 10^{-2}$

**Table 6.3.** The table shows the maximal errors  $e_{Z_e}^\infty$  for the reconstructions of the respective function  $f(x_1, x_2)$  that were obtained without and with a transformation of the problem to the sub cube  $\tilde{\Omega}$ , see Section 5.3. For the values of the further parameters see figures 6.4, 6.5, 6.6, and 6.7. All functions do not vanish at the boundary.



**Figure 6.4.** Reconstructions  $f_\ell$  (left) of  $f(x_1, x_2) = x_2$  from  $P = 8000$  data points with  $I = 8192$ ,  $m = 8$ , and  $\nu = 10^{-7}$  without (top) and with (bottom) transformation to a sub cube  $\tilde{\Omega}$ . The right column shows the respective residuals  $f - f_\ell$ .



**Figure 6.5.** Here,  $f(x_1, x_2) = \exp(-x_1^2 + x_2)$  was reconstructed with the parameters  $P = 8000$ ,  $I = 8192$ ,  $m = 8$ ,  $\nu = 10^{-8}$ . The plots show the respective graphs of  $f_\ell$  (left) and  $f - f_\ell$  (right) as computed without (top) and with (bottom) transformation.

$\Phi$  that resulted from the transformed problems. In Figure 6.9 the modulus  $|\widehat{\Phi}_d(k)|$  of the respective outer functions are plotted. The underlying function  $f$  is given in each picture.

For all functions  $f$  we observe that the vector of Fourier coefficients of the corresponding outer function  $\Phi$  is sparse, i.e. many coefficients are nearly zero.

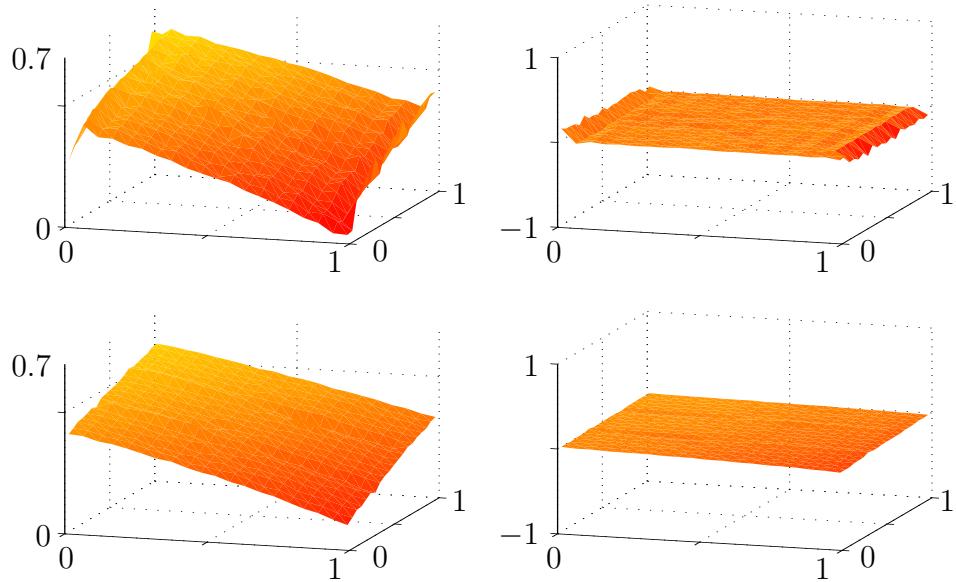
Obviously, the use of the Fourier basis is advantageous to represent and resolve oscillating functions, while basis functions with local support like B-splines have the problem that the mesh resolution has to be very fine to resolve the frequencies.

### 6.1.2 Trigonometric polynomials

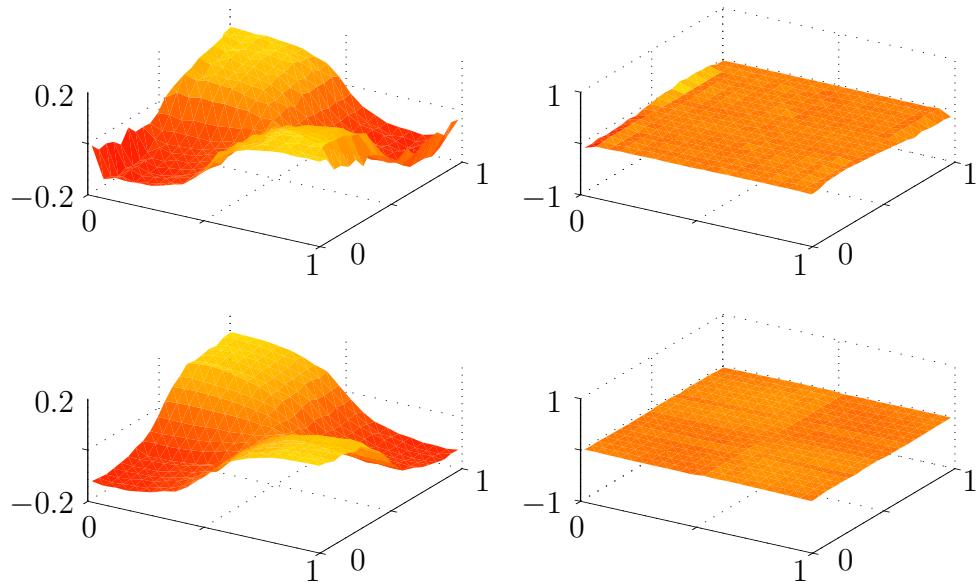
We have seen in the previous section that the outer function  $\Phi$  which is computed in Problem 4.18 typically oscillates. Consequently, an approximation with basis functions which have local support is inefficient. This is due to the fact that the spatial resolution of an expansion in such a basis has to capture the highest oscillations.

---

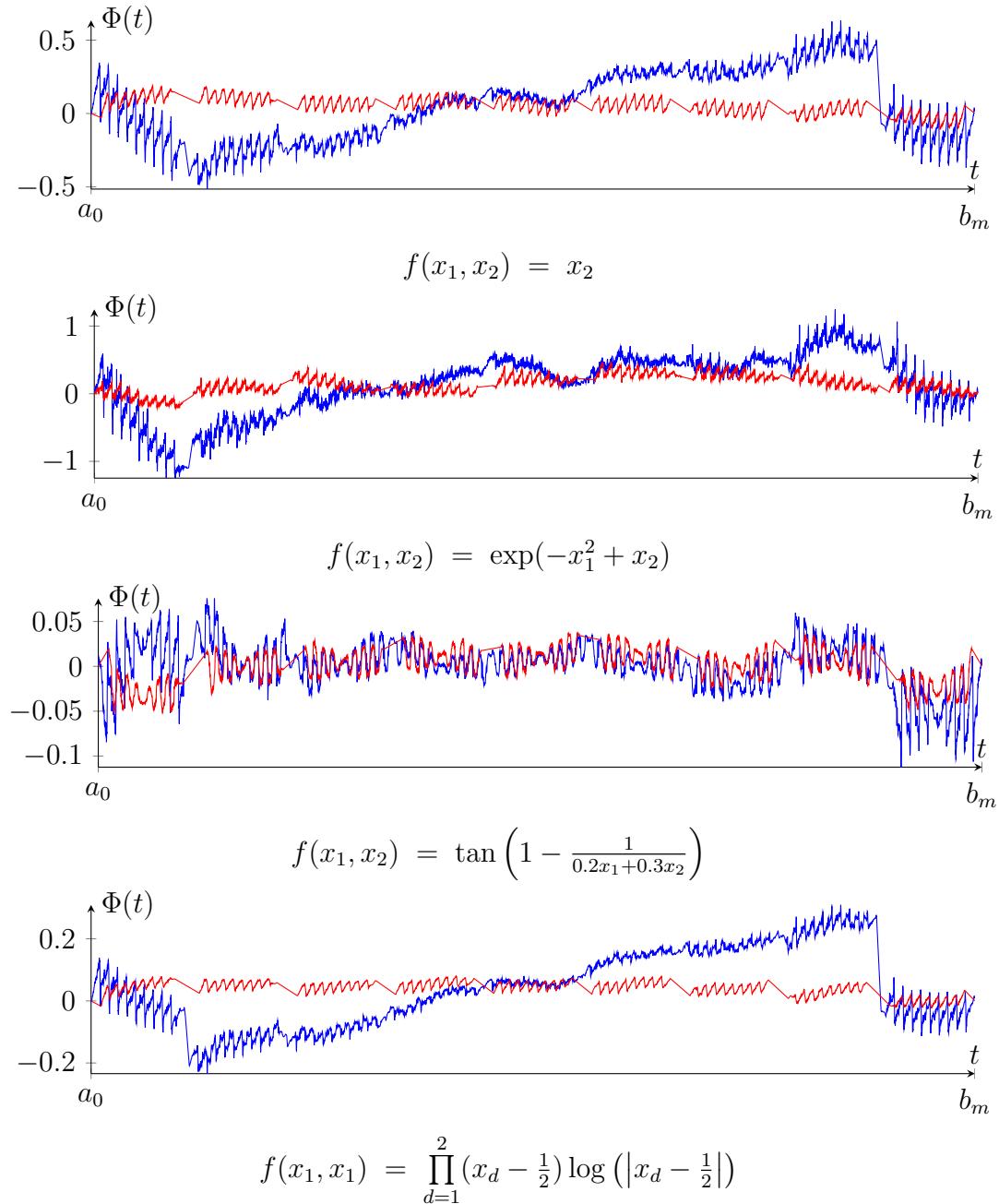
$\phi(t_0), \dots, \phi(t_{M-1})$ , the discrete Fourier transform of an  $\omega$ -periodic function  $\phi : \mathbb{R} \rightarrow \mathbb{C}$  is given by  $\widehat{\phi}_d(k) = \sum_{j=0}^{M-1} \phi(t_j) \exp\left(-\frac{2\pi}{M} i k j\right)$ ,  $k = -M/2, \dots, M/2 - 1$ . Conversely, for  $j = 0, \dots, M - 1$  it holds  $\phi(t_j) = \frac{1}{M} \sum_{k=-M/2}^{M/2-1} \widehat{\phi}_d(k) \exp\left(\frac{2\pi}{\omega} i k t_j\right)$ .



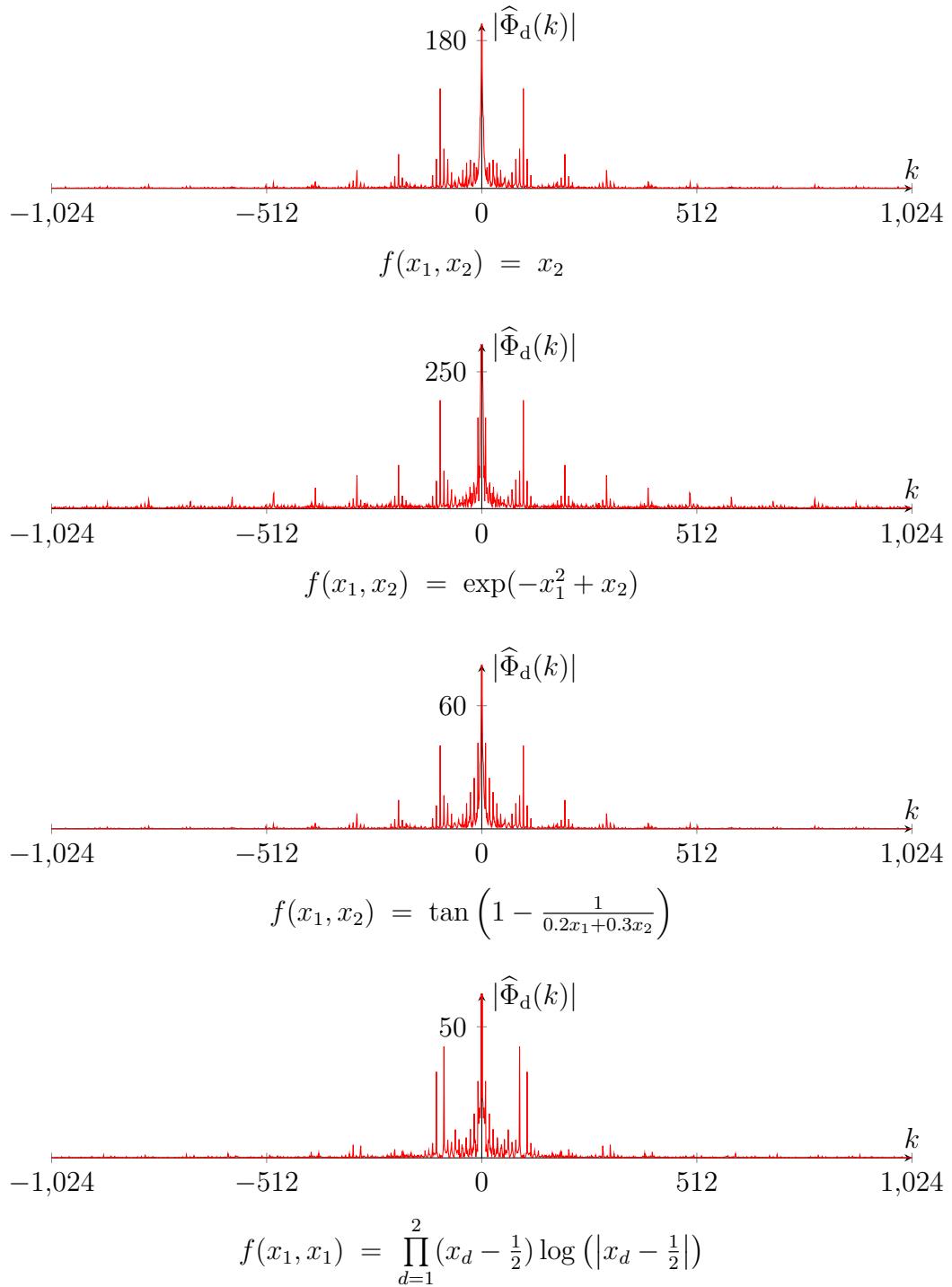
**Figure 6.6.** Results for  $f(x_1, x_2) = \tan(1 - (0.2x_1 + 0.3x_2)^{-1})$  with  $P = 8000$ ,  $I = 8192$ ,  $m = 8$ , and  $\nu = 10^{-7}$ . The left column shows  $f_\ell$  and the right column  $f - f_\ell$ . The top row is the reconstruction without transformation while the bottom row shows the result from the transformed problem.



**Figure 6.7.** Solution of the regular and the transformed problem for  $f(x_1, x_2) = \prod_{d=1}^2 (x_d - \frac{1}{2}) \log(|x_d - \frac{1}{2}|)$  from  $P = 8000$  data points and  $I = 8192$ ,  $m = 8$ ,  $\nu = 10^{-7}$ . The pictures are arranged as previously.



**Figure 6.8.** Outer functions  $\Phi$  for a reconstruction of the respective function  $f$  from  $P = 8000$  data points and  $I = 8192$ . Here, there computations were made without (blue) and with (red) a previous transformation of the problem to a sub cube  $\tilde{\Omega} \subsetneq [0, 1]^2$ .



**Figure 6.9.** Absolute values of the Fourier coefficients  $|\widehat{\Phi}_d(k)|$  for the outer functions  $\Phi$  from Figure 6.8. The different pictures show the results for the transformed problems and the respective functions  $f$ . Note that for real valued functions it holds  $\widehat{\Phi}_d(-k) = \overline{\widehat{\Phi}_d(k)}$ .

We will see in the following that the oscillations are even amplified if the dimension increases.

However, the vector of discrete Fourier coefficients of  $\Phi$  turned out to be sparse. We remark that the discrete Fourier coefficients  $\widehat{\Phi}_d(k)$  are approximations to the Fourier coefficients of  $\Phi$ , see [13]. Thus, it is obvious to use trigonometric polynomials as basis functions and directly calculate the Fourier coefficients within the Regularization Network approach.

To explain this, we define  $\omega_m := b_m - a_0$  and assume without loss of generality that each outer function  $\Phi$  is an  $\omega_m$ -periodic function  $\Phi : \mathbb{R} \rightarrow \mathbb{C}$ . Then,  $\Phi$  can be represented by means of its Fourier series expansion which will be briefly recalled in the following. For more details on Fourier analysis we refer to [13, 60]. At this point the constructions in Section 5.3 are essential, since otherwise the assumption of periodicity would not hold.

The Fourier series expansion of an  $\omega_m$ -periodic function  $\phi : \mathbb{R} \rightarrow \mathbb{C}$  is given by

$$\phi(t) = \sum_{k=-\infty}^{\infty} \widehat{\phi}(k) \exp\left(\frac{2\pi}{\omega_m} i k t\right), \quad (6.5)$$

and the Fourier coefficients are defined for  $k \in \mathbb{N}$  by

$$\widehat{\phi}(k) := \int_0^{\omega_m} \phi(s) \exp\left(-\frac{2\pi}{\omega_m} i k s\right) ds \in \mathbb{C}. \quad (6.6)$$

Moreover, for real valued functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  it holds

$$\widehat{\phi}(-k) = \overline{\widehat{\phi}(k)}, \quad \text{which implies} \quad |\widehat{\phi}(-k)| = |\widehat{\phi}(k)|. \quad (6.7)$$

Now, to compute the Fourier coefficients  $\widehat{\phi}(k)$  directly within the Regularization Network approach, let  $N \in \mathbb{N}$  and  $\ell = 2N$ . Then, with a shift of the index set  $\{0, \dots, 2N\} \mapsto \{-N, \dots, N\}$  in (4.38) we can define

$$\mathcal{F}_\ell^0 := \mathcal{F}_N^0 := \{\varphi_{-N}^0, \dots, \varphi_N^0\}$$

with basis functions

$$\varphi_k^0(t) := \exp\left(\frac{2\pi}{\omega_m} i k t\right), \quad \omega_m := b_m - a_0, \quad k = -N, \dots, N. \quad (6.8)$$

Remember that in (4.39) we defined  $\mathcal{H}_\ell^0 := \text{span } \mathcal{F}_\ell^0$  and that a function  $\phi_* \in \mathcal{H}_\ell^0$  has the representation

$$\phi_*(t) = \sum_{k=-N}^N c_k \exp\left(\frac{2\pi}{\omega_m} i k t\right), \quad c_k \in \mathbb{C}.$$

The Fourier coefficients (6.6) of this function are given by

$$\widehat{\phi}_*(k) = \begin{cases} c_k, & k = -N, \dots, N, \\ 0, & \text{else.} \end{cases}$$

This shows that the choice of the basis  $\mathcal{F}_N^0$  allows for a direct computation of the Fourier coefficients  $\widehat{\phi}_*(k)$  for  $k = -N, \dots, N$ . Note that the index set is chosen such that the index  $k$  of  $c_k$  is simply its location in Fourier space. Next, we remember that for real valued functions (6.7) holds, and we a priori assume for  $\phi_*$  that

$$c_{-k} = \bar{c}_k, \quad \text{for } k = 0, \dots, N. \quad (6.9)$$

This defines the real valued function

$$\phi_*(t) = \phi(t) + \overline{\phi}(t) = 2\phi_r(t),$$

with

$$\phi(t) := \frac{c_0}{2} + \sum_{k=1}^N c_k \exp\left(\frac{2\pi}{\omega_m} i k t\right).$$

Conversely, we can compute  $\phi_r$  by determining the coefficients  $c_0, \dots, c_N$  and then derive the remaining coefficients  $c_{-N}, \dots, c_{-1}$  by (6.9).

Note that the complex valued constructions in Section 4.5 allow for this choice of basis, and that the functions (6.8) form an orthogonal basis with respect to the scalar product from (6.3):

$$\int_{a_0}^{b_m} \varphi_j^0(t) \varphi_k^0(t) dt + \int_{a_0}^{b_m} (\varphi_j^0(t))' (\varphi_k^0(t))' dt = \delta_{j,k} (b_m - a_0) \left(1 + \left(\frac{2\pi}{\omega_m}\right)^2 |k|^2\right).$$

Furthermore, we remark that in this basis the computation of the  $H^s$ -norm for arbitrary  $s \in \mathbb{R}$  is easy and can be accomplished by changing the scalar product for  $j, k = -N, \dots, N$  to

$$\langle \varphi_j^0, \varphi_k^0 \rangle_{\mathcal{H}^0} := \delta_{j,k} \gamma_k^0, \quad \text{with} \quad \gamma_k^0 := (b_m - a_0) \left(1 + \left(\frac{2\pi}{\omega_m}\right)^2 |k|^2\right)^s. \quad (6.10)$$

Then, for  $s > 1/2$  the Sobolev embedding theorem guarantees that the embedding  $(\mathcal{H}_\ell^0, \|\cdot\|_{\mathcal{H}^0}) \hookrightarrow (\mathcal{C}^0([a_0, b_m]), \|\cdot\|_\infty)$  is continuous and the constructions from Section 4.5 apply. Numerical experiments have shown that the choice  $s = 1/2$  does not lead to additional instabilities, although theoretically we have to assume  $s > 1/2$ . For numerical simplicity, we will take  $s = 1/2$  in our computations. The space  $\mathcal{H}_\ell^0$  consists of  $\omega_m$ -periodic continuous functions.

The data matrix which is given by (4.47) can be computed by applying Euler's formula

$$\varphi_k^0(t) = \exp\left(\frac{2\pi}{\omega_m} i k t\right) = \cos\left(\frac{2\pi}{\omega_m} k t\right) + i \sin\left(\frac{2\pi}{\omega_m} k t\right) = \varphi_{k,\text{r}}^0(t) + i \varphi_{k,\text{i}}^0(t),$$

which leads to

$$\mathbf{B} := \sum_{q=0}^m \begin{pmatrix} 1 & \dots & \cos\left(\frac{2\pi}{\omega_m} N \Psi_q(\mathbf{x}_1)\right) & \sin\left(\frac{2\pi}{\omega_m} N \Psi_q(\mathbf{x}_1)\right) \\ \vdots & & & \vdots \\ 1 & \dots & \cos\left(\frac{2\pi}{\omega_m} N \Psi_q(\mathbf{x}_P)\right) & \sin\left(\frac{2\pi}{\omega_m} N \Psi_q(\mathbf{x}_P)\right) \end{pmatrix}.$$

Due to the global support of the functions  $\varphi_{k,\text{r}}^0$  and  $\varphi_{k,\text{i}}^0$  it is densely populated, but the regularization matrix is the diagonal matrix  $\mathbf{S} := \text{diag}(\gamma_0^0, \dots, \gamma_N^0, \gamma_N^0)$ .

**Remark 6.1.** *By our choice of the basis (6.8) we have  $\varphi_0^0 \equiv \mathbb{1} \equiv \varphi_{0,\text{r}}^0$ , and the matrix entries  $\varphi_{0,\text{i}}^0(\mathbf{x}_j) = 0$ ,  $j = 1, \dots, P$  could be omitted in  $\mathbf{B}$ . The same holds for the corresponding entry  $S_{0,0}^{\text{i},\text{i}} = 0$  in  $\mathbf{S}$ . Furthermore, we only have to compute the coefficients  $c_0, \dots, c_N$ , and thus the total number of real valued degrees of freedom (DOF) for the first model  $f_{\ell,\text{r}}$  with trigonometric polynomials as basis functions is  $\ell^r = 2N + 1$ . The matrices have the dimensions  $\mathbf{B} \in \mathbb{R}^{(2N+1) \times P}$  and  $\mathbf{S} \in \mathbb{R}^{(2N+1) \times (2N+1)}$ .*

### Simple example

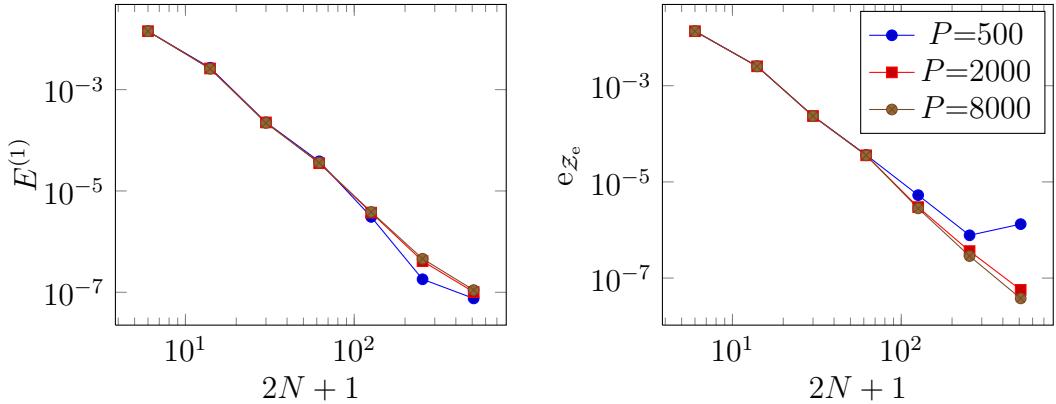
We start our analysis of this model with the simple example (6.4) from Section 6.1.1. Similar to the calculations there, we reconstructed  $f_p$  from  $P = 500, 2000$  and  $8000$  data points with the regularization parameter  $\nu = 10^{-9}$  and chose  $m = 8$ . We considered increasing values  $N = 4, 8, 16, \dots, 256$ . The results are shown in Figure 6.10 and Table 6.4.

For all data sets we observe a rapid decay of  $E^{(1)}$  and  $e_{\mathcal{Z}_e}$ , where the rates slightly improve with increasing number of points  $P$ . However, for the highest resolution  $N = 256$  with  $P = 500$  points, the error increases by a small amount. This is due to an overfitting effect. Altogether, we achieve similar results as in Section 6.1.1 for this artificial example with smooth outer function.

### General example

Next, to analyze the model in a more realistic setting, we choose the tensorproduct hatfunction

$$f_h(\mathbf{x}) := \prod_{d=1}^n \left(1 - 2|x_d - 1/2|\right), \quad \mathbf{x} \in [0, 1]^n \quad (6.11)$$



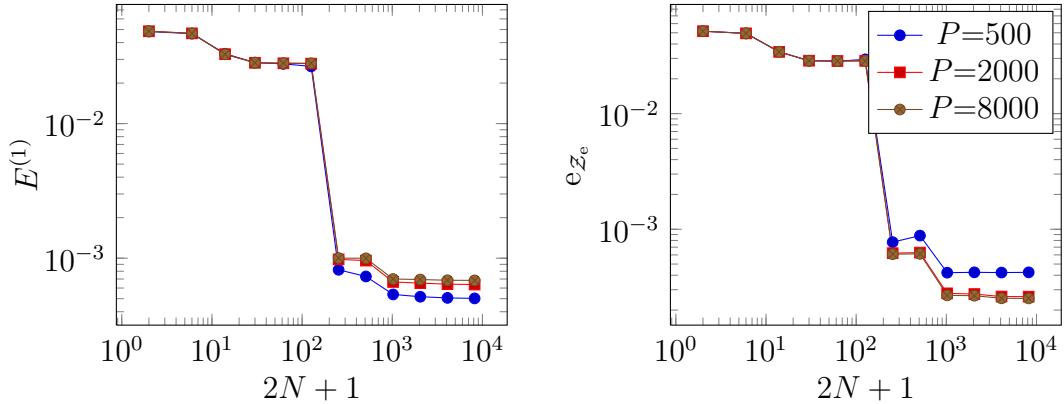
**Figure 6.10.** Reconstruction of  $f_p$  from (6.4) with  $P = 500, 2000$  and  $8000$  data points and the simple model (3.5) using trigonometric polynomials as basis functions. The plots show the value of the energy functional  $E^{(1)}$  with  $\nu = 10^{-9}$  and the MSE depending on the total number of degrees of freedom  $2N + 1$ .

$2N + 1$	$P = 500$		$P = 2000$		$P = 8000$	
	$e_{Z_e}$	slope	$e_{Z_e}$	slope	$e_{Z_e}$	slope
9	$1.36 \cdot 10^{-2}$	—	$1.36 \cdot 10^{-2}$	—	$1.36 \cdot 10^{-2}$	—
17	$2.54 \cdot 10^{-3}$	-1.98	$2.54 \cdot 10^{-3}$	-1.98	$2.53 \cdot 10^{-3}$	-1.98
33	$2.36 \cdot 10^{-4}$	-3.12	$2.35 \cdot 10^{-4}$	-3.12	$2.34 \cdot 10^{-4}$	-3.13
65	$3.64 \cdot 10^{-5}$	-2.58	$3.57 \cdot 10^{-5}$	-2.59	$3.57 \cdot 10^{-5}$	-2.59
129	$5.30 \cdot 10^{-6}$	-2.72	$2.99 \cdot 10^{-6}$	-3.50	$2.81 \cdot 10^{-6}$	-3.58
257	$7.76 \cdot 10^{-7}$	-2.74	$3.64 \cdot 10^{-7}$	-3.00	$2.87 \cdot 10^{-7}$	-3.26
513	$1.32 \cdot 10^{-6}$	+0.77	$5.69 \cdot 10^{-8}$	-2.66	$3.78 \cdot 10^{-8}$	-2.91

**Table 6.4.** Values of the MSE  $e_{Z_e}$  from (6.1) for the reconstruction of  $f_p$  from  $P = 500, 2000, 8000$  data points and regularization parameter  $\nu = 10^{-9}$ . Here, the basis functions from (6.8) have been used. Again, the slope is computed between two successive points in the right graph of Figure 6.10.

as test function and fix  $n = 2$ . Note that we do not consider the sinus example from the previous section due to the fact that our basis functions are now trigonometric polynomials. Therefore, by this choice we prevent any correlations between the basis and the test function. For the same reason we did not consider the piecewise linear function  $f_h$  in Section 6.1.1.

Figure 6.11 and Table 6.5 show the results for a reconstruction of  $f_h$  from  $P = 500, 2000$  and  $P = 8000$  data points. We chose  $m = 8$ ,  $\nu = 10^{-6}$ , and then tested



**Figure 6.11.** Values of  $E^{(1)}$  and  $e_{\mathcal{Z}_e}$  ( $m = 8$ ,  $\nu = 10^{-6}$ ) for the reconstructions of the two-dimensional hatfunction  $f_h$  depending on the value  $2N + 1$ . The graphs show the results for different sizes  $P$  of the test set  $\mathcal{Z}$ .

$2N + 1$	$P = 500$		$P = 2000$		$P = 8000$	
	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope
5	$5.16 \cdot 10^{-2}$	–	$5.16 \cdot 10^{-2}$	–	$5.16 \cdot 10^{-2}$	–
9	$4.94 \cdot 10^{-2}$	-0.04	$4.94 \cdot 10^{-2}$	-0.04	$4.94 \cdot 10^{-2}$	-0.04
17	$3.43 \cdot 10^{-2}$	-0.43	$3.42 \cdot 10^{-2}$	-0.44	$3.42 \cdot 10^{-2}$	-0.44
33	$2.86 \cdot 10^{-2}$	-0.24	$2.87 \cdot 10^{-2}$	-0.23	$2.87 \cdot 10^{-2}$	-0.23
65	$2.84 \cdot 10^{-2}$	-0.01	$2.85 \cdot 10^{-2}$	-0.01	$2.85 \cdot 10^{-2}$	-0.01
129	$2.95 \cdot 10^{-2}$	+0.05	$2.86 \cdot 10^{-2}$	+0.00	$2.85 \cdot 10^{-2}$	+0.00
257	$7.73 \cdot 10^{-4}$	-5.19	$6.21 \cdot 10^{-4}$	-5.46	$6.09 \cdot 10^{-4}$	-5.49
513	$8.81 \cdot 10^{-4}$	+0.19	$6.28 \cdot 10^{-4}$	+0.02	$6.13 \cdot 10^{-4}$	+0.01
1025	$4.21 \cdot 10^{-4}$	-1.06	$2.79 \cdot 10^{-4}$	-1.17	$2.68 \cdot 10^{-4}$	-1.19
2049	$4.24 \cdot 10^{-4}$	+0.01	$2.75 \cdot 10^{-4}$	-0.02	$2.66 \cdot 10^{-4}$	-0.01
4097	$4.22 \cdot 10^{-4}$	-0.01	$2.62 \cdot 10^{-4}$	-0.07	$2.53 \cdot 10^{-4}$	-0.07
8193	$4.24 \cdot 10^{-4}$	+0.01	$2.61 \cdot 10^{-4}$	+0.00	$2.52 \cdot 10^{-4}$	+0.00

**Table 6.5.** Errors between the two-dimensional hatfunction  $f_h$  and its numerical reconstructions with different  $N$  and  $P$ . Further parameter choices were  $m = 8$  and  $\nu = 10^{-6}$ . The slopes are computed as in the tables from Section 6.1.1.

the cases  $N = 2^j$ , for  $j = 1, 2, \dots, 12$ . Here, we turn back to the average error  $e_{\mathcal{Z}_e}$  on the test set to evaluate the results.

Again, the values  $E^{(1)}$  and  $e_{\mathcal{Z}_e}$  decay in two steps: First, for low resolutions they are both diminished only by a small amount. Then, the errors are considerably

reduced for the values  $2N + 1 = 257, 513, 1025$  until they finally reach a saturation, see Figure 6.11. As before, we observe the necessity of a minimal resolution  $N$  which captures the high frequencies of the outer function.

Until now, the results only showed that the choice of the Fourier basis produces in general the same results as the B-spline basis, but we did not benefit from its application in any form. To take advantage from the trigonometric basis, especially the sparseness of the coefficient vector, we have to investigate the frequency structure of  $\Phi$  in more detail. For this reason, we consider its dependence on different model parameters and the function  $f$  in the next section. Note that finally this will lead to the definition of the second model, and the results from the previous sections will serve as a reference for the quality of the second model.

### 6.1.3 Locations of the relevant frequencies

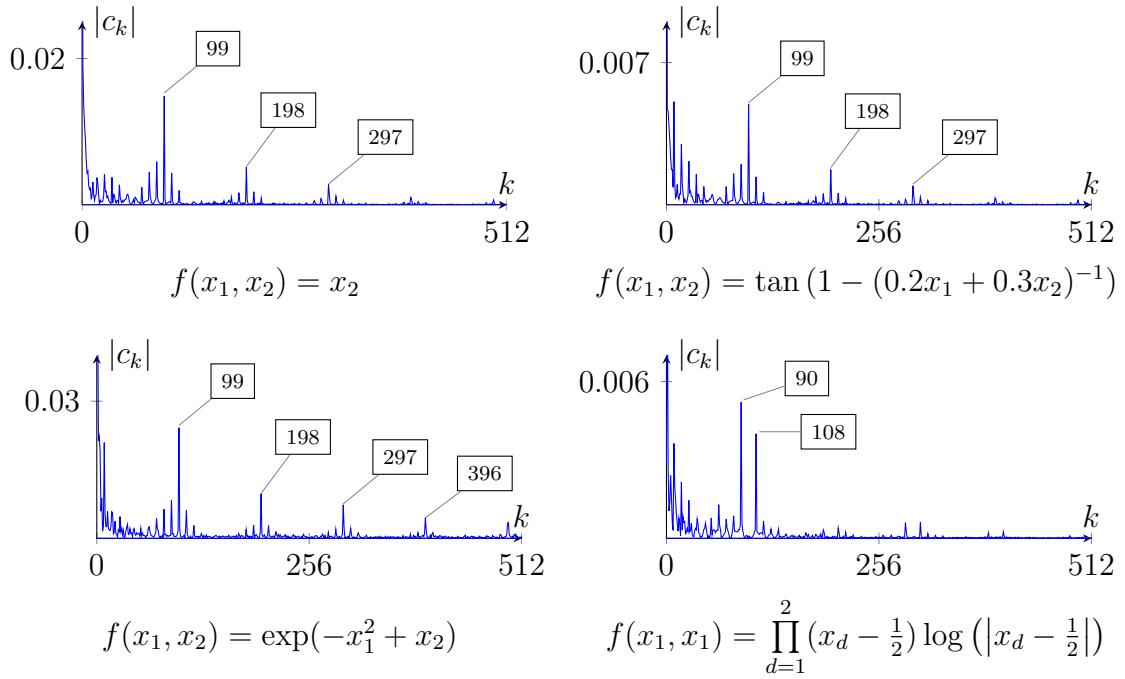
In Section 6.1.1 we concluded that the use of the Fourier basis is advantageous to represent and resolve oscillating functions. This was motivated by an analysis of the discrete Fourier transforms of several outer functions, see Figure 6.9, where we observed that the vector of discrete Fourier coefficients is sparse, i.e. many coefficients are nearly zero. Then, we defined the basis functions (6.8), such that for each  $k = -N, \dots, N$  the coefficient  $c_k$  is simply the Fourier coefficient  $\widehat{\Phi}(k)$  of the outer function  $\Phi$ . However, to benefit from the fact that many coefficients are nearly zero, further knowledge on the frequency structure is useful.

In this section, we will see that it is possible to determine a priori the locations of the frequencies in Fourier space which correspond to large coefficient entries. In particular, they only depend on the dimension  $n$  of the problem and the model parameters  $m, \gamma$ .

To show this, we next present further numerical results for the first model and trigonometric basis functions (6.8). Here, we do not consider the errors  $e_{\mathcal{Z}_e}$  or  $e_{\mathcal{Z}_e}^\infty$ , but rather the magnitudes of the coefficients  $c_k \in \mathbb{C}$  from the extended version of (4.35). This will give more insight since we are interested in the locations  $k$  of the relevant frequencies in Fourier space. Note that in this section we do not assume  $\gamma = m + 2$ .

First, we reconstructed the two-dimensional test functions  $f$  from Table 6.3 for fixed parameters  $m = 8$ ,  $\gamma = 10$ , and  $N = 4096$  from  $P = 8000$  data points. The regularization parameter  $\nu$  was according to Table 6.3. Figure 6.12 shows the magnitudes of the Fourier coefficients  $c_k$ ,  $k = 0, \dots, 512$  for the respective functions. Remember that by (6.9) we only have to consider the positive frequency domain. Additionally, the  $k$ -values for the largest coefficients  $c_k$  are marked by flags in each plot. These values correspond to the dominating frequencies of the oscillating outer function  $\Phi$ .

A comparison of the results in Figure 6.12 with the discrete Fourier transforms

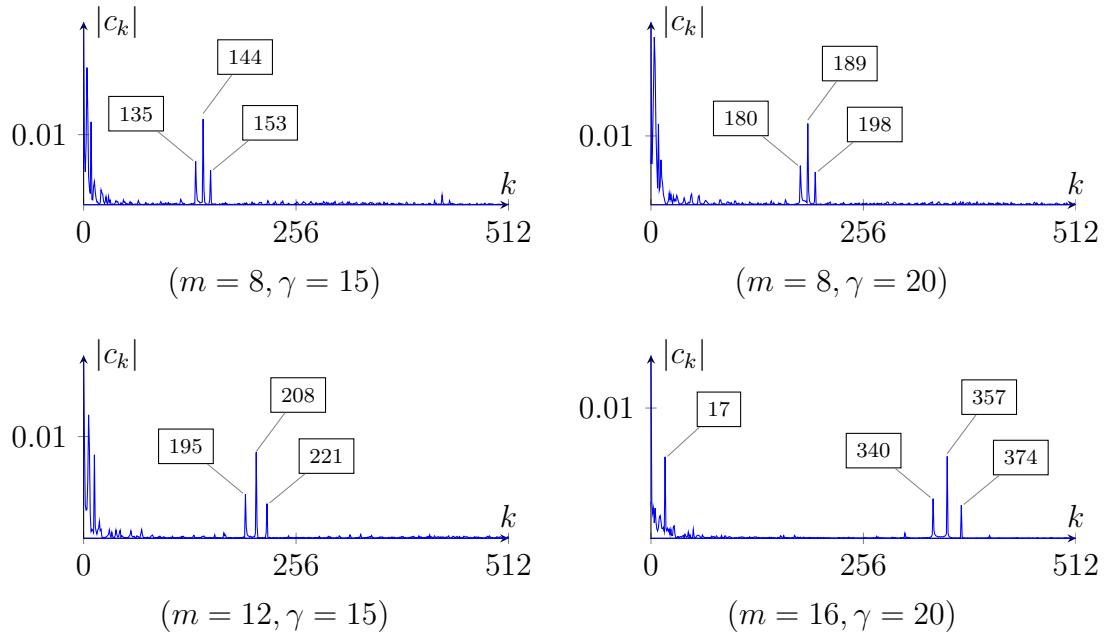


**Figure 6.12.** Relevant (positive) frequencies for the reconstructions ( $m = 8, \gamma = 10$ ) of the functions from Figure 6.8 using trigonometric basis functions, cf. Figure 6.9. The flags indicate the  $k$ -values of the respective frequencies.

from Figure 6.9 shows, that the direct computation of the Fourier coefficients results in the same frequency pattern as before. Note that by definition the magnitudes of the values  $c_k$  differ from the absolute values of the discrete Fourier coefficients  $\widehat{\Phi}_d(k)$ . We observe that in three out of four cases the main frequencies are located at the same positions on the  $k$ -axis. This suggests that they are independent of the function  $f$  which is considered.

In the next tests we investigate varying parameters  $m$  and  $\gamma$ . To this end, we reconstructed the two-dimensional hatfunction (6.11) with  $N = 4096$  from  $P = 8000$  points samples. Figure 6.13 shows the magnitudes of the coefficients  $c_0, \dots, c_{512} \in \mathbb{C}$  for different combinations of  $m$  and  $\gamma$ . As before, the  $k$ -values of the largest coefficients are marked with flags in each figure. We tested the combinations  $(m = 8, \gamma = 15)$ ,  $(m = 8, \gamma = 20)$ ,  $(m = 12, \gamma = 15)$ , and  $(m = 16, \gamma = 20)$ . Remember that in Theorem 2.14 one has to respect the condition  $\gamma \geq m + 2$ , and that all previous examples used  $(m = 8, \gamma = 10)$ . In our computations, the regularization parameter in the first three cases was  $\nu = 10^{-8}$  and  $\nu = 10^{-7}$  in the last case.

From Figure 6.13 we see that the  $k$ -values differ between the different examples. If we additionally take into account the previous tests, then we observe that the



**Figure 6.13.** Magnitudes of the Fourier coefficients  $c_k$ ,  $k = 0, \dots, 512$  in reconstructions of the function  $h$  from (6.11) for different combinations of  $(m, \gamma)$ . The flags mark the  $k$ -values of the largest Fourier coefficients.

main frequencies are always located at

$$k = j \cdot (m+1)(\gamma+1), \quad j \in \mathbb{N},$$

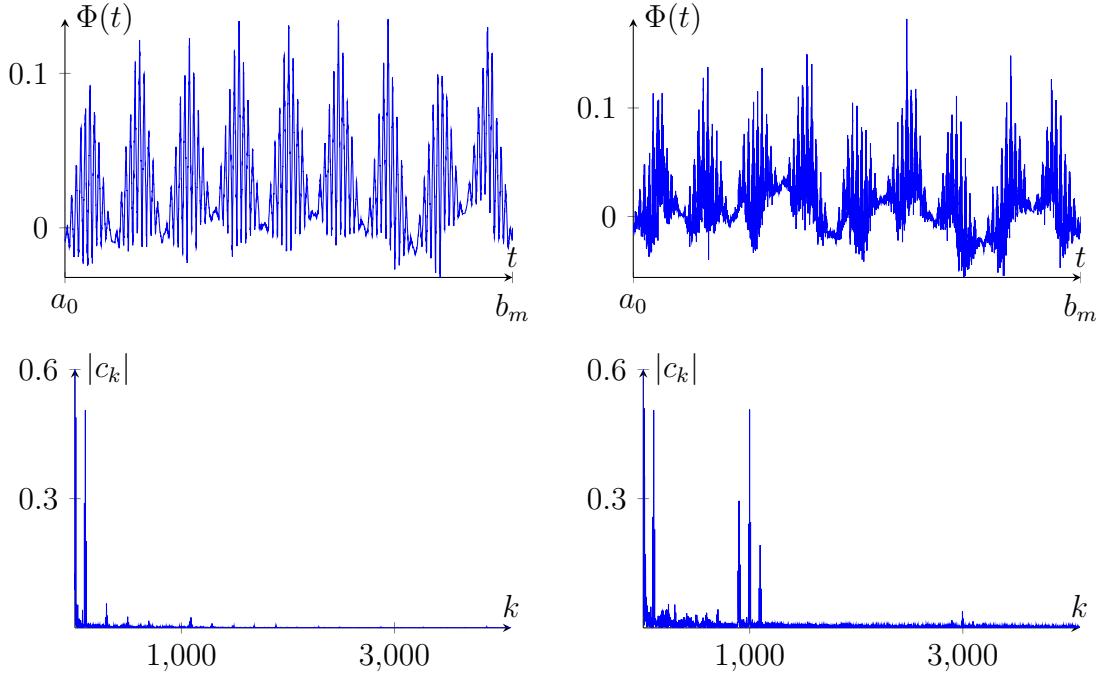
in Fourier space. Additionally, we find that there exist neighbouring peaks for these values, and thus we get the more general locations

$$k = \pm j_1 \cdot (m+1) \pm j_2 \cdot (m+1)(\gamma+1), \quad j_1, j_2 \in \mathbb{N}.$$

Note that this includes the case  $k < 0$  which is reasonable, since the negative  $k$ -axis merely was omitted in our plots due to (6.9).

Until now, all computations dealt with the reconstruction of a two-dimensional continuous function  $f$ . Next, to gain more insight into the dependency of the frequency pattern on the dimension  $n$ , we present the result for a reconstruction of the function  $h$  from (6.11) in dimensions  $n = 2, 3$ . Both functions have been reconstructed from  $P = 8000$  data points with  $N = 8192$ ,  $m = 8$ , and  $\gamma = 10$ . The regularization parameter in the two-dimensional reconstruction was  $\nu = 10^{-7}$  and in the three-dimensional case we used  $\nu = 10^{-8}$ .

Figure 6.14 (top) shows the outer function  $\Phi$  for our sample function for  $n = 2, 3$ . We observe that with rising  $n$  the one-dimensional function is overlayed with a



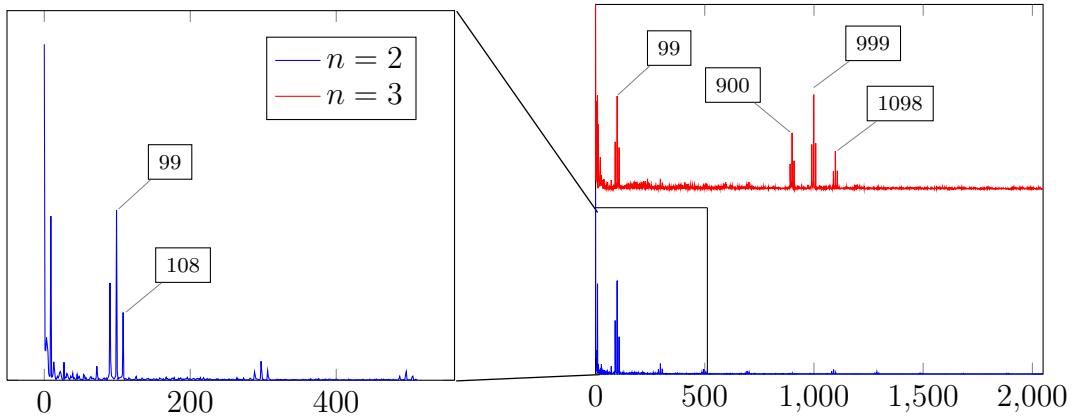
**Figure 6.14.** The outer functions in (3.5) for  $f_h(\mathbf{x})$  from (6.11) (top) and the absolute values of the respective coefficients (bottom). The left column shows the case  $n = 2$  while the right column shows  $n = 3$ .

rescaled function of the same shape. The additional dimension thus causes increasingly oscillations in the outer function. The magnitudes of the corresponding Fourier coefficients  $c_k$ ,  $k = 0, \dots, 4095$  are shown in the bottom row of Figure 6.14. To provide a more detailed insight into the structures of these patterns, Figure 6.15 shows zooms into the pictures. There, flags mark the locations of the main frequencies on the  $k$ -axis.

The left picture in Figure 6.15 shows the location of the main frequencies on the  $k$ -axis for the two-dimensional test case. The absolute values of the complex Fourier coefficients are shown on the ordinate. Anyhow, their values, which depend on  $f$ , have been rescaled for clarity as we are only interested in the location independent of  $f$ . In the right picture we additionally show the three-dimensional case. We observe that the frequencies associated to lower dimensions appear as translated and scaled blocks in the higher dimensions: The main frequencies are now located at the values

$$k = \pm j_1 \cdot (m+1) \pm j_2 \cdot (m+1)(\gamma+1) \pm j_3 \cdot (m+1)(\gamma^2 + \gamma + 1),$$

with  $j_1, j_2, j_3 \in \mathbb{N}$ .



**Figure 6.15.** Locations of the main positive frequencies in Fourier space for the two- and three-dimensional test function  $h$ . The respective values of  $k$  are marked by the flags. The left image is a zoom into the marked region of the right picture.

To generalize this result, we first define for  $d, m, \gamma \in \mathbb{N}$  the values

$$K_d(m, \gamma) := \begin{cases} 1, & d = 0, \\ (m+1) \sum_{k=0}^{d-1} \gamma^k, & d > 0. \end{cases} \quad (6.12)$$

Then, the occurring frequencies in each test case were of the form

$$k = \sum_{d=0}^n \pm j_d K_d(m, \gamma), \quad j_d \in \mathbb{N}.$$

Finally, we remark that the values  $|c_k|$  decayed for increasing  $|k|$ . To account for this fact, we choose for  $d = 0, \dots, n$  the numbers  $N_d \in \mathbb{N}$  and assume that

$$j_d \in \{-N_d, \dots, -1, 0, 1, \dots, N_d\}.$$

In conclusion we can make the following conjecture which is based on the numerical observations in this section: For the reconstruction of an  $n$ -dimensional function  $f$  with model parameters  $m, \gamma \in \mathbb{N}$ , the relevant frequency numbers are given by the set

$$\mathcal{K}_n(m, \gamma) := \left\{ k = \sum_{d=0}^n k_d : k_d = j_d K_d(m, \gamma), j_d \in \{-N_d, \dots, N_d\} \right\}. \quad (6.13)$$

Truncating the Fourier basis to those frequencies in  $\mathcal{K}_n$  substantially reduces the complexity of the model (3.5). A careful analysis however shows that there is

still an exponential dependence of the number of active Fourier coefficients on the dimension  $n$ .

The tests in this section showed that the frequencies depend on the number of terms  $m$ , the basis  $\gamma$  in (2.10), and the dimension  $n$ . These parameters are model parameters and merely cause the general structure of the appearing frequencies, while the specific function  $f$  affects the amplitudes of the frequencies. This can be explained by the inner function which results to an unfolding of the dimensions in the mentioned Z-curve manner as it was explained in Section 2.4. From these observations we conclude that the dimensionality of the function  $f$  translates into structured oscillation of the outer function  $\Phi$ .

## 6.2 Second model

The definition of the second model is based on the observations from Section 6.1.3, where we could identify the relevant frequencies in Fourier space in form of the set (6.13). We start this section with a derivation of the product model (4.36) for the special choice of the Fourier basis. Note that we could also fix the basis functions in (4.38) instantaneously, but the following analysis will provide a motivation for the definitions and constructions from Section 3.2.2 and Section 4.5.2. It will point out the choice of a product ansatz to approximate the outer function.

### 6.2.1 Trigonometric polynomials

The unfolding of dimensions which is caused by the inner functions  $\Psi_q$ ,  $q = 0, \dots, m$  was explained in Section 2.4. Here, we will see that this approximately results in a separation of the dimensions into a product of functions with different frequencies.

We start with the definition of the basis functions in (4.38) for  $d = 0, \dots, n$ . To this end, let the numbers  $N_0, \dots, N_n \in \mathbb{N}$  be given. For  $d = 0, \dots, n$  we set  $\ell_d = 2N_d$  and shift the index sets  $\{0, \dots, 2N_d\} \mapsto \{-N_d, \dots, N_d\}$  in (4.38). Then, for each set

$$\mathcal{F}_{\ell_d}^d := \mathcal{F}_{N_d}^d := \{\varphi_{-N_d}^d, \dots, \varphi_{N_d}^d\}, \quad d = 0, \dots, n,$$

the basis functions are defined for  $j = -N_d, \dots, N_d$  by

$$\varphi_j^d(t) := \exp\left(\frac{2\pi}{\omega_m} i k_j^d t\right), \quad \text{with } k_j^d := j \cdot K_d(m, \gamma). \quad (6.14)$$

Note that for  $d = 0, \dots, n$  the functions in  $\mathcal{H}_{\ell_d}^d = \text{span } \mathcal{F}_{\ell_d}^d$  are periodic with period  $\omega_m/K_d(m, \gamma)$ . They have the form

$$\phi^d(t) := \sum_{j=-N_d}^{N_d} c_j^d \exp\left(\frac{2\pi}{\omega_m} i k_j^d t\right), \quad c_j^d \in \mathbb{C}, \quad (6.15)$$

and their Fourier coefficients (6.6) are given by

$$\widehat{\phi}^d(k_j^d) = \begin{cases} c_j^d, & j = -N_d, \dots, N_d, \\ 0, & \text{else.} \end{cases}$$

In other words, for  $d = 0, \dots, n$ ,  $j_d = -N_d, \dots, N_d$  the coefficient  $c_{j_d}^d \in \mathbb{C}$  is simply the Fourier coefficient of  $\phi^d$  which is located at  $k_{j_d}^d \in \mathbb{Z}$  in Fourier space.

Next, we consider a function  $\Phi \in \mathcal{H}_L^{(n)}$ , see (4.41), which has the special form

$$\Phi(t) := \prod_{d=0}^n \left( \sum_{j=-N_d}^{N_d} c_j^d \exp\left(\frac{2\pi}{\omega_m} i k_j^d t\right) \right), \quad c_j^d \in \mathbb{C},$$

and remark that the outer functions in the set  $\mathcal{K}_L$  from (4.48) are exactly of this type, see Section 4.5.2.

Now, by the fact that multiplication of exponentials is actually an addition of frequency-numbers

$$\exp\left(\frac{2\pi}{\omega_m} i k_1 t\right) \exp\left(\frac{2\pi}{\omega_m} i k_2 t\right) = \exp\left(\frac{2\pi}{\omega_m} i (k_1 + k_2) t\right),$$

one can see that the frequency-pattern of  $\Phi$  is precisely given by the set  $\mathcal{K}_n(m, \gamma)$  in (6.13). This property motivated the definition of the product model in Section 3.2.2 and Section 4.5.2. However, note that conversely not any function which is defined on this frequency-pattern in Fourier space is necessarily a product of functions. Therefore, our ansatz is merely an approximation of the outer function which is made first by the choice of the frequency set  $\mathcal{K}_n(m, \gamma)$ , and second by the use of a product of functions.

Finally, to compute the Fourier coefficients of the function  $\Phi$ , we first define for each  $k \in \mathcal{K}_n(m, \gamma)$  the set of multiindices

$$\mathcal{J}(k) := \left\{ \mathbf{j} := (j_0, \dots, j_n) : k = \sum_{d=0}^n k_{j_d}^d, \quad |j_d| \leq N_d, \quad d = 0, \dots, n \right\},$$

and see that the  $k$ -th Fourier coefficient of  $\Phi = \prod_{d=0}^n \phi^d$  is given by

$$\widehat{\Phi}(k) = \begin{cases} \sum_{\mathbf{j} \in \mathcal{J}(k)} (c_{j_0}^0 \cdots c_{j_n}^n), & k \in \mathcal{K}_n(m, \gamma), \\ 0, & \text{else.} \end{cases} \quad (6.16)$$

Thus, each product  $(c_{j_0}^0 \cdots c_{j_n}^n)$ ,  $\mathbf{j} \in \mathcal{J}(k)$ ,  $k \in \mathcal{K}_n(m, \gamma)$  is a contribution to the Fourier coefficient  $\widehat{\Phi}(k_{j_0}^0 + \cdots + k_{j_n}^n)$ . Clearly, for  $\#\mathcal{J}(k) = 1$  it holds  $(c_{j_0}^0 \cdots c_{j_n}^n) = \widehat{\Phi}(k_{j_0}^0 + \cdots + k_{j_n}^n)$ .

Altogether we end up with a complex valued extended model  $f_L \in \mathcal{K}_L$  which has the form

$$f_L(\mathbf{x}) := \sum_{q=0}^m \prod_{d=0}^n \left( \sum_{j=-N_d}^{N_d} c_j^d \exp \left[ 2\pi i k_j^d \left( \sum_{p=1}^n \alpha_p \psi(x_p + qa) + \Delta_q \right) \right] \right). \quad (6.17)$$

Now, to get a real valued function from  $f_L$ , in Section 4.5.2 we argued that the definition of  $f_{L,\mathbf{r}}$  is less restrictive than the choice  $\tilde{f}_L^{(n)}$ , see Remark 4.20 for a definition. To explain this, we exemplarily consider  $\Phi$  as before, and therewith the Fourier coefficients of the real valued functions

$$\tilde{\Phi}(t) := \prod_{d=0}^n \operatorname{Re}(\phi^d(t)), \quad \text{and} \quad \Phi_{\mathbf{r}}(t) := \operatorname{Re} \left( \prod_{d=0}^n \phi^d(t) \right),$$

where for each  $d = 0, \dots, n$ ,  $\phi^d$  is defined by (6.15).

To point out the difference between  $\tilde{\Phi}$  and  $\Phi_{\mathbf{r}}$ , we first split up each index set  $\{-N_d, \dots, N_d\}$ ,  $d = 0, \dots, n$  into disjoint sets

$$\mathcal{I}_0^d := \{-N_d, \dots, -1\}, \quad \text{and} \quad \mathcal{I}_1^d := \{0, \dots, N_d\},$$

which only contain positive or negative numbers<sup>4</sup>, respectively. Furthermore, let  $\mathbf{j} := (j_0, \dots, j_n) \in \mathbb{Z}^{(n+1)}$  and  $\mathbf{s} := (s_0, \dots, s_n) \in \{0, 1\}^{(n+1)}$  be multiindices, and define the index set  $\mathcal{T} := \{0, 1\}^{(n+1)} \setminus \{(0, \dots, 0), (1, \dots, 1)\}$ . Then, one can split up  $\Phi$  into three parts

$$\Phi(t) = \prod_{d=0}^n \phi^d(t) = \left( \prod_{d=0}^n \Sigma_0^d(t) \right) + \left( \prod_{d=0}^n \Sigma_1^d(t) \right) + \sum_{\mathbf{s} \in \mathcal{T}} \left( \prod_{d=0}^n \Sigma_{s_d}^d(t) \right), \quad (6.18)$$

where we abbreviated for  $d = 0, \dots, n$  and  $s \in \{0, 1\}$

$$\Sigma_s^d(t) := \sum_{j \in \mathcal{I}_s^d} c_j^d \exp \left( \frac{2\pi}{\omega_m} i k_j^d t \right).$$

Next, we define for  $\mathbf{s} \in \{0, 1\}^{(n+1)}$  the set  $\mathcal{I}_{\mathbf{s}} = \mathcal{I}_{s_0}^0 \times \dots \times \mathcal{I}_{s_n}^0$ . Then, each term in (6.18) can written as

$$\prod_{d=0}^n \Sigma_{s_d}^d(t) = \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{s}}} (c_{j_0}^0 \cdots c_{j_n}^n) \exp \left( \frac{2\pi}{\omega_m} i (k_{j_0}^0 + \dots + k_{j_n}^n) t \right). \quad (6.19)$$

---

<sup>4</sup>Here, for simplicity and in abuse of the usual terminology, we make the convention that 0 is a positive number.

Note that the first term in (6.18) corresponds to  $\mathbf{s} = (0, \dots, 0)$  and solely contains negative indices  $j_d < 0$ . The second term results from  $\mathbf{s} = (1, \dots, 1)$  and only contains positive  $j_d \geq 0$ , while for the third term the signs of the indices  $j_d$  vary depending on the multiindex  $\mathbf{s} \in \mathcal{T}$ .

Now, to compute  $\tilde{\Phi} = \prod_{d=0}^n \operatorname{Re}(\phi^d)$ , remember that each factor  $\operatorname{Re}(\phi^d)$ ,  $d = 0, \dots, n$  is a real valued function and by (6.7) we can assume for  $d = 0, \dots, n$  that

$$c_{-j}^d = \bar{c}_j^d, \quad \text{for } j = 0, \dots, N_d, \quad (6.20)$$

see also (6.9). This dependency is sketched in Figure 6.16 (a) and (b). There, for  $d \in \{0, 1\}$  and  $j = -2, \dots, 2$  the values  $|c_j^d|$  are marked blue for  $j \geq 0$  and red for  $j < 0$ . From (6.20) we can see that the red coefficients are defined from the blue ones. Next, Figure 6.16 (c) sketches the frequency pattern of the product  $\prod_{d=0}^n \phi^d$  for this case. Note that in our example  $\#\mathcal{J}(k) = 1$ . Here, the values  $|c_{j_0}^0 \cdots c_{j_n}^n|$  for coefficients in the first term of (6.18), i.e.  $\mathbf{j} \in \mathcal{I}_{\mathbf{s}}$ ,  $\mathbf{s} = (0, \dots, 0)$  are marked red, while the respective values for  $\mathbf{j} \in \mathcal{I}_{\mathbf{s}}$ ,  $\mathbf{s} = (1, \dots, 1)$  in the second term are colored blue, see also (6.19). Again, the red values are fixed through their blue counterparts by (6.20). Finally, the magnitudes of  $(c_{j_0}^0 \cdots c_{j_n}^n)$  from the third term in (6.18) are marked brown, i.e.  $\mathbf{j} \in \mathcal{I}_{\mathbf{s}}$ ,  $\mathbf{s} \in \mathcal{T}$ . Here, for at least one factor  $c_{j_d}^d$ ,  $d \in \{0, \dots, n\}$  it holds  $j_d < 0$ , and thus it is given by (6.20).

Altogether we see that only in the second term of (6.18) there is no restriction on the corresponding coefficients  $c_j^d$  through (6.20). These values are marked blue in Figure 6.16 (c). For all remaining coefficients (brown, red) there is at least some restriction from the assumption (6.20) which comes along with the computation of  $\tilde{\Phi} = \prod_{d=0}^n \operatorname{Re}(\phi^d)$ .

In contrast to this, the dependencies of coefficients in  $\Phi_r = \operatorname{Re}(\prod_{d=0}^n \phi^d)$  are less restrictive. To see this, the splitting of the index sets  $\{-N_d, \dots, N_d\}$  is not necessary: Since the factors  $\phi^d$  are complex valued functions, there is no a priori restriction on their coefficients  $c_j^d$ , cf. (6.20). This is depicted in Figure 6.17 (A) and (B), where all coefficients are marked blue.

Next, to compute  $\Phi_r$  we rather define the sets of multiindices

$$\mathcal{J}_- := \{\mathbf{j} \in \mathcal{J}(k) : k \in \mathcal{K}_n(m, \gamma), \quad k < 0\}$$

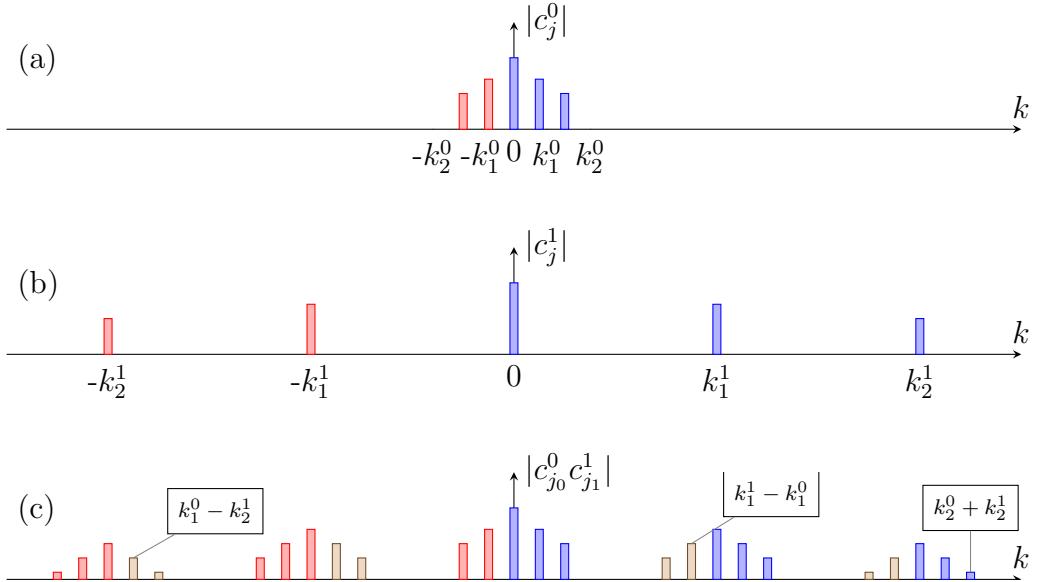
and

$$\mathcal{J}_+ := \{\mathbf{j} \in \mathcal{J}(k) : k \in \mathcal{K}_n(m, \gamma), \quad k \geq 0\},$$

which allow for the splitting of

$$\Phi_r(t) = \operatorname{Re} \left( \prod_{d=0}^n \phi^d(t) \right) = \sum_{s \in \{-, +\}} \sum_{\mathbf{j} \in \mathcal{J}_s} (c_{j_0}^0 \cdots c_{j_n}^n) \exp \left( \frac{2\pi}{\omega_m} i (k_{j_0}^0 + \dots + k_{j_n}^n) t \right)$$

into two sums which run over the index sets  $\mathcal{J}_-$ ,  $\mathcal{J}_+$ . Now, to account for the fact that  $\Phi_r = \operatorname{Re}(\prod_{d=0}^n \phi^d)$  is a real valued function and therefore (6.7) holds, we can



**Figure 6.16.** In picture (a) and (b), the relation between the Fourier coefficients  $c_j^d$  of  $\phi^d$  is shown for  $d = 0, 1$ ,  $j = -2, \dots, 2$ : The red values are given by the blue ones through (6.9). Part (c) shows the dependencies for  $\tilde{\Phi} = \prod_{d=0}^n \text{Re}(\phi^d)$ . The brown values are partly determined by (6.9), while the red coefficients again depend on their blue counterparts.

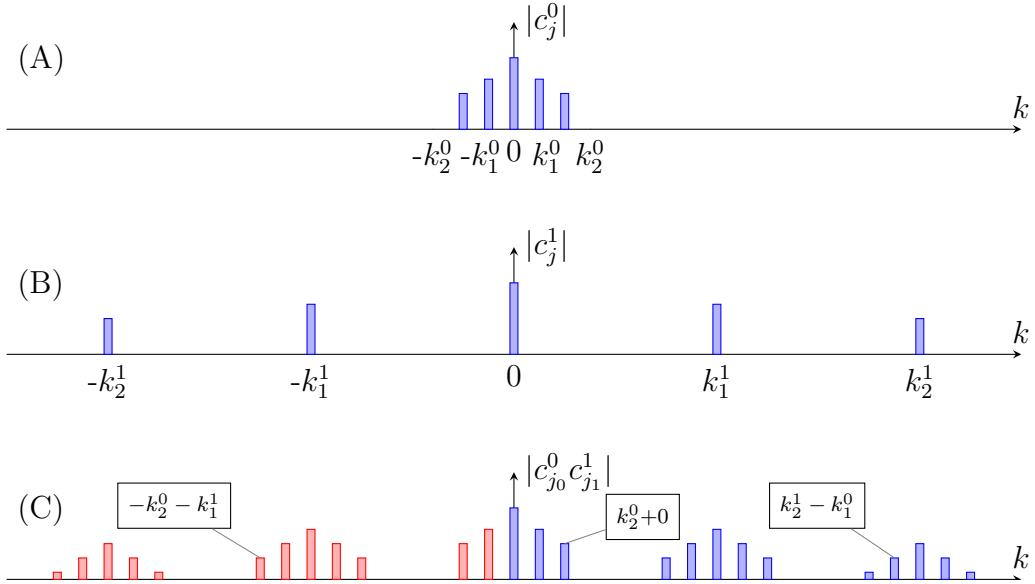
assume that

$$c_{-j_0}^0 \cdots c_{-j_n}^n = \overline{c_{j_0}^0 \cdots c_{j_n}^n}, \quad \mathbf{j} = (j_0, \dots, j_n) \in \mathcal{I}_+, \quad (6.21)$$

which now replaces assumption (6.20). Finally, the fact that the second sum over all indices  $\mathbf{j} \in \mathcal{J}_+$  contains the second term  $\prod_{d=0}^n \Sigma_1^d$  from (6.18) plus additional terms shows that  $\Phi_r$  is less restrictive than  $\tilde{\Phi}$ . This is also shown in Figure 6.17 (C), where the magnitudes  $|c_{j_0}^0 \cdots c_{j_n}^n|$  are marked blue for  $\mathbf{j} \in \mathcal{J}_+$  and red for  $\mathbf{j} \in \mathcal{J}_-$ . Again, the red values depend on the blue ones. We conclude with the remark that the same arguments apply to the functions  $\tilde{f}_L^{(n)}$  and  $f_{L,r}$ . For this reason we chose the second option in Section 4.5.2, cf. Remark 4.20.

After the foregoing remarks on the definition of the function  $f_{L,r}$ , we next define the norm for the regularization term in (4.50). To this end, remember that the norm  $\|\cdot\|_{\mathcal{H}(K)}$  is defined by the values  $\gamma_j^d$  in (4.19), see (4.34). For the first model, i.e.  $d = 0$ , these values have been defined in (6.10) as the  $H^s$ -Norm of the basis functions  $\varphi_j^0$ . Next, we generalize this definition and set for  $s > 1/2$ ,  $d = 0, \dots, n$

$$\gamma_j^d := (b_m - a_0) \left( 1 + \left( \frac{2\pi}{\omega_m} \right)^2 K_d(m, \gamma)^{-2} |k_j^d|^2 \right)^s, \quad j = -N_d, \dots, N_d. \quad (6.22)$$

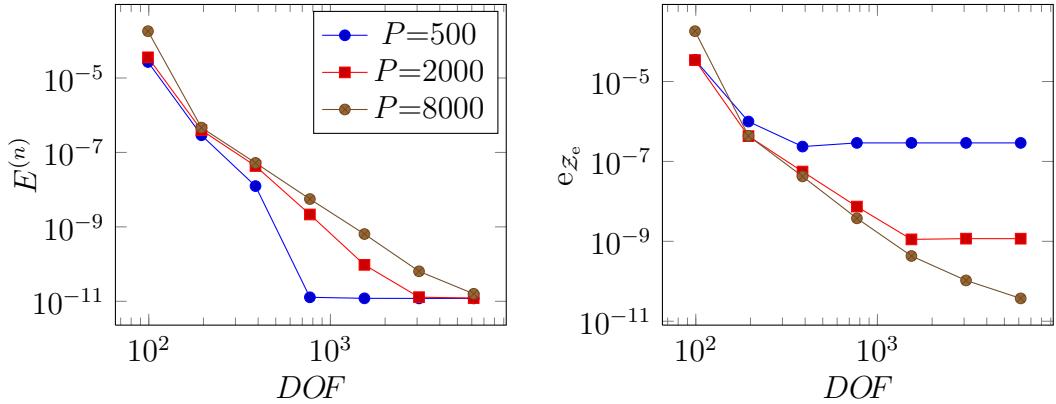


**Figure 6.17.** Dependencies of the Fourier coefficients of  $\Phi_r = \text{Re}(\prod_{d=0}^d \phi^d)$ . (A), (B) sketch the coefficients of  $\phi^0, \phi^1$  without restriction. In (C) the red values can be computed from the blue coefficients through (6.21) to account for the fact that  $\Phi_r$  is real valued.

First, remember that  $K_0(m, \gamma) = 1$  which implies that (6.10) and (6.22) coincide if  $d = 0$ . However, for  $d > 0$ , (6.22) defines norms  $\|\cdot\|_{\mathcal{H}^d}$  which are merely equivalent to the  $H^s$ -norm, but account for the intrinsic oscillations of the basis functions. To explain this, consider definition (6.14). There, the period of functions in  $\mathcal{H}_{\ell_d}^d$  is scaled from  $\omega_m$  to the smaller value  $\omega_m/K_d(m, \gamma)$ . Clearly, the norm  $\|\cdot\|_{\mathcal{H}^d}$  compensates for the multiplication of the frequency numbers by the factor  $K_d(m, \gamma)$  and the resulting additional oscillations. Next, remember from Section 2.4 that the oscillations are due to the unfolding of dimensions which is caused by the inner functions  $\Psi_q$ ,  $q = 0, \dots, m$ . However, the inner functions are independent of the function  $f$ , and thus a definition of the norm through (6.22) does not penalize the additional oscillations which have been introduced into the model to account for the unfolding of dimension.

### Simple example

For the so defined model  $f_{L,r}$  we start the presentation of numerical results for the solution of Problem 4.19 with the simple example from Section 6.1.1. Remember that there, the outer function is a fixed polynomial  $p$  which defines  $f_p$  through (6.4). Again, we reconstructed the two-dimensional function  $f_p$  from  $P = 500, 2000, 8000$



**Figure 6.18.** Value of the error functional (4.50) (left) and the MSE (6.1) (right) in the reconstruction of  $f_p$  from (6.4) for different values of  $DOF$ . The reconstructions have been computed from  $P = 500, 2000, 8000$  data points with  $m = 8$  and  $\nu = 10^{-12}$ .

points with  $m = 8$  terms in (6.17). Note that in the following we will again use  $\gamma = m + 2$  in all calculations. For each factor (6.15) we chose the values  $N_0 = \dots = N_n = N$  to be equal. Consequently, the total number of complex degrees of freedom is

$$L = \sum_{d=0}^n (\ell_d + 1) = \sum_{d=0}^n (2N_d + 1) = (n+1)(2N+1), \quad (6.23)$$

which implies for this example ( $n = 2$ ) that the total number of real degrees of freedom is

$$DOF := 2L = 2(n+1)(2N+1) = 6(2N+1).$$

Figure 6.18 shows the results for increasing values  $N = 8, \dots, 512$  and the regularization parameter  $\nu = 10^{-12}$ . In the left picture the value of the error functional  $E^{(n)}(f_{L,r})$  from (4.50) is plotted against  $DOF$  while the right picture shows the mean-square-error (MSE)  $e_{Z_e}(f_{L,r})$  depending on  $DOF$ . Table 6.6 gives the corresponding numerical values of the MSE.

Similar to the reconstructions of  $f_p$  with the first model, we observe a rapid decay of the MSE. However, depending on the value of  $P$ , it saturates at some level, and the error is not reduced any more, or even slightly increases. From Table 6.6 it becomes clear that the turning point is when  $DOF$  exceeds the number of learning points  $P$ . This shows that the amount of information which is contained in the learning data set  $Z$  determines the quality of the reconstruction that can be achieved at best. In conclusion, this simple example shows that the second model leads to similar results as the first model although the outer function has no product structure. This is due to the fact that  $f_{\ell,r} \in \mathcal{H}_\ell^{(1)} \subset \mathcal{K}_L \ni f_{L,r}$ .

DOF	$P = 500$		$P = 2000$		$P = 8000$	
	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope	$e_{\mathcal{Z}_e}$	slope
102	$3.49 \cdot 10^{-5}$	—	$3.44 \cdot 10^{-5}$	—	$1.82 \cdot 10^{-4}$	—
198	$9.96 \cdot 10^{-7}$	-5.25	$4.29 \cdot 10^{-7}$	-6.47	$4.37 \cdot 10^{-7}$	-8.90
390	$2.35 \cdot 10^{-7}$	-2.10	$5.58 \cdot 10^{-8}$	-2.98	$4.22 \cdot 10^{-8}$	-3.41
774	$2.91 \cdot 10^{-7}$	+0.31	$7.40 \cdot 10^{-9}$	-2.93	$3.77 \cdot 10^{-9}$	-3.50
1542	$2.91 \cdot 10^{-7}$	+0.00	$1.12 \cdot 10^{-9}$	-2.73	$4.29 \cdot 10^{-10}$	-3.14
3078	$2.91 \cdot 10^{-7}$	+0.00	$1.17 \cdot 10^{-9}$	+0.06	$1.05 \cdot 10^{-10}$	-2.03
6150	$2.91 \cdot 10^{-7}$	+0.00	$1.16 \cdot 10^{-9}$	+0.00	$3.73 \cdot 10^{-11}$	-1.50

**Table 6.6.** For each  $P = 500, 200, 8000$  the value of the MSE for the reconstruction of  $f_p$  is given. The functions  $f_{L,r}$  have been computed with  $m = 8$  terms in (6.17) and regularization parameter  $\nu = 10^{-12}$ . The slopes have been computed for successive points of the respective graph in Figure 6.18.

### Numerical analysis of the nonlinear minimizers

For the previous simple example we used the BFGS–method to compute the minima of the non–convex error functional  $E^{(n)}$ . To explain this choice, we next analyze the iterative nonlinear minimizers from Section 5.1 numerically. For this purpose, we define the test function

$$f_d(\mathbf{x}) := \sqrt{\sum_{d=1}^n (x_d - d(\mathbf{x}))^2}, \quad \text{with} \quad d(\mathbf{x}) := \frac{1}{n} \sum_{d=1}^n x_d, \quad (6.24)$$

and consider in the following the case  $n = 2$ . The function value  $f_d(\mathbf{x})$  is simply the distance of the point  $\mathbf{x} \in [0, 1]^n$  to the diagonal of  $[0, 1]^n$ . We remark that this function has no product structure, and by that choice we prevent any correlations between the test function, the product structure of the outer function, and the basis for each factor. Furthermore,  $f_d$  is continuous but not differentiable on the diagonal of the  $n$ –dimensional unit cube.

To set up a test case which provides enough information to reconstruct (6.24), the training set  $\mathcal{Z}$  always contains  $P = 8000$  samples of  $f_d$ . Furthermore, we fix the parameters  $m = 8$  and  $\nu = 10^{-6}$ .

For this test problem, we next investigate the BFGS–method, the PRcg–method, and the FRcg–method separately with respect to their dependency on the choice of the start value for the iteration.<sup>5</sup> To this end, we randomly sample an initial

<sup>5</sup>We do not analyze the steepest descent (SD) method here but remark that it reduces  $E^{(n)}$  independently from the start value. However, the reduction is very slow, see below.

coefficient  $\mathbf{c}_0 \in [0, 0.5]^{2L}$  which will then be used as start value for each method  $\text{ITER}(\cdot) = \text{BFGS}, \text{PRcg}, \text{FPcg}$ . Remember from Section 5.1 that  $\text{ITER}(\cdot)$  performs one iteration step of the respective nonlinear minimizer, i.e. we compute for  $k = 0, 1, 2 \dots$

$$\mathbf{c}_{k+1} = \text{ITER}(\mathbf{c}_k).$$

The method stops, if either a previously defined maximum number of iterations  $k_{\max} \in \mathbb{N}$  is reached, or

$$\frac{|E^{(n)}(\mathbf{c}_{k+1}) - E^{(n)}(\mathbf{c}_k)|}{|E^{(n)}(\mathbf{c}_{k+1})|} < \varepsilon$$

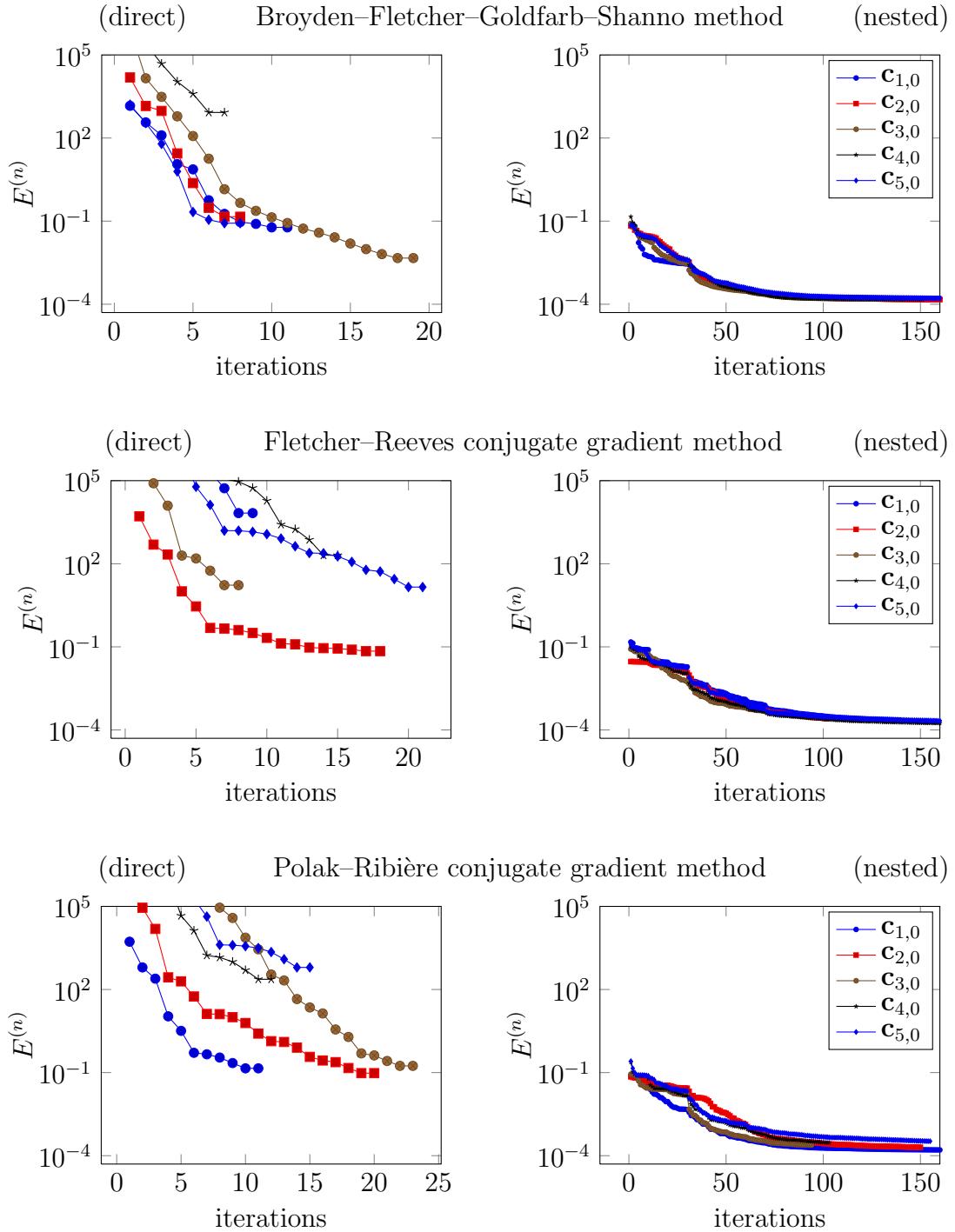
holds for some given accuracy  $\varepsilon > 0$ . Numerical tests have shown that reasonable choices for the stopping parameters are  $\varepsilon := 10^{-8}$ ,  $k_{\max} := 1000$ , and  $\delta := 10^{-2}$  for the line search, see (5.4). In the following we will refer to this method as the *direct* application of the respective nonlinear minimizer.

Our numerical tests showed that there exists a strong dependency of all methods on the choice of the start value. In the left column of Figure 6.19 we give five examples for which none of the minimizers led to a satisfactory result.

Figure 6.19 (left) shows the value of the error functional  $E^{(n)}(\mathbf{c}_k)$  for each iteration step of the BFGS–method, the FRcg–method, and the PRcg–method which result from five different random start values  $\mathbf{c}_{j,0} \in [0, 0.5]^{2L}$ ,  $j = 1, \dots, 5$ . Clearly, the iteration histories reveal that a direct application of the minimizers is not possible. In fact, they either rapidly ran into a poor local minimum or the iteration even stopped without success due to a failing line search (e.g.  $\mathbf{c}_{4,0}$ ).

To remedy this, we introduced the nested iteration Algorithm 5.1 which recursively precomputes the start values for the respective iterative nonlinear minimizer as the solution of the same problem at a lower resolution, see Section 5.1. Now, in order to generate a sequence of nested approximation spaces  $\mathcal{K}_{L_1} \subset \mathcal{K}_{L_2} \subset \dots \subset \mathcal{K}_{L_J}$  for our test problem, let  $J = 7$ , and  $N^{(j)} := 2^{j+1}$  which implies  $L_j = (n+1)(2N^{(j)} + 1)$ ,  $j = 1, 2, \dots, J$ , cf. (6.23). Furthermore, we choose for each lower resolution  $k_{\max}^1 := \dots := k_{\max}^{J-1} := 10$ , and  $k_{\max}^J := k_{\max} := 1000$ . All further stopping parameters  $\varepsilon := 10^{-8}$ , and  $\delta := 10^{-2}$  are as before. The start values  $\mathbf{c}_{j,0}^1 \in \mathbb{R}^{2L_1}$  for the nested iteration scheme at the lowest resolution are derived for  $j = 1, \dots, 5$  from the corresponding  $\mathbf{c}_{j,0} \in \mathbb{R}^{2L_J}$  in the following way: Let  $f_{L_J}^j \in \mathcal{K}_{L_J}$  be given by (6.17) with the coefficient vector  $\mathbf{c}_{j,0}$ . Then,  $\mathbf{c}_{j,0}^1$  is defined to be the coefficient vector of the orthogonal projection of  $f_{L_J}^j$  onto  $\mathcal{K}_{L_1}$  with respect to the scalar product (4.34) in  $\mathcal{H}(K)$ .

The right column of Figure 6.19 shows the results for the previous test cases and the nonlinear minimizers  $\text{ITER}(\cdot) = \text{BFGS}, \text{PRcg}, \text{FPcg}$  but with an application of the nested iteration Algorithm 5.1. Clearly, this stabilized the iteration processes. For all methods, and independent of the start values  $\mathbf{c}_{j,0}^1$ ,  $j = 1, \dots, 5$ , the error



**Figure 6.19.** The left pictures show the iteration histories of five test cases for which a direct application of the respective nonlinear minimizer to reconstruct (6.24) fails. The graphs in the right column result from the same test cases but with an application of the nested iteration Algorithm 5.1.

functional  $E^{(n)}$  could be significantly reduced to an approximately similar value. Thus, an application of the nested iteration scheme reduces the dependency of all methods from the choice of the start value.

Next, we compare the performance of the SD–method, the FRcg–method, the PRcg–method, and the BFGS–method. To this end, we first directly apply each method to the previous test problem, i.e. we reconstruct  $f_d$  with the parameters  $P = 8000$ ,  $m = 8$ ,  $\nu = 10^{-6}$ , and  $N = 256$ . However, in the following we always use the start vector

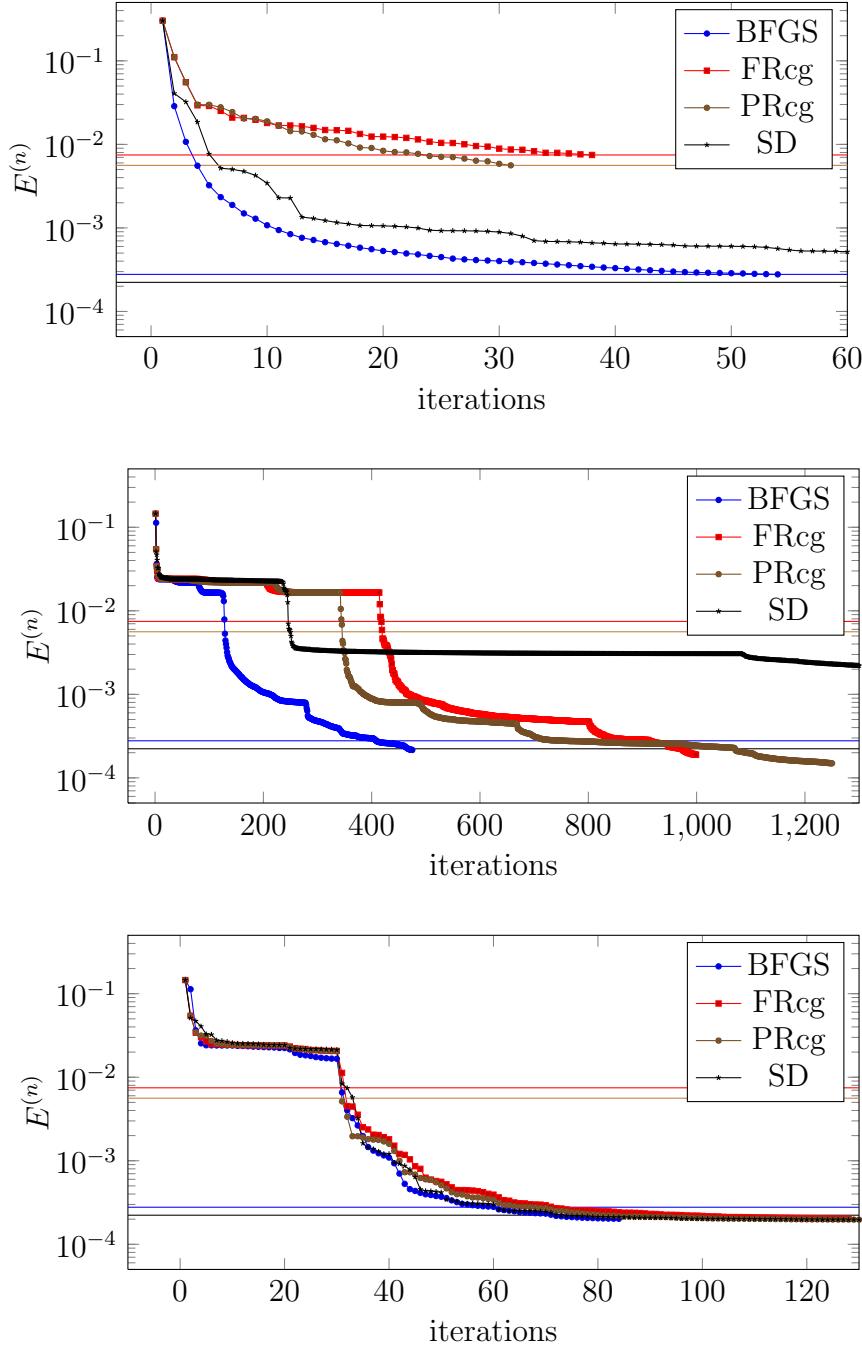
$$\mathbf{c}_0 := (c_{-N,\mathfrak{r}}^0, c_{-N,\mathfrak{t}}^0, \dots, c_{N,\mathfrak{r}}^0, c_{N,\mathfrak{t}}^0, \dots, c_{-N,\mathfrak{r}}^n, c_{-N,\mathfrak{t}}^n, \dots, c_{N,\mathfrak{r}}^n, c_{N,\mathfrak{t}}^n)^T \in \mathbb{R}^{2L},$$

where for  $d = 0, \dots, n$  the coordinates are given by

$$c_{j,\mathfrak{t}}^d := 0, \quad j = -N, \dots, N, \quad \text{and} \quad c_{j,\mathfrak{r}}^d := \begin{cases} 1, & j = 0, \\ 0, & \text{else.} \end{cases}$$

Then, we additionally apply each method together with the nested iteration algorithm using, as before, the orthogonal projection of this start value onto  $\mathcal{K}_{L_1}$ . Here, we employ the same maximal iteration number  $k_{\max}^1 := \dots := k_{\max}^J := 1000$  for each level, while all other parameters are as described previously. In a final test, we reduce the values to  $k_{\max}^j := 10$  for the lower levels  $j = 1, \dots, J-1$ . In Figure 6.20 the iteration history of each solver in the respective test is given.

Figure 6.20 shows that the quality of the results for a direct application differs for each solver with respect to the minimal value of  $E^{(n)}$ , which is marked by a horizontal line, and the number of iterations that are required to achieve this minimum. Here, the difference between PRcg–method and FRcg–method is rather small. Both minimizers stop after a small number of iterations, however at a large value of  $E^{(n)}$ . In contrast, the SD–method computes the lowest value  $E^{(n)}$  but converges very slowly. In fact, the value which is marked by the black line is attained after  $k_{\max} = 1000$  iterations. The direct application of the BFGS–method led to the best results. After a small number of iterations it attained a minimum which is close to the SD minimum. With an additional application of the nested iteration scheme and  $k_{\max}^j := 1000$ ,  $j = 1, \dots, J$ , see the second picture of Figure 6.20, all minimizers achieved this minimum. We remark that this also holds for the SD–method although this is not visible from the plot due to its slow convergence. Finally, we observe that the number of total iterations can be reduced if  $k_{\max}^j$  is small for the lower resolutions  $j < J$  in the nested iteration algorithm. Figure 6.20 (bottom) shows that for this test problem and any nonlinear minimizer,  $k_{\max}^j = 10$ ,  $j = 1, \dots, J-1$  suffices to achieve good results with few iterations. We conclude our numerical investigations of the nonlinear minimizer with the remark that an application of the nested iteration Algorithm 5.1 is essential to solve Problem 4.19. For all numerical results in this thesis it was used in combination with the BFGS–method.



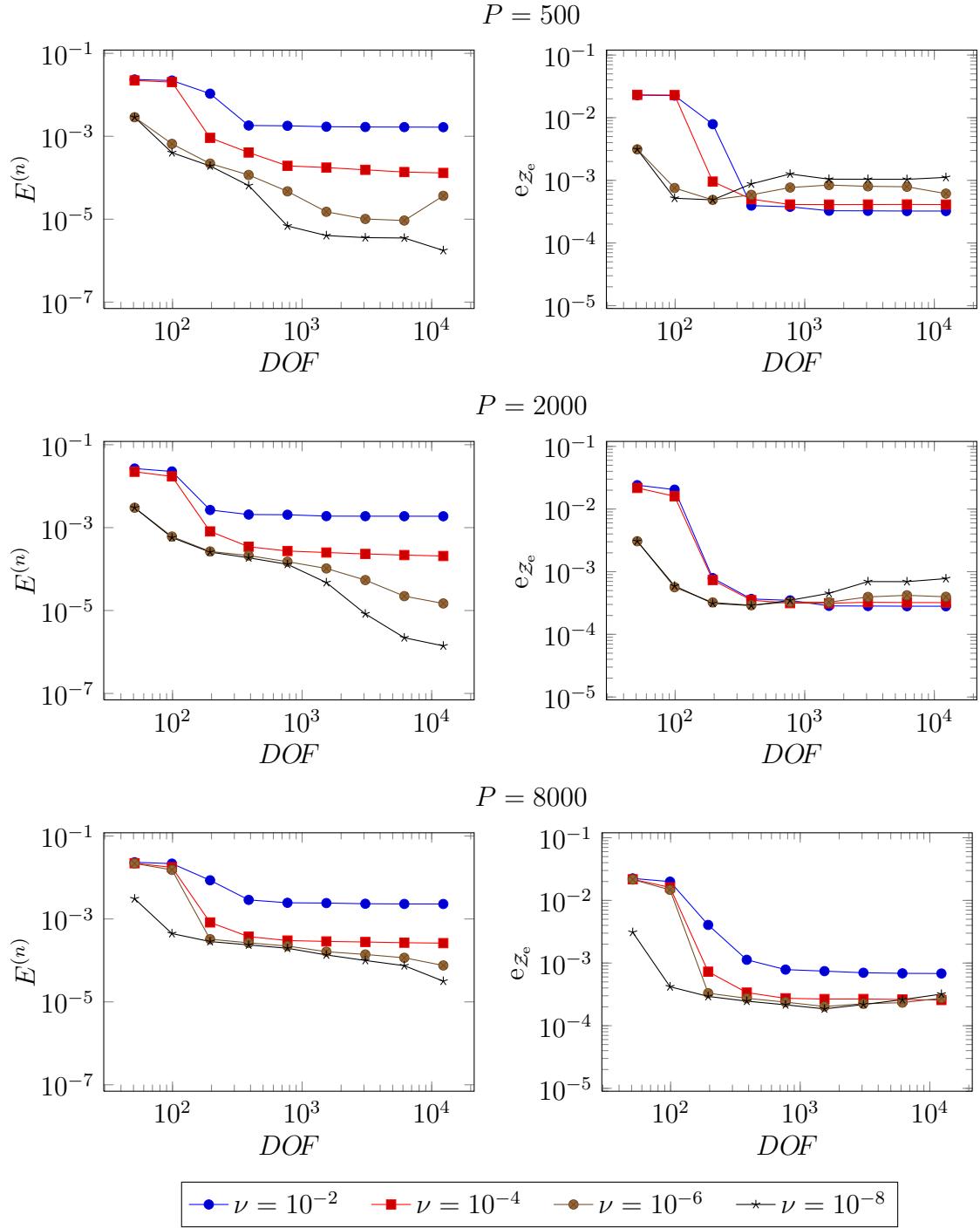
**Figure 6.20.** Iteration histories of SD-, FRcg-, PRcg-, and BFGS-method for a direct application (top) to the test problem, and for the nested iteration algorithm with  $k_{\max}^j := 1000$ ,  $j = 1, \dots, J$ , (middle) and  $k_{\max}^j := 10$ ,  $j = 1, \dots, J - 1$ ,  $k_{\max}^J := 1000$  (bottom). The horizontal lines mark the minimal value of  $E^{(n)}$  which is achieved by the direct application, respectively.

### Numerical analysis of the model parameters

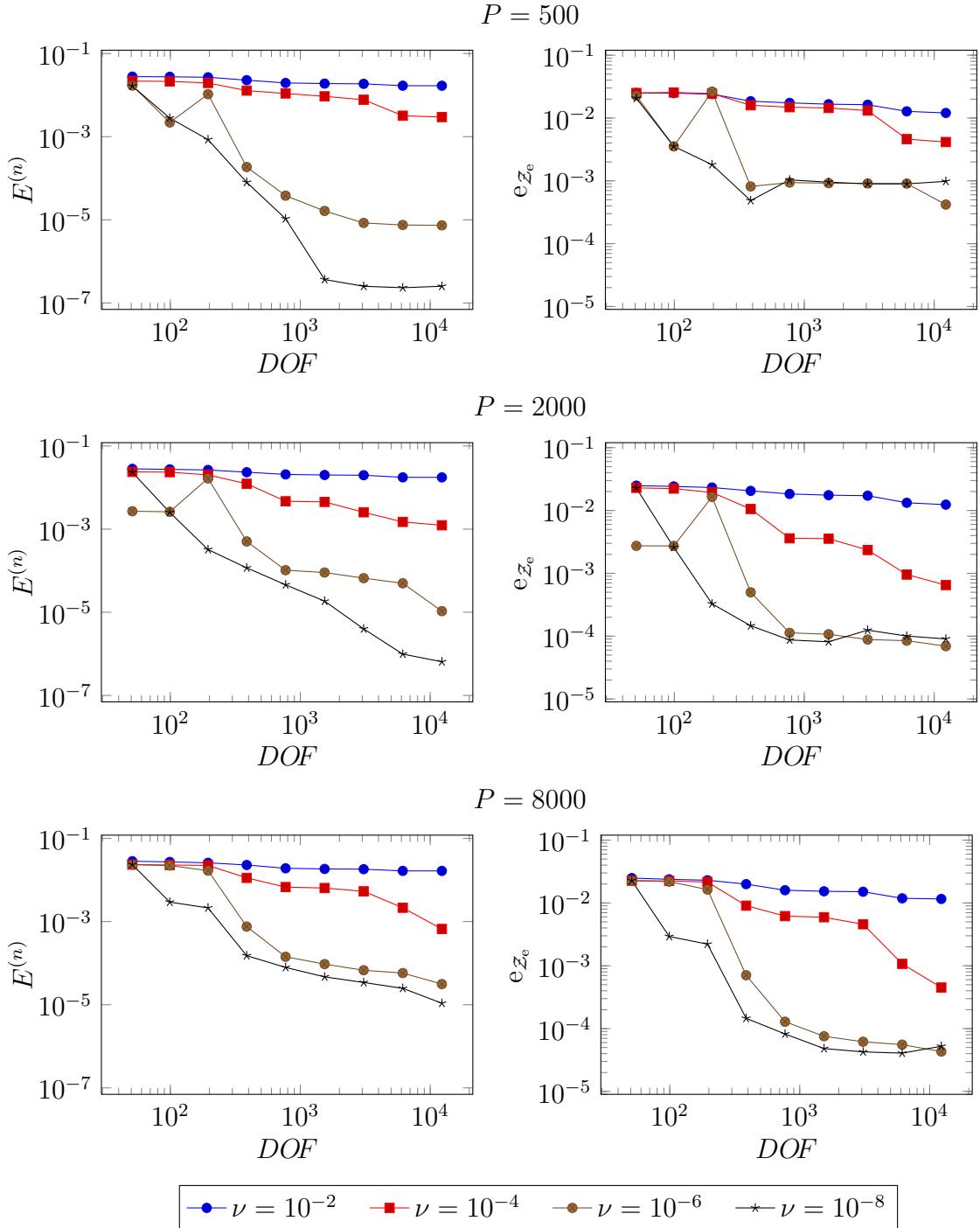
After the previous considerations on the nonlinear minimizers, next our emphasis is on a more detailed numerical analysis of the second model (6.17). To this end, we reconstructed  $f_d$  for  $n = 2$ , see (6.24), and different parameter setups of  $f_L$  from  $P = 500, 2000, 8000$  data points. To be more precise, the reconstructions were computed for an increasing number of terms  $m = 8, 16, 32$  and different regularization parameters  $\nu = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ . With  $N^{(j)} := 2^{j+1}$ , we calculated the solution of Problem 4.19 in each setup for increasing resolutions  $N = N^{(J)}$ ,  $J = 1, \dots, 9$  for which (6.23) implies that the total number of real degrees of freedom is  $DOF = 6(2^{J+1} + 1)$ . Here, to stabilize the nonlinear minimizer, we used  $k_{\max}^j = 100$ ,  $j = 1, \dots, J - 1$  iterations at the lower resolutions in the nested iteration scheme, and  $\delta = 0.1$  in (5.4).

Figures 6.21–6.23 show the results of these tests for  $m = 8, 16, 32$ , respectively. There, the value of the error functional  $E^{(n)}$  (4.50) is shown against  $DOF$  in the left columns. The graphs in the right columns show the MSE (6.1) depending on  $DOF$ . In each figure, the number of learning points is increased from  $P = 500$  (top row) to  $P = 2000$  (middle row), and  $P = 8000$  (bottom row). Finally in all graphs, for both  $E^{(n)}$ , and MSE the results are marked with separate colors for each regularization parameter  $\nu$  which was used.

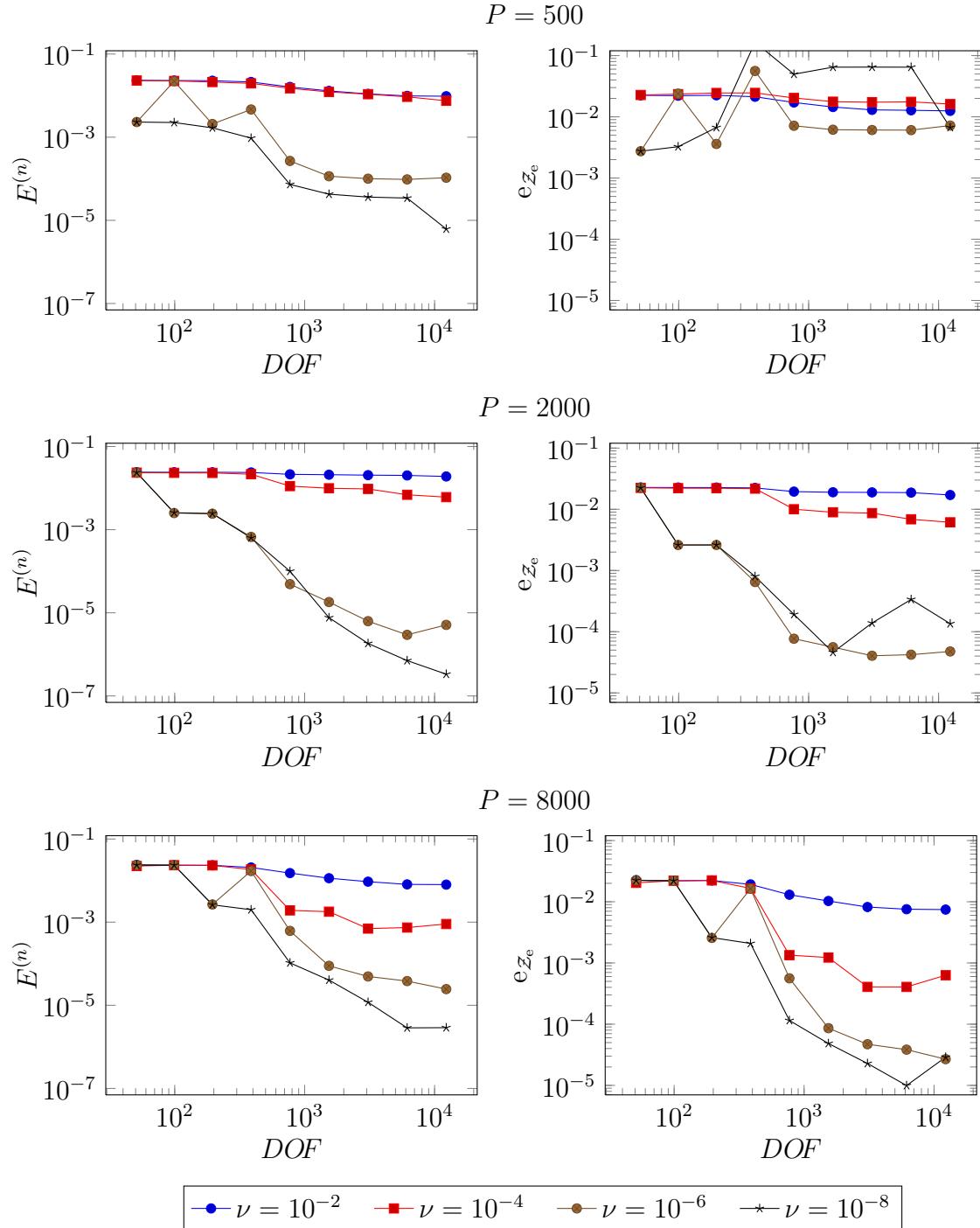
We clearly see that in general the value of the error functional decreases for increasing resolution  $N$  and then saturates for fine resolutions. However, the decay of  $E^{(n)}$  is not monotone in all cases which shows that the minimizer still can run into a poor local minimum although we increased the iteration numbers  $k_{\max}^j = 100$ ,  $j = 1, \dots, J - 1$  for the lower resolutions in the nested iteration scheme. Thus it might be preferable to also use e.g.  $k_{\max}^j = k_{\max}^J$  if  $j < J$ . Anyhow, since the value of the error functional is an upper bound for the MSE on the training data we can conclude that the model approaches the training set. In particular for small training data sets this leads to an overfitting of the data. This can be seen in detail in the cases  $m = 8$ ,  $P = 500, 2000$  where the MSE slightly increases for larger values of  $DOF$  while the value of  $E^{(n)}$  becomes small. This effect is postponed if  $P$  is larger. Furthermore, we see that increasing the number of terms  $m$  leads to better approximations of  $f_d$  if the training set is large ( $P = 2000, 8000$ ). This is consistent with the theoretical argument (2.62) that a larger value  $m$  leads to a faster convergence of Algorithm 2.1, see the remarks at the end of Section 2.3.7. But for larger  $m$  the outer function, i.e. its oscillations, becomes more difficult to approximate and more degrees of freedom and more learning points are needed. This is indicated by the later onset of the decay, by the earlier onset of the saturation of the error ( $m = 16, 32$ ,  $P = 2000, 8000$ ,  $\nu = 10^{-6}$ ), and by the first row in Figure 6.23 ( $P = 500$ ,  $m = 32$ ) where, based on the small learning set, the error  $e_{Z_e}$  could not be reduced. This is also the case for large regularization parameters  $\nu$ . While in



**Figure 6.21.** Values of the error functional  $E^{(n)}$  (left) and the MSE (6.1) (right) for increasing resolutions in the reconstruction of  $f_d$  from  $P = 500, 2000, 8000$  data points with parameters  $m = 8, \gamma = 10$  and different regularization parameters.



**Figure 6.22.** Results for the reconstructions of  $f_d$  with parameters  $m = 16$ ,  $\gamma = 18$ , and varying regularization parameter  $\nu$ . Again,  $E^{(n)}$  and  $e_{\mathcal{Z}_e}$  are plotted against  $DOF$  for increasing resolutions. The rows show different sizes  $P$  of the training set  $\mathcal{Z}$ .



**Figure 6.23.** The graphs show the results for the same test cases as in Figure 6.21 and Figure 6.22 but for the model parameters  $m = 32$ ,  $\gamma = 34$ .

general we observe that the overfitting effect can be reduced by a larger  $\nu$ , the MSE remains large for  $m = 16, 32$  and  $\nu = 10^{-2}, 10^{-4}$ . This is due to the fact that the regularization term in (4.50) has more impact on the value of the error functional which enforces higher regularity of the approximand  $f_L$ . Here, it even dominates  $E^{(n)}$ . In contrast, the use of a larger value of  $\nu$  is preferable for smaller training data sets and small values of  $m$  to avoid overfitting of the data ( $m = 8, P = 500, 2000$ ). Finally we remark that the saturation of the error for large resolutions is due to the limited information that is contained in the finite number of learning points.

Altogether we conclude that the product model (6.17) produces good results and behaves as expected from comparable regression methods, cf. [38]. Furthermore, the product model (6.17) allows to also approximate functions that are not smooth and have no tensor–product structure. This indicates numerically that the product structure of the outer function  $\Phi$  is indeed induced by the outer mappings  $\Psi_q$  and not by the function  $f$  itself. However, to achieve good results, special care has to be taken of the minimization of the non–convex error functional  $E^{(n)}$ .

### Higher dimensional examples

Until now, our numerical analysis was based two–dimensional examples. In the following we aim at the reconstruction of higher dimensional functions with the product model (6.17). In particular, we consider the test function

$$f_s(\mathbf{x}) := \prod_{d=1}^n \sin(\pi x_d), \quad \mathbf{x} \in [0, 1]^n$$

in dimensions  $n = 4, 6, 8, 10$ . To reconstruct these functions, the training data set was enlarged for increasing dimension, while the resolution was constant  $N = 32$ . Following the results from the previous section we chose a large number of terms in (6.17) to get better approximations, i.e. we take  $m = 32$ . Remember also that  $m \geq 2n$  is necessary. Furthermore, we used the parameters from Table 6.7. There, the size  $P$  of the training set, the number of real degrees of freedom  $DOF$ , and the regularization parameter  $\nu$  which led to the best results are listed. Furthermore, the values of the error functional and the MSE are given for each test. However, when comparing the results, note that both the number of training points  $P$ , and test points  $P_e$  which is used to calculate  $e_{\mathcal{Z}_e}$  depend on the dimension. See Table A.4 in the appendix for a list of the values  $P_e$  depending on the dimension  $n$ .

Figures 6.24–6.27 show the graphs of the reconstructions in the respective dimensions on different two–dimensional hyperplanes. Here, for a  $(x_i, x_j)$ –plane, the remaining coordinates of  $\mathbf{x} \in [0, 1]^n$  are chosen to be  $x_k = 0.5$ . This corresponds to the maximum of  $f_s$  with respect to this coordinate. For reasons of simplicity we only present four different hyperplanes. Note however that the results do not differ

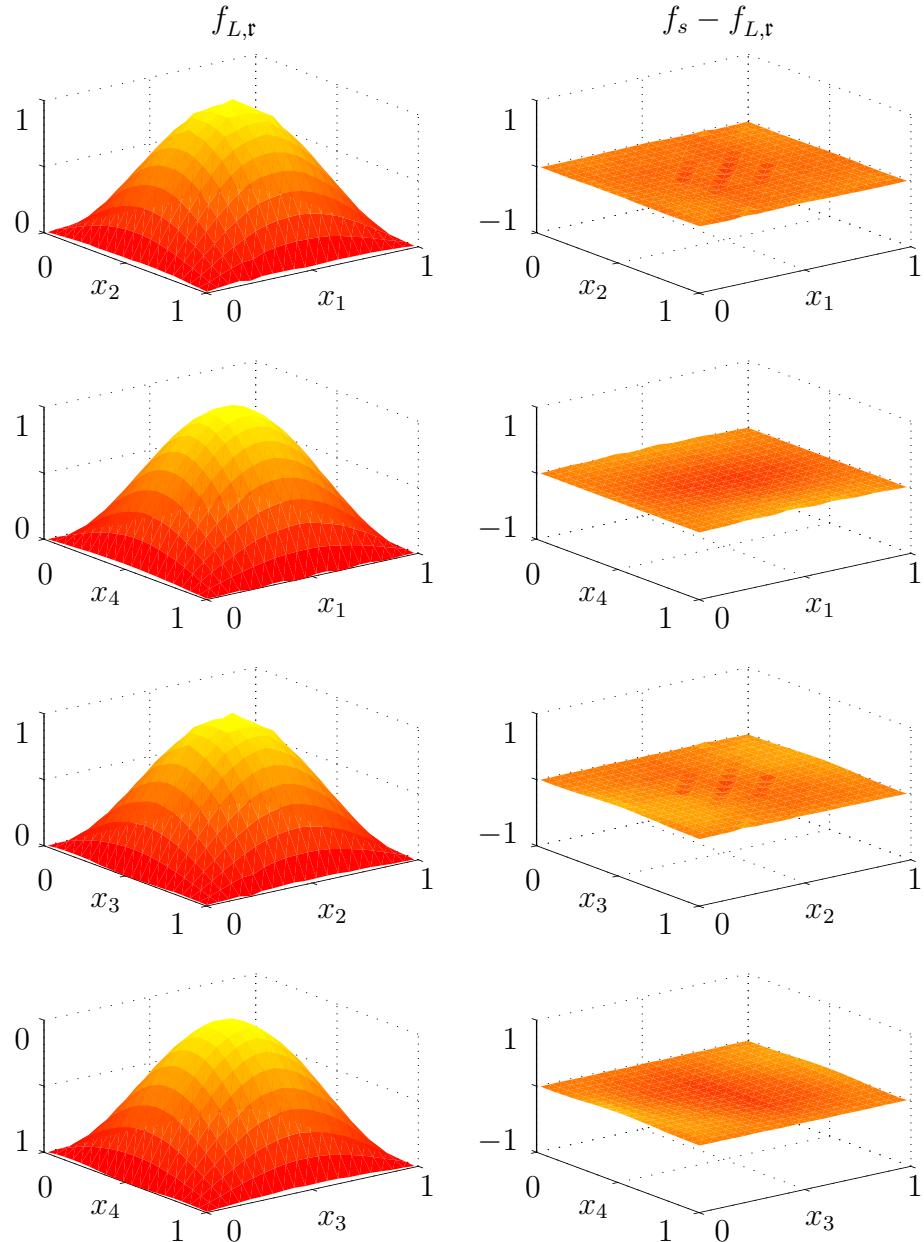
$n$	$P$	$DOF$	$\nu$	$E^{(n)}$	$e_{\mathcal{Z}_e}$
4	20 000	650	$10^{-6}$	$6.38 \cdot 10^{-5}$	$6.32 \cdot 10^{-5}$
6	50 000	910	$10^{-10}$	$3.15 \cdot 10^{-5}$	$3.65 \cdot 10^{-5}$
8	80 000	1170	$10^{-10}$	$1.28 \cdot 10^{-5}$	$1.55 \cdot 10^{-5}$
10	100 000	1430	$10^{-9}$	$7.41 \cdot 10^{-6}$	$1.21 \cdot 10^{-5}$

**Table 6.7.** Parameters for the reconstructions of  $f_s$  in the respective dimensions. Additionally, the values of the error functional and the MSE are given for each test case.

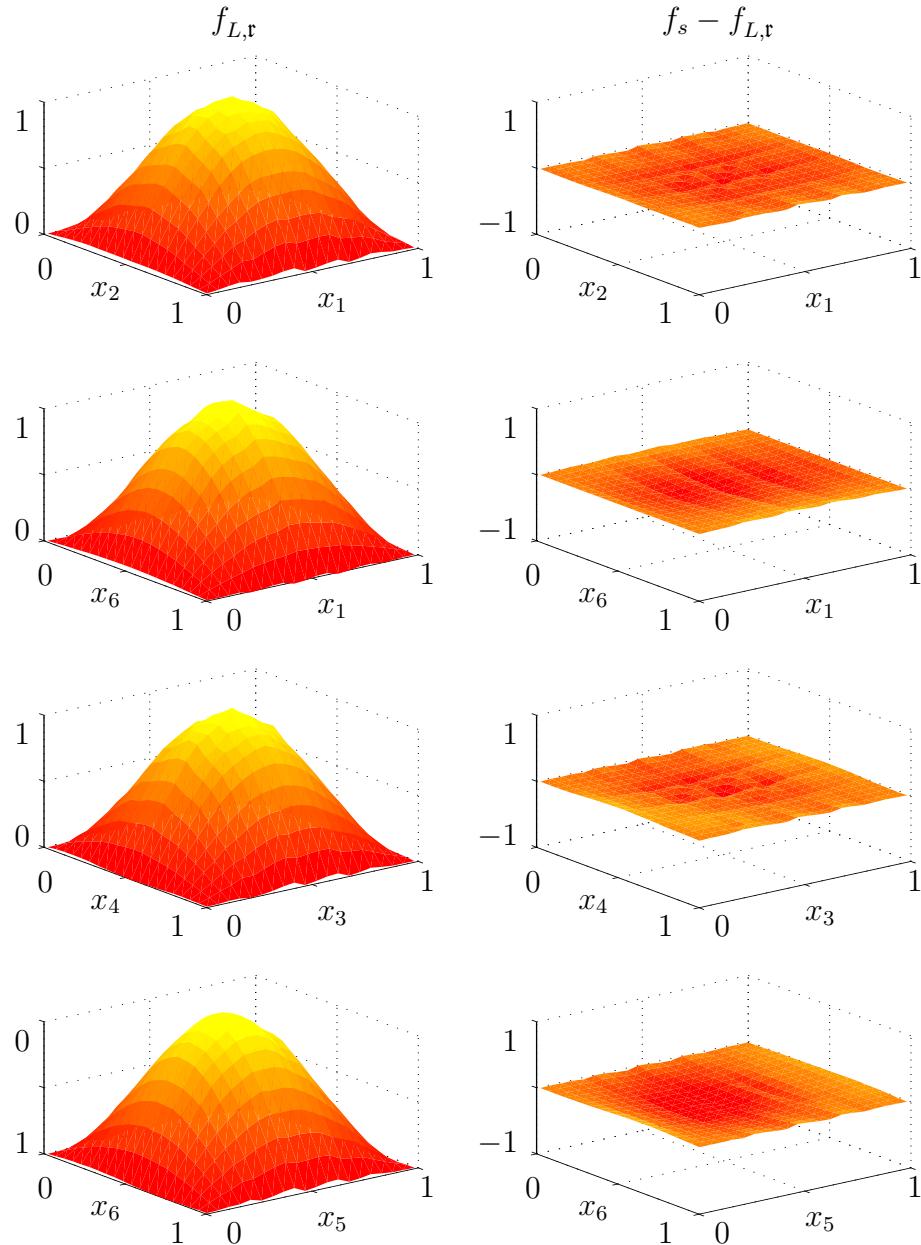
considerably for the other cases. Additionally, the residual  $f_s - f_{L,r}$  is plotted for the respective hyperplane.

The results demonstrate that we obtain a good approximation of  $f_s$  with very few degrees of freedom. The residuals are close to 0 in all hyperplanes. Note that this includes the  $(x_{n-1}, x_n)$ -planes which correspond to the highest dimensions of  $f_s$ , i.e., where the highest frequencies of the outer function have to be resolved. We roughly achieve the same order of magnitude for the value of the error functional and the MSE in all dimensions while the number of degrees of freedom increases linearly with  $n$  the Fourier expansion remains the same. However, we remark again that a direct comparison of the error  $e_{\mathcal{Z}_e}$  for different dimensions has to be considered with care. Furthermore, the examples show that our ansatz is suited for large data sets  $\mathcal{Z}$ .

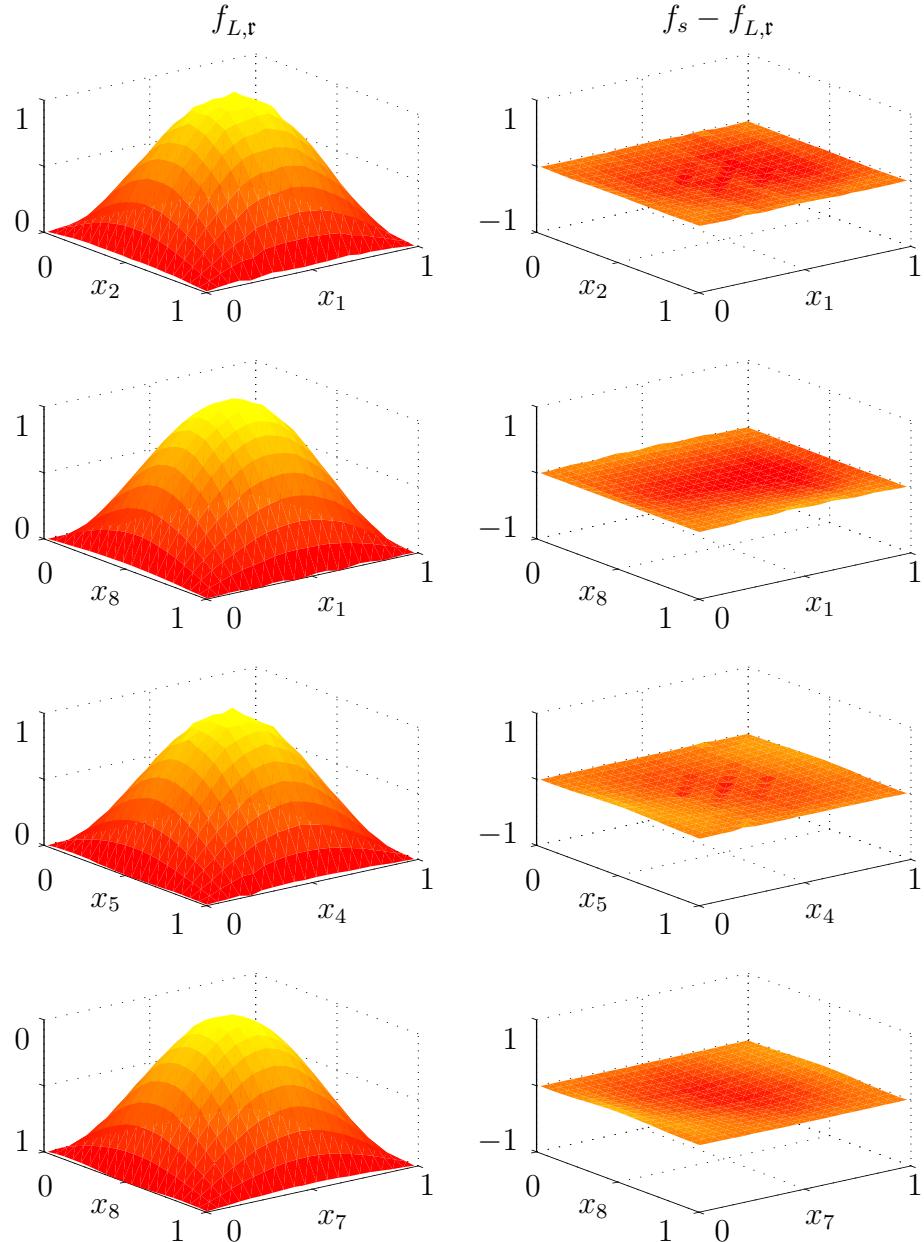
Altogether, this shows that the second model  $f_{L,r}$  resulting from (6.17) is capable of approximating higher dimensional functions and that the product ansatz for the outer function in Kolmogorov's representation is an appropriate choice. This finally allows to circumvent the exponential growth of the number of degrees of freedom with the dimensionality of the function  $f$  that has to be approximated. Also, we have seen in Section 5.1.1 that the numerical complexity for one iteration step of the BFGS-method is of the order  $\mathcal{O}(n^2 N^2) + \mathcal{O}(P n^2 N)$ , i.e. it is linear in the number of points and in particular not exponential in the dimension. However, note that a multiple precision arithmetic is required to evaluate the model and that this (second) term dominates the costs. Furthermore, the total number of iterations which are needed to compute the solution at the desired resolution might increase rapidly with  $n$ , although we did not observe this in our tests. We conclude with the remark that these are potential reasons which might still cause an occurrence of the curse of dimensionality.



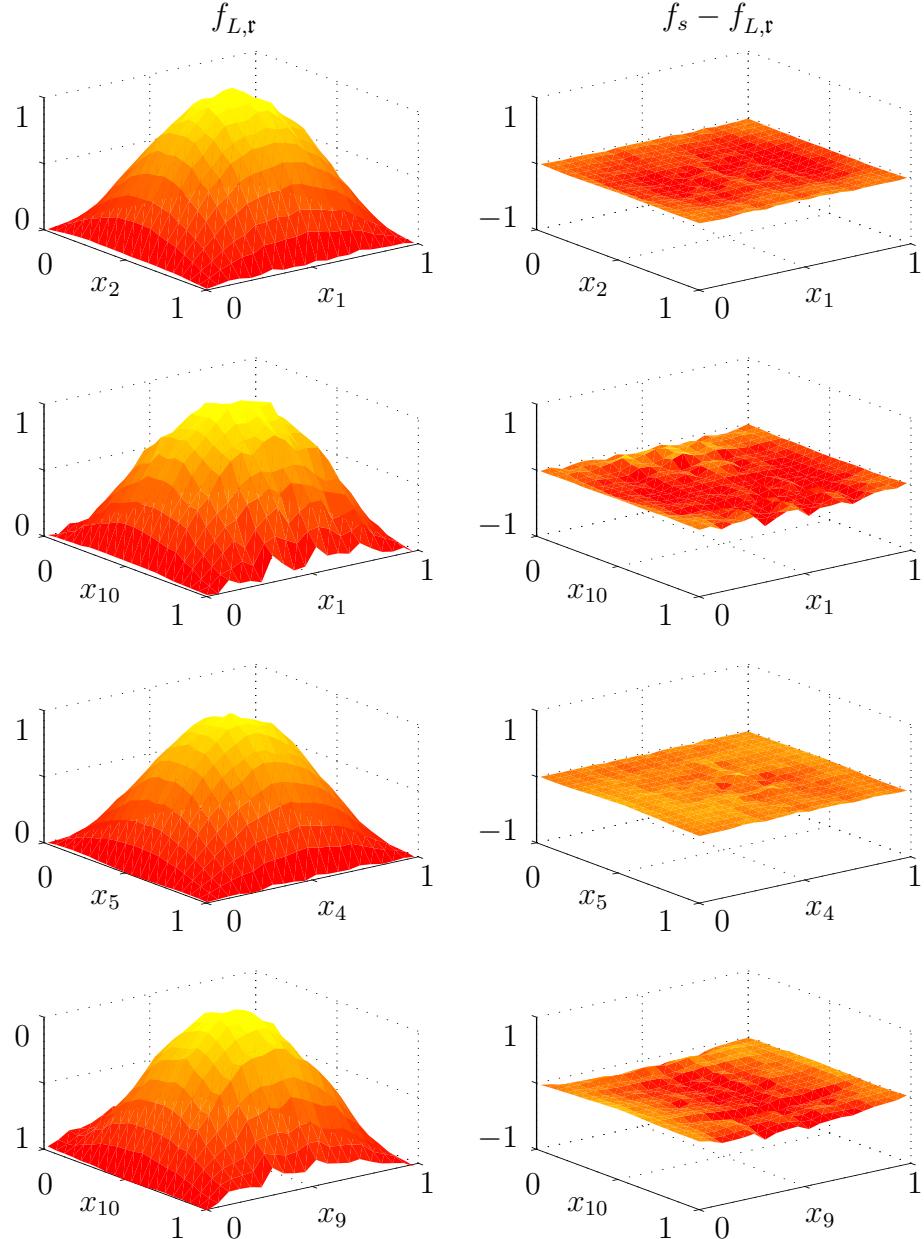
**Figure 6.24.** Reconstruction of  $f_s$  in dimension  $n = 4$  from  $P = 20000$  data points, and  $N = 32$ ,  $m = 32$ . The regularization parameter was  $\nu = 10^{-6}$ . The figures show two-dimensional hyperplanes of the graph (left) and the residual (right) for which the remaining coordinates of  $\mathbf{x} \in [0, 1]^n$  are chosen to be  $x_k = 0.5$ .



**Figure 6.25.** Graphs of  $f_{L,r}$  (left) and  $f_s - f_{L,r}$  (right) on different two-dimensional hyperplanes for the reconstruction of  $f_s$ ,  $n = 6$  with  $P = 50000$ ,  $N = 32$ ,  $m = 32$ , and  $\nu = 10^{-10}$ . For a  $(x_i, x_j)$ -plane, the remaining coordinates of  $\mathbf{x} \in [0, 1]^n$  are chosen to be  $x_k = 0.5$ .



**Figure 6.26.** The left column shows the graphs  $f_{L,r}$  for the reconstruction of  $f_s$  on two-dimensional hyperplanes in dimension  $n = 8$ . All remaining coordinates of  $\mathbf{x} \in [0, 1]^n$  are  $x_k = 0.5$ . The corresponding residuals  $f_s - f_{L,r}$  are shown in the right column. Here,  $f_{L,r}$  was computed from  $P = 80000$  data points with  $N = 32$ ,  $m = 32$ , and  $\nu = 10^{-10}$ .



**Figure 6.27.** Reconstruction of  $f_s$ ,  $n = 10$  from  $P = 100000$  points, with parameters  $N = 32$ ,  $m = 32$ , and  $\nu = 10^{-9}$ . The pictures show the function  $f_{L,r}$  (left) and the residual  $f_s - f_{L,r}$  on different two-dimensional hyperplanes. Here, the remaining coordinates of the points  $\mathbf{x} \in [0, 1]^n$  are fixed to  $x_k = 0.5$ .



# Chapter 7

# Conclusions

In this thesis we presented a Regularization Network approach to reconstruct a continuous function  $f : [0, 1]^n \rightarrow \mathbb{R}$  from its samples  $f(\mathbf{x}_j)$  at random points  $\mathbf{x}_j \in [0, 1]^n$ ,  $j = 1, \dots, P$ . The method is directly based on a new constructive version of Kolmogorov's superposition theorem which represents an  $n$ -dimensional continuous function as superposition of one inner and one outer univariate function. No higher regularity conditions on the function  $f$  had to be assumed.

To build the theoretical foundation of our model, we modified a recent result by Sprecher who derived a numerical algorithm for the implementation of  $m + 1$  external univariate functions in Kolmogorov's representation where  $m \geq 2n$ . This algorithm was adapted in a suitable way such that its convergence towards one single outer function  $\Phi$  could be proved which replaces Sprecher's  $m + 1$  outer functions. Moreover, the inner functions  $\psi_q$  are defined as translations of a single function  $\psi$  that was recently defined by Köppen<sup>1</sup> as an extension of a function which is defined on a dense subset of the real line. For this  $\psi$  existence, as well as the essential properties of continuity and monotonicity were shown.

Based on the resulting new version of Kolmogorov's superposition theorem we introduced our model to approximate  $f$  in two steps. To define the first model we replaced the outer function  $\Phi$  by an expansion in a finite basis with unknown coefficients which resulted in a linear representation of the approximand. The coefficients were determined via a variational formulation, i.e. the minimization of a regularized convex cost functional  $E^{(1)}(\cdot)$  which measures the empirical error on the random sample set. A detailed numerical analysis of this first model then revealed that the dimensionality of the function is transformed by Kolmogorov's representation into a structured oscillation of the outer function. As a consequence of this, the use of locally supported basis functions like B-splines was shown to be disadvantageous since the one-dimensional mesh resolution had to resolve the strong oscillations. However,

---

<sup>1</sup>In fact, Köppen corrected the earlier definition of Sprecher, whose function  $\psi$  was neither continuous nor monotone increasing.

the vector of Fourier coefficients turned out to be sparse which suggested the use of the Fourier basis and led to further insight into the structure of the outer function  $\Phi$ . In particular, further numerical tests revealed a product structure of the outer function. These observations complemented previous theoretical arguments which already showed that the inner sums cause an unfolding of the dimensions into a product of functions. Another important insight was the fact that the relevant frequency numbers could be computed a priori. In fact, their locations in Fourier space only depend on several model parameters, independent of the function  $f$ .

These findings motivated the definition of the second model. Here, we replaced  $\Phi$  by a product of functions  $\prod_{d=0}^n \phi^d$  and expanded each factor  $\phi^d$  in a Fourier basis with appropriate frequency numbers to capture the structured oscillations of the outer function. Since the number of basis functions was similar for each factor, the total number of degrees of freedom for the model depends linearly on the dimension  $n$ . Note that due to the strong oscillations a multiple precision arithmetic is required to evaluate both models.

The constructions for the two models were formalized by means of reproducing kernel Hilbert spaces such that the corresponding Regularization Network approaches could be defined in this general setting. For the second model the product ansatz then resulted in a cost functional  $E^{(n)}(\cdot)$  which was no longer convex and required the use of a nonlinear iterative minimizer. An analysis of the nonlinear minimizers showed that the numerical costs for one iteration step depends linearly on the number of points  $P$  and is quadratic in the dimensionality  $n$ .

The numerical analysis for the second model then revealed that the minimization of the non-convex error functional  $E^{(n)}(\cdot)$  requires a nested iteration scheme to pre-compute sufficiently good start values for the iterative solvers. For the total number of iterations we did not observe a rapid, or even exponential, increase with the dimension  $n$  in our experiments. However, we remark that this and the necessity of a multiple precision arithmetic potentially still comprise the curse of dimensionality. In conclusion, our numerical tests proved that the product model is an appropriate choice for function reconstruction and that it is capable of approximating functions up to dimension ten. Moreover, it can be used for reconstructions from large data sets.

In addition to our own contributions, we presented a comprehensive overview of the existing versions and improvements of Kolmogorov's superposition theorem which were made by different authors. This included Sprecher's constructive version, which was the basis for our considerations, and we proved that his original inner function  $\psi$  indeed is discontinuous. Furthermore, we presented a geometric interpretation of the theorem. Also, examples of approximation schemes that are closely related to Kolmogorov's superposition theorem were given. To provide the theoretical background for the construction of our Regularization Network approach, we repeated some basic facts on reproducing kernel Hilbert spaces. Then, we in-

---

troduced the Regularization Network approach in general by means of Statistical learning theory and Structural risk minimization. Moreover, we gave a Bayesian interpretation of the method.

In conclusion, the scope of the present thesis was the development of a new model for function reconstruction. This included the formulation and proof of a new constructive version of Kolmogorov's superposition theorem similar to Sprecher's result, i.e. with Köppen's explicit definition of the inner function  $\psi$  but only one single outer function. Furthermore, the underlying reproducing kernel Hilbert spaces for our models were derived which then defined the corresponding Regularization Network approaches. Finally, numerical results from an implementation of both models were presented. Here, to analyze the new models, we set up artificial test cases where we could ensure that the data set  $\mathcal{Z}$  stemmed from an underlying function  $f$ . By the choice of the function we could control the structure of the data which was important for an analysis of the models. This way we could eliminate any effects resulting from the data. In particular, these tests were important to analyze the general behaviour of the product model and prove its general capability to reconstruct functions in higher dimensions.

As an outline for further investigations, we mention the application of our Regularization Network approach to the analysis of real world data from data mining problems. Here, the next step in this direction would be to test the product model for the case when noise is added to the response variable, i.e. to consider values  $y_j = f(\mathbf{x}_j) + \xi$ , where  $\xi$  is some normally distributed random variable with expectation 0, see [127]. Then, a comparison with known methods from literature for well defined (artificial) test cases is at hand, [35, 38]. However, these tests would have significantly blown up the extend of this thesis and are the topic of future work.



## Appendix A

### A.1 Duality between RKHS's and stochastic processes

In Section 4.4 the Maximum A Posteriori interpretation of the Regularization Network approach has been introduced. To explain the duality between reproducing kernel Hilbert spaces (RKHS's) and stochastic processes that was mentioned there in more detail, we follow [127]. For details on RKHS's see Section 4.1 and the references therein.

Let  $\Omega$  be a general index set. Now, let  $X(t)$ ,  $t \in \Omega$  be real valued Gaussian random variables with zero mean and covariance

$$\mathbf{E}[X(s)X(t)] =: K(s, t), \quad (\text{A.1})$$

where  $\mathbf{E}[\xi]$  denotes the expected value of a random variable  $\xi$ , see Section 4.2 for a definition. Furthermore, we define the linear vector space of random variables

$$\mathcal{X}^o := \text{span} \{X(t) : t \in \Omega\} = \left\{ \sum_{i=1}^p a_i X(t_i) : p \in \mathbb{N}, t \in \Omega, a_i \in \mathbb{R} \right\}.$$

For random variables  $\xi_1, \xi_2 \in \mathcal{X}^o$  we define their inner product to be the value

$$\langle \xi_1, \xi_2 \rangle_{\mathcal{X}} := \mathbf{E}[\xi_1(t)\xi_2(t)].$$

The completion of  $\mathcal{X}^o$  with respect to the induced norm is a Hilbert space which will be denoted by  $\mathcal{X}$ .

Now, from its definition (A.1) and the properties of the covariance matrix we can conclude that  $K : \Omega \times \Omega \rightarrow \mathbb{R}$  is a symmetric positive definite function. Then, by Theorem 4.8, there exists a RKHS  $H(K)$  with reproducing kernel  $K$  which is the completion of

$$H^o = \text{span} \{K(\cdot, t) : t \in \Omega\} = \left\{ \sum_{i=1}^p a_i K(s, t_i) : p \in \mathbb{N}, t_i \in \Omega, a_i \in \mathbb{R} \right\}.$$

For the scalar products in the spaces  $\mathcal{X}$  and  $H(K)$  it holds

$$\langle X(s), X(t) \rangle_{\mathcal{X}} = \mathbf{E}[X(s)X(t)] = K(s, t) = \langle K(\cdot, t), K(\cdot, s) \rangle_{H(K)}. \quad (\text{A.2})$$

We define a linear mapping  $L : H(K) \rightarrow \mathcal{X}$  by

$$L \left( \sum_{i=1}^p a_i K(s, t_i) \right) := \sum_{i=1}^p a_i X(t_i) \quad \text{for } p \in \mathbb{N}, a_i \in \mathbb{R}, t_i \in \Omega,$$

and for

$$\xi := \lim_{k \rightarrow \infty} \sum_{i=1}^p a_i^k K(s, t_i^k) \quad \text{with } p \in \mathbb{N}, a_i^k \in \mathbb{R}, t_i^k \in \Omega,$$

we set

$$L(\xi) := \lim_{k \rightarrow \infty} \sum_{i=1}^p a_i^k X(t_i^k).$$

We claim the following: The mapping  $L$  is an isometric isomorphism between  $H(K)$  and  $\mathcal{X}$ .

Clearly,  $L$  is surjective and isometric. To show that  $L$  is also injective, let  $f(s) := \sum_{i=1}^p a_i K(s, t_i) \in H(K)$  and  $L(f) = 0$ . Then, since  $\langle \cdot, \cdot \rangle_{\mathcal{X}}$  is a scalar product, we get by (A.2) and (A.1)

$$0 = \langle L(f), X(s) \rangle_{\mathcal{X}} = \sum_{i=1}^p a_i \langle X(t_i), X(s) \rangle_{\mathcal{X}} = \sum_{i=1}^p a_i K(s, t_i) = f(s),$$

which shows that  $L$  is injective. For the limit case the assertion holds due to the continuity of the scalar product. Thus, the space  $H(K)$  is indeed isometrically isomorphic to  $\mathcal{X}$ .

Next, we comment on the prior  $\mathcal{P}(f)$  as it is used in Section 4.4, following the lines of [28]. To this end, let  $\{\varphi_1, \dots, \varphi_\ell\}$  be a, possibly infinite, orthogonal basis of  $H(K)$  (with respect to the scalar product  $\langle \cdot, \cdot \rangle_{H(K)}$ ), and norm

$$\|\varphi_i\|_{H(K)} = \sqrt{\gamma_i}.$$

For a function  $f \in H(K)$  given by

$$f(s) := \sum_{i=0}^{\ell} a_i \varphi_i(s), \quad a_i \in \mathbb{R},$$

the squared norm is simply

$$\|f\|_{H(K)}^2 = \sum_{i=0}^{\ell} \gamma_i a_i^2 = \mathbf{a}^T \mathbf{M} \mathbf{a},$$

where  $\mathbf{a} := (a_1, \dots, a_\ell)^T$  is the coefficient vector of  $f$  and  $\mathbf{M}$  is the diagonal matrix  $\mathbf{M} := \text{diag}(\gamma_1, \dots, \gamma_\ell)$ . Clearly, the stabilizer can also be expressed in any other reference system  $\{\tilde{\varphi}_1, \dots, \tilde{\varphi}_\ell\}$  by

$$\|f\|_{H(K)}^2 = \mathbf{b}^T \tilde{\mathbf{M}} \mathbf{b},$$

which suggests that  $\tilde{\mathbf{M}}$  can be interpreted as the covariance matrix in the alternative reference system. Thus,  $\|f\|_{H(K)}^2$  can be regarded as the Mahalanobis distance of  $f$  from the mean of the functions. Therefore, the prior

$$\mathcal{P}(f) \propto \exp(-\|f\|_{H(K)}^2) = \exp(-\mathbf{b}^T \tilde{\mathbf{M}} \mathbf{b})$$

is a multivariate Gaussian with zero mean in the RKHS  $H(K)$ , and can be related to a Gaussian prior on the function space. This captures the idea that the stabilizer effectively constraints the desired function to be in  $H(K)$ .

## A.2 Real valued formulation of first model

In Section 4.5.1 we used a real valued formulation (4.45) of the complex valued error functional  $E^{(1)}$  from (4.43). Here, we will show more details on the derivation of this formula.

To this end, we first compute the expansion of

$$\begin{aligned} 2 f_{\ell,\mathfrak{r}}(\mathbf{x}) &= \sum_{q=0}^m \left( \sum_{j=0}^{\ell} c_j \varphi_j \right) \circ \Psi_q(\mathbf{x}) + \overline{\sum_{q=0}^m \left( \sum_{j=0}^{\ell} c_j \varphi_j \right) \circ \Psi_q(\mathbf{x})} \\ &= \sum_{q=0}^m \left( \sum_{j=0}^{\ell} (c_j \varphi_j + \bar{c}_j \bar{\varphi}_j) \right) \circ \Psi_q(\mathbf{x}) = 2 \sum_{q=0}^m \phi_{\ell,\mathfrak{r}} \circ \Psi_q(\mathbf{x}) \end{aligned}$$

in terms of the real and imaginary parts of the coefficients  $c_j$  and the basis functions  $\varphi_j$ , given by

$$c_j = c_{j,\mathfrak{r}} + i c_{j,\mathfrak{i}}, \quad \varphi_j(s) = \varphi_{j,\mathfrak{r}}(s) + i \varphi_{j,\mathfrak{i}}(s), \quad j = 0, \dots, \ell.$$

This computation is straightforward. It holds

$$(c_{j,\mathfrak{r}} \pm i c_{j,\mathfrak{i}})(\varphi_{j,\mathfrak{r}}(s) \pm i \varphi_{j,\mathfrak{i}}(s)) = (c_{j,\mathfrak{r}} \varphi_{j,\mathfrak{r}}(s) - c_{j,\mathfrak{i}} \varphi_{j,\mathfrak{i}}(s)) \pm i (c_{j,\mathfrak{i}} \varphi_{j,\mathfrak{r}}(s) + c_{j,\mathfrak{r}} \varphi_{j,\mathfrak{i}}(s)),$$

and thus

$$\frac{c_j \varphi_j(s) + \bar{c}_j \bar{\varphi}_j(s)}{2} = c_{j,\mathfrak{r}} \varphi_{j,\mathfrak{r}}(s) - c_{j,\mathfrak{i}} \varphi_{j,\mathfrak{i}}(s).$$

Changing the order of summation in  $f_{\ell,\mathfrak{r}}$  directly results in the representation

$$f_{\ell,\mathfrak{r}}(\mathbf{x}) = \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{j=0}^{\ell} c_{j,e} \sum_{q=0}^m \tilde{\varphi}_{j,e}(\Psi_q(\mathbf{x})),$$

where we defined the real valued functions

$$\tilde{\varphi}_{j,\mathfrak{r}}(s) := \varphi_{j,\mathfrak{r}}(s), \quad \tilde{\varphi}_{j,\mathfrak{i}}(s) := -\varphi_{j,\mathfrak{i}}(s), \quad j = 0, \dots, \ell.$$

Next, we will derive an expansion of the regularization term  $\|f_{\ell,\mathfrak{r}}\|_{\mathcal{H}(K)}^2$  of  $E^{(1)}$  in terms of the real and imaginary parts of the coefficients. We first note that for any complex scalar product  $\langle \cdot, \cdot \rangle$  and  $0 \neq a, b \in \mathbb{C}$  the following equations hold:

$$\langle a, b \rangle = \langle (\bar{a}/\bar{a}) a, (\bar{b}/\bar{b}) b \rangle = (a/\bar{a})(\bar{b}/b) \langle \bar{a}, \bar{b} \rangle, \quad (\text{A.3})$$

and

$$\langle a, \bar{b} \rangle = \langle (b/b) a, (\bar{a}/\bar{a}) \bar{b} \rangle = (a/b)(b/a) \langle b, \bar{a} \rangle = \langle b, \bar{a} \rangle. \quad (\text{A.4})$$

If  $a = 0$  or  $b = 0$ , then  $0 = \langle a, b \rangle = \langle \bar{a}, \bar{b} \rangle = \langle a, \bar{b} \rangle = \langle b, \bar{a} \rangle$  trivially holds.

Now, with  $a = \varphi_j$ ,  $b = \varphi_k$ ,  $\langle \varphi_j, \varphi_k \rangle = \delta_{j,k} \gamma_k$ , and using (A.3) one can compute for  $f_{\ell,\mathfrak{r}} \in \mathcal{H}_\ell^{(1)}$

$$\begin{aligned} \|f_{\ell,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 &= \frac{1}{4} \left\langle f_\ell + \bar{f}_\ell^{(1)}, f_\ell + \bar{f}_\ell^{(1)} \right\rangle_{\mathcal{H}(K)} \\ &= \frac{1}{4} \langle f_\ell, f_\ell \rangle_{\mathcal{H}(K)} + \frac{1}{4} \left\langle \bar{f}_\ell^{(1)}, \bar{f}_\ell^{(1)} \right\rangle_{\mathcal{H}(K)} + \frac{1}{2} \operatorname{Re} \left\langle f_\ell, \bar{f}_\ell^{(1)} \right\rangle_{\mathcal{H}(K)} \\ &= \frac{1}{4} \sum_{k,l=0}^{\ell} c_k \bar{c}_l \langle \varphi_k, \varphi_l \rangle + \frac{1}{4} \sum_{k,l=0}^{\ell} \bar{c}_k c_l \langle \bar{\varphi}_k, \bar{\varphi}_l \rangle + \frac{1}{2} \operatorname{Re} \left( \sum_{k,l=0}^{\ell} c_k c_l \langle \varphi_k, \bar{\varphi}_l \rangle \right) \\ &= \frac{1}{2} \sum_{k,l=0}^{\ell} \delta_{k,l} c_{k,\mathfrak{r}}^2 \gamma_k + \delta_{k,l} c_{k,\mathfrak{i}}^2 \gamma_k + c_{k,\mathfrak{r}} c_{l,\mathfrak{r}} \operatorname{Re}(\langle \varphi_k, \bar{\varphi}_l \rangle) - c_{k,\mathfrak{i}} c_{l,\mathfrak{i}} \operatorname{Re}(\langle \varphi_k, \bar{\varphi}_l \rangle) \\ &\quad - c_{k,\mathfrak{i}} c_{l,\mathfrak{r}} \operatorname{Im}(\langle \varphi_k, \bar{\varphi}_l \rangle) - c_{k,\mathfrak{r}} c_{l,\mathfrak{i}} \operatorname{Im}(\langle \varphi_k, \bar{\varphi}_l \rangle) \\ &= \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k,l=0}^{\ell} c_{k,e_1} c_{l,e_2} S_{k,l}^{e_1, e_2}, \end{aligned}$$

where we abbreviated  $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{\mathcal{H}^0(K^0)}$  and defined for  $j, k = 0, \dots, \ell$ ,  $e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}$  the values

$$S_{j,k}^{e_1, e_2} := \frac{1}{2} \begin{cases} \delta_{j,k} \gamma_k + \operatorname{Re}(\langle \varphi_j, \bar{\varphi}_k \rangle_{\mathcal{H}^0(K^0)}), & e_1 = e_2 = \mathfrak{r}, \\ \delta_{j,k} \gamma_k - \operatorname{Re}(\langle \varphi_j, \bar{\varphi}_k \rangle_{\mathcal{H}^0(K^0)}), & e_1 = e_2 = \mathfrak{i}, \\ -\operatorname{Im}(\langle \varphi_j, \bar{\varphi}_k \rangle_{\mathcal{H}^0(K^0)}), & e_1 \neq e_2. \end{cases}$$

Note that (A.4) implies

$$S_{j,k}^{e_1, e_2} = S_{k,j}^{e_2, e_1}.$$

We have shown that the error functional (4.43) is given by

$$\begin{aligned} E^{(1)}(f_{\ell,\mathfrak{r}}) &= E^{(1)}((c_{0,\mathfrak{r}}, c_{0,\mathfrak{i}}, \dots, c_{\ell,\mathfrak{r}}, c_{\ell,\mathfrak{i}})) \\ &= \frac{1}{P} \sum_{j=1}^P \left( \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k=0}^{\ell} c_{k,e} \sum_{q=0}^m \tilde{\varphi}_{k,e}(\Psi_q(\mathbf{x}_j)) - y_j \right)^2 + \nu \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k,l=0}^{\ell} c_{k,e_1} c_{l,e_2} S_{k,l}^{e_1, e_2}, \end{aligned}$$

which is a convex functional with respect to the vector  $\mathbf{c}_\ell^{(1)} := (c_{0,\mathfrak{r}}, c_{0,\mathfrak{i}}, \dots, c_{\ell,\mathfrak{r}}, c_{\ell,\mathfrak{i}})^T \in \mathbb{R}^{2(\ell+1)}$ .

Now, a minimizer of  $E^{(1)}(\mathbf{c}_\ell^{(1)})$  has to fulfill for all  $\mu = 0, \dots, \ell$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$  the necessary and sufficient conditions

$$\begin{aligned} 0 &= \frac{\partial}{\partial c_{\mu,\theta}} E^{(1)}(\mathbf{c}_\ell^{(1)}) \\ &= \frac{2}{P} \sum_{j=1}^P \left( \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k=0}^{\ell} c_{k,e} \sum_{q=0}^m \tilde{\varphi}_{k,e}(\Psi_q(\mathbf{x}_j)) - y_j \right) \left( \sum_{q=0}^m \tilde{\varphi}_{\mu,\theta}(\Psi_q(\mathbf{x}_j)) \right) \\ &\quad + 2\nu \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k=0}^{\ell} c_{k,e} S_{k,\mu}^{e,\theta}, \end{aligned}$$

or equivalently

$$\begin{aligned} \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} \sum_{k=0}^{\ell} c_{k,e} \left[ P\nu S_{k,\mu}^{e,\theta} + \sum_{j=1}^P \left( \sum_{q=0}^m \tilde{\varphi}_{k,e}(\Psi_q(\mathbf{x}_j)) \right) \left( \sum_{q=0}^m \tilde{\varphi}_{\mu,\theta}(\Psi_q(\mathbf{x}_j)) \right) \right] \\ = \sum_{j=1}^P y_j \left( \sum_{q=0}^m \tilde{\varphi}_{\mu,\theta}(\Psi_q(\mathbf{x}_j)) \right). \end{aligned}$$

Defining the data matrix  $\mathbf{B} \in \mathbb{R}^{P \times 2(\ell+1)}$  by

$$\mathbf{B} := \sum_{q=0}^m \begin{pmatrix} \tilde{\varphi}_{0,\mathfrak{r}}(\Psi_q(\mathbf{x}_1)) & \tilde{\varphi}_{0,\mathfrak{i}}(\Psi_q(\mathbf{x}_1)) & \dots & \tilde{\varphi}_{\ell,\mathfrak{r}}(\Psi_q(\mathbf{x}_1)) & \tilde{\varphi}_{\ell,\mathfrak{i}}(\Psi_q(\mathbf{x}_1)) \\ \vdots & & & & \vdots \\ \tilde{\varphi}_{0,\mathfrak{r}}(\Psi_q(\mathbf{x}_P)) & \tilde{\varphi}_{0,\mathfrak{i}}(\Psi_q(\mathbf{x}_P)) & \dots & \tilde{\varphi}_{\ell,\mathfrak{r}}(\Psi_q(\mathbf{x}_P)) & \tilde{\varphi}_{\ell,\mathfrak{i}}(\Psi_q(\mathbf{x}_P)) \end{pmatrix},$$

the regularization matrix by

$$\mathbf{S} := \begin{pmatrix} S_{0,0}^{\mathfrak{r},\mathfrak{r}} & S_{0,0}^{\mathfrak{r},\mathfrak{i}} & \dots & S_{0,\ell}^{\mathfrak{r},\mathfrak{r}} & S_{0,\ell}^{\mathfrak{r},\mathfrak{i}} \\ S_{0,0}^{\mathfrak{i},\mathfrak{r}} & S_{0,0}^{\mathfrak{i},\mathfrak{i}} & \dots & S_{0,\ell}^{\mathfrak{i},\mathfrak{r}} & S_{0,\ell}^{\mathfrak{i},\mathfrak{i}} \\ \vdots & & & & \vdots \\ S_{\ell,0}^{\mathfrak{r},\mathfrak{r}} & S_{\ell,0}^{\mathfrak{r},\mathfrak{i}} & \dots & S_{\ell,\ell}^{\mathfrak{r},\mathfrak{r}} & S_{\ell,\ell}^{\mathfrak{r},\mathfrak{i}} \\ S_{\ell,0}^{\mathfrak{i},\mathfrak{r}} & S_{\ell,0}^{\mathfrak{i},\mathfrak{i}} & \dots & S_{\ell,\ell}^{\mathfrak{i},\mathfrak{r}} & S_{\ell,\ell}^{\mathfrak{i},\mathfrak{i}} \end{pmatrix} \in \mathbb{R}^{2(\ell+1) \times 2(\ell+1)},$$

and the data vector  $\mathbf{y} = (y_1, \dots, y_P)^T \in \mathbb{R}^P$ , these conditions are equivalent to the following system of linear equations:

$$(\mathbf{B}^T \mathbf{B} + \nu P \mathbf{S}) \mathbf{c}_\ell^{(1)} = \mathbf{B}^T \mathbf{y}.$$

Thus, the minimizer of  $E^{(1)}(\mathbf{c}_\ell^{(1)})$  is given by the solution of this system.

### A.3 Real valued formulation of second model

Next, we will compute the expansion of  $f_{L,\mathfrak{r}} = (f_L + \bar{f}_L^{(n)})/2$  that results from a splitting of all complex coefficients  $c_j^d$  and basis functions  $\varphi_j^d$ ,  $d = 0, \dots, n$ ,  $j = 0, \dots, \ell_d$  into their real and imaginary part:

$$c_j^d = c_{j,\mathfrak{r}}^d + i c_{j,\mathfrak{i}}^d, \quad \text{and} \quad \varphi_j^d(s) = \varphi_{j,\mathfrak{r}}^d(s) + i \varphi_{j,\mathfrak{i}}^d(s).$$

Here,  $f_L$  is given by (4.36).

We start with some useful formulas. Remember that for a product of general sums it holds

$$\prod_{d=0}^n \left( \sum_{j=0}^N A_j^d \right) = \sum_{j_0, \dots, j_n=0}^N A_{j_0}^0 \cdots A_{j_n}^n.$$

Using this fact and defining  $e(0) := \mathfrak{r}$ , and  $e(1) := \mathfrak{i}$ , one can compute the product of  $(n+1)$  complex numbers  $(C_{\mathfrak{r}}^0 + i C_{\mathfrak{i}}^0), \dots, (C_{\mathfrak{r}}^n + i C_{\mathfrak{i}}^n) \in \mathbb{C}$  by

$$\begin{aligned} \prod_{d=0}^n (C_{\mathfrak{r}}^d + i C_{\mathfrak{i}}^d) &= \sum_{j_0, \dots, j_n=0}^1 i^{(j_0+\dots+j_n)} \prod_{d=0}^n C_{e(j_d)}^d \\ &= \sum_{\substack{j_0, \dots, j_n=0 \\ (\sum_{d=0}^n j_d) \text{ even}}}^1 (-1)^{\frac{1}{2}(\sum_{d=0}^n j_d)} \prod_{d=0}^n C_{e(j_d)}^d + i \sum_{\substack{j_0, \dots, j_n=0 \\ (\sum_{d=0}^n j_d) \text{ odd}}}^1 (-1)^{\frac{1}{2}((\sum_{d=0}^n j_d)-1)} \prod_{d=0}^n C_{e(j_d)}^d. \end{aligned} \tag{A.5}$$

Note that we always define 0 to be an even number since a division by 2 produces no remainder.

Now, consider the functions from (4.27)

$$\begin{aligned}
\Phi_{j_0, \dots, j_n}(\mathbf{x}) &= \sum_{q=0}^m \left( \prod_{d=0}^n (\varphi_{j_d, \mathfrak{r}}^d + i \varphi_{j_d, \mathfrak{i}}^d) \circ \Psi_q(\mathbf{x}) \right) \\
&= \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ even}}}^1 (-1)^{\frac{1}{2}(\sum_{d=0}^n k_d)} \sum_{q=0}^m \left( \prod_{d=0}^n \varphi_{j_d, e(k_d)}^d (\Psi_q(\mathbf{x})) \right) \\
&\quad + i \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ odd}}}^1 (-1)^{\frac{1}{2}((\sum_{d=0}^n k_d)-1)} \sum_{q=0}^m \left( \prod_{d=0}^n \varphi_{j_d, e(k_d)}^d (\Psi_q(\mathbf{x})) \right) \\
&= \operatorname{Re}(\Phi_{j_0, \dots, j_n}(\mathbf{x})) + \operatorname{Im}(\Phi_{j_0, \dots, j_n}(\mathbf{x})).
\end{aligned}$$

For an element  $f_L \in \mathcal{K}_L$ , the coefficient of  $\Phi_{j_0, \dots, j_n}$  in its expansion with respect to the generating system  $\{\Phi_{j_0, \dots, j_n} : j_d = 0, \dots, \ell_d, d = 0, \dots, n\}$  is the product  $c_{j_0, \dots, j_n} = c_{j_0}^0 \cdots c_{j_n}^n \in \mathbb{C}$ , see (4.48). Thus, it holds with (A.5)

$$\begin{aligned}
\operatorname{Re}(c_{j_0, \dots, j_m} \Phi_{j_0, \dots, j_n}) &= (\operatorname{Re}(c_{j_0, \dots, j_m}) \operatorname{Re}(\Phi_{j_0, \dots, j_n})) - (\operatorname{Im}(c_{j_0, \dots, j_m}) \operatorname{Im}(\Phi_{j_0, \dots, j_n})) \\
&= \left( \sum_{\substack{l_0, \dots, l_n=0 \\ (\sum_{d=0}^n l_d) \text{ even}}}^1 (-1)^{\frac{1}{2}(\sum_{d=0}^n l_d)} \prod_{d=0}^n c_{j_d, e(l_d)}^d \right) \\
&\quad \left( \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ even}}}^1 (-1)^{\frac{1}{2}(\sum_{d=0}^n k_d)} \sum_{q=0}^m \left( \prod_{d=0}^n \varphi_{j_d, e(k_d)}^d (\Psi_q(\mathbf{x})) \right) \right) \\
&\quad - \left( \sum_{\substack{l_0, \dots, l_n=0 \\ (\sum_{d=0}^n l_d) \text{ odd}}}^1 (-1)^{\frac{1}{2}((\sum_{d=0}^n l_d)-1)} \prod_{d=0}^n c_{j_d, e(l_d)}^d \right) \\
&\quad \left( \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ odd}}}^1 (-1)^{\frac{1}{2}((\sum_{d=0}^n k_d)-1)} \sum_{q=0}^m \left( \prod_{d=0}^n \varphi_{j_d, e(k_d)}^d (\Psi_q(\mathbf{x})) \right) \right).
\end{aligned}$$

Next, we define the expression

$$\delta_{k_0, \dots, k_n}^{l_0, \dots, l_n} := \begin{cases} 1 & \text{if } (\sum_{d=0}^n l_d) \text{ and } (\sum_{d=0}^n k_d) \text{ even,} \\ -1 & \text{if } (\sum_{d=0}^n l_d) \text{ and } (\sum_{d=0}^n k_d) \text{ odd,} \\ 0 & \text{else,} \end{cases}$$

and finally get for the function  $f_{L,\mathfrak{r}} = (f_L + \bar{f}_L^{(n)})/2$  with  $f_L \in \mathcal{K}_L$  the following expansion in terms of the real valued functions  $(\varphi_{j,\mathfrak{r}}^d \circ \Psi_q)$  and  $(\varphi_{j,\mathfrak{i}}^d \circ \Psi_q)$ :

$$\begin{aligned}
f_{L,\mathfrak{r}}(\mathbf{x}) &= \operatorname{Re} \left( \sum_{j_0, \dots, j_n=0}^{\ell_0, \dots, \ell_n} c_{j_0, \dots, j_m} \Phi_{j_0, \dots, j_n} \right) = \sum_{q=0}^m \operatorname{Re} \left( \prod_{d=0}^n \phi^d (\Psi_q(\mathbf{x})) \right) \\
&= \sum_{\substack{l_0, \dots, l_n=0 \\ (\sum_{d=0}^n l_d) \text{ even}}}^1 \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ even}}}^1 (-1)^{\frac{1}{2} \sum_{d=0}^n (l_d + k_d)} \sum_{q=0}^m \sum_{j_0, \dots, j_n=0}^{\ell_0, \dots, \ell_n} \prod_{d=0}^n c_{j_d, e(l_d)}^d \varphi_{j_d, e(k_d)}^d (\Psi_q(\mathbf{x})) \\
&\quad - \sum_{\substack{l_0, \dots, l_n=0 \\ (\sum_{d=0}^n l_d) \text{ odd}}}^1 \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ odd}}}^1 (-1)^{\frac{1}{2} \sum_{d=0}^n (l_d + k_d)} \sum_{q=0}^m \sum_{j_0, \dots, j_n=0}^{\ell_0, \dots, \ell_n} \prod_{d=0}^n c_{j_d, e(l_d)}^d \varphi_{j_d, e(k_d)}^d (\Psi_q(\mathbf{x})) \\
&= \sum_{l_0, \dots, l_n=0}^1 \sum_{k_0, \dots, k_n=0}^1 (-1)^{\frac{1}{2} \sum_{d=0}^n (l_d + k_d)} \delta_{k_0, \dots, k_n}^{l_0, \dots, l_n} \sum_{q=0}^m \prod_{d=0}^n \left( \sum_{j_d=0}^{\ell_d} c_{j_d, e(l_d)}^d \varphi_{j_d, e(k_d)}^d (\Psi_q(\mathbf{x})) \right). \tag{A.6}
\end{aligned}$$

For the derivative of  $f_{L,\mathfrak{r}}$  with respect to the real coefficients  $c_{\mu,\theta}^\delta$ ,  $\delta = 0, \dots, n$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$ ,  $\mu = 0, \dots, \ell_d$ , one can compute

$$\begin{aligned}
\frac{\partial}{\partial c_{\mu,\theta}^\delta} f_{L,\mathfrak{r}}(\mathbf{x}) &= \frac{\partial}{\partial c_{\mu,\theta}^\delta} \operatorname{Re} \left[ \sum_{q=0}^m \phi^\delta (\Psi_q(\mathbf{x})) \prod_{\substack{d=0 \\ d \neq \delta}}^n \phi^d (\Psi_q(\mathbf{x})) \right] \\
&= \frac{1}{2} \frac{\partial}{\partial c_{\mu,\theta}^\delta} \sum_{q=0}^m \left[ \operatorname{Re} (\phi^\delta (\Psi_q(\mathbf{x}))) \operatorname{Re} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \phi^d (\Psi_q(\mathbf{x})) \right) \right. \\
&\quad \left. - \operatorname{Im} (\phi^\delta (\Psi_q(\mathbf{x}))) \operatorname{Im} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \phi^d (\Psi_q(\mathbf{x})) \right) \right] \\
&= \sum_{q=0}^m \left[ \left\{ \begin{array}{ll} \varphi_{\mu,\mathfrak{r}}^\delta (\Psi_q(\mathbf{x})), & \theta = \mathfrak{r} \\ -\varphi_{\mu,\mathfrak{i}}^\delta (\Psi_q(\mathbf{x})), & \theta = \mathfrak{i} \end{array} \right\} \operatorname{Re} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \phi^d (\Psi_q(\mathbf{x})) \right) \right. \\
&\quad \left. - \left\{ \begin{array}{ll} \varphi_{\mu,\mathfrak{i}}^\delta (\Psi_q(\mathbf{x})), & \theta = \mathfrak{r} \\ \varphi_{\mu,\mathfrak{r}}^\delta (\Psi_q(\mathbf{x})), & \theta = \mathfrak{i} \end{array} \right\} \operatorname{Im} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \phi^d (\Psi_q(\mathbf{x})) \right) \right] =: \sum_{q=0}^m D_{q,\mu}^{\delta,\theta}(\mathbf{x}). \tag{A.7}
\end{aligned}$$

Here, the terms

$$\operatorname{Re} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \phi^d(\Psi_q(\mathbf{x})) \right) \quad \text{and} \quad \operatorname{Im} \left( \prod_{\substack{d=0 \\ d \neq \delta}}^n \phi^d(\Psi_q(\mathbf{x})) \right)$$

can be computed analog to (A.6).

Next, we derive a representation of the regularization term  $\|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2$  from (4.50). To this end, remember (A.3), (4.49) and again (A.5). Furthermore, define for  $e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}$ ,  $d = 0, \dots, n$ , and  $k, l = 0, \dots, \ell_d$  the real numbers

$$S_{d,k,l}^{e_1, e_2, \mathfrak{r}} := \begin{cases} \operatorname{Re}(\langle \varphi_k^d, \bar{\varphi}_l^d \rangle_{\mathcal{H}^d(k_d)}), & e_1 = e_2 = \mathfrak{r}, \\ -\operatorname{Re}(\langle \varphi_k^d, \bar{\varphi}_l^d \rangle_{\mathcal{H}^d(k_d)}), & e_1 = e_2 = \mathfrak{i}, \\ \operatorname{Im}(\langle \varphi_k^d, \bar{\varphi}_l^d \rangle_{\mathcal{H}^d(k_d)}), & e_1 \neq e_2, \end{cases}$$

and

$$S_{d,k,l}^{e_1, e_2, \mathfrak{i}} := \begin{cases} \operatorname{Im}(\langle \varphi_k^d, \bar{\varphi}_l^d \rangle_{\mathcal{H}^d(k_d)}), & e_1 = e_2 = \mathfrak{r}, \\ -\operatorname{Im}(\langle \varphi_k^d, \bar{\varphi}_l^d \rangle_{\mathcal{H}^d(k_d)}), & e_1 = e_2 = \mathfrak{i}, \\ \operatorname{Re}(\langle \varphi_k^d, \bar{\varphi}_l^d \rangle_{\mathcal{H}^d(k_d)}), & e_1 \neq e_2. \end{cases}$$

Note that (A.4) implies  $S_{d,k,l}^{e_1, e_2, \mathfrak{r}} = S_{d,l,k}^{e_2, e_1, \mathfrak{r}}$  and  $S_{d,k,l}^{e_1, e_2, \mathfrak{i}} = S_{d,l,k}^{e_2, e_1, \mathfrak{i}}$ . Then it holds

$$\begin{aligned} \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 &= \frac{1}{4} \left\langle f_L + \bar{f}_L^{(n)}, f_L + \bar{f}_L^{(n)} \right\rangle_{\mathcal{H}(K)} = \frac{1}{2} \langle f_L, f_L \rangle_{\mathcal{H}(K)} + \frac{1}{2} \operatorname{Re} \left\langle f_L, \bar{f}_L^{(n)} \right\rangle_{\mathcal{H}(K)} \\ &= \frac{1}{2} \prod_{d=0}^n \left( \sum_{k=0}^{\ell_d} (c_k^d)^2 \gamma_k^d \right) + \frac{1}{2} \operatorname{Re} \left( \prod_{d=0}^n \left( \sum_{k,l=0}^{\ell_d} c_k^d c_l^d \langle \varphi_k^d, \bar{\varphi}_l^d \rangle \right) \right) \\ &= \frac{1}{2} \prod_{d=0}^n \left( \sum_{k=0}^{\ell_d} ((c_{k,\mathfrak{r}}^d)^2 + (c_{k,\mathfrak{i}}^d)^2) \gamma_k^d \right) \\ &\quad + \frac{1}{2} \operatorname{Re} \left( \prod_{d=0}^n \left[ \sum_{k,l=0}^{\ell_d} \left( \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} c_{k,e_1}^d c_{l,e_2}^d S_{d,k,l}^{e_1, e_2, \mathfrak{r}} \right) + i \left( \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} c_{k,e_1}^d c_{l,e_2}^d S_{d,k,l}^{e_1, e_2, \mathfrak{i}} \right) \right] \right) \\ &= \frac{1}{2} \prod_{d=0}^n \left( \sum_{k=0}^{\ell_d} ((c_{k,\mathfrak{r}}^d)^2 + (c_{k,\mathfrak{i}}^d)^2) \gamma_k^d \right) \\ &\quad + \frac{1}{2} \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ even}}}^1 (-1)^{\frac{1}{2}(\sum_{d=0}^n k_d)} \prod_{d=0}^n \left( \sum_{k,l=0}^{\ell_d} \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} c_{k,e_1}^d c_{l,e_2}^d S_{d,k,l}^{e_1, e_2, e(k_d)} \right). \end{aligned}$$

Here, for the third equality we used the fact that  $\bar{f}_L^{(n)} \in \mathcal{K}_L$ . The partial derivatives with respect to the real coefficients  $c_{\mu,\theta}^\delta$ ,  $\delta = 0, \dots, n$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$ , and  $\mu = 0, \dots, \ell_\delta$  are given by

$$\begin{aligned} \frac{\partial}{\partial c_{\mu,\theta}^\delta} \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 &= c_{\mu,\theta}^\delta \gamma_\mu^\delta \prod_{\substack{d=0 \\ d \neq \delta}}^n \left( \sum_{k=0}^{\ell_d} ((c_{k,\mathfrak{r}}^d)^2 + (c_{k,\mathfrak{i}}^d)^2) \gamma_k^d \right) \\ &+ \sum_{\substack{k_0, \dots, k_n=0 \\ (\sum_{d=0}^n k_d) \text{ even}}}^1 (-1)^{\frac{1}{2} \sum k_d} \left( \sum_{k=0}^{\ell_\delta} \sum_{e \in \{\mathfrak{r}, \mathfrak{i}\}} c_{k,e}^\delta S_{\delta,\mu,k}^{\theta,e,e(k_d)} \right) \prod_{\substack{d=0 \\ d \neq \delta}}^n \left( \sum_{k,l=0}^{\ell_d} \sum_{e_1, e_2 \in \{\mathfrak{r}, \mathfrak{i}\}} c_{k,e_1}^d c_{l,e_2}^d S_{d,k,l}^{e_1, e_2, e(k_d)} \right), \end{aligned} \quad (\text{A.8})$$

where, as in the previous cases, the sum  $\sum k_d$  in the exponent is built over all  $d = 0, \dots, n$ . Thus, a minimizer of  $E^{(n)}(\mathbf{c}_L^{(n)})$  has to fulfill for all  $d = 0, \dots, n$ ,  $\theta \in \{\mathfrak{r}, \mathfrak{i}\}$ ,  $\mu = 0, \dots, \ell_\delta$  the necessary condition

$$0 = \frac{\partial}{\partial c_{\mu,\theta}^\delta} E^{(n)}(\mathbf{c}_L^{(n)}) = \nu \left( \frac{\partial}{\partial c_{\mu,\theta}^\delta} \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 \right) + \frac{2}{P} \sum_{j=1}^P (f_{L,\mathfrak{r}}(\mathbf{x}_j) - y_j) \left( \sum_{q=0}^m D_{q,\mu}^{\delta,\theta}(\mathbf{x}_j) \right), \quad (\text{A.9})$$

where  $f_{L,\mathfrak{r}}$ ,  $(\partial \|f_{L,\mathfrak{r}}\|_{\mathcal{H}(K)}^2 / \partial c_{\mu,\theta}^\delta)$  and  $\sum_{q=0}^m D_{q,\mu}^{\delta,\theta}(\mathbf{x}_j)$  are given by (A.6), (A.7), and (A.8), respectively. Note that all terms depend nonlinearly on the real coefficient vector

$$\mathbf{c}_L^{(n)} := (c_{0,\mathfrak{r}}^0, c_{0,\mathfrak{i}}^0, \dots, c_{\ell_0,\mathfrak{r}}^0, c_{\ell_0,\mathfrak{i}}^0, \dots, c_{0,\mathfrak{r}}^n, c_{0,\mathfrak{i}}^n, \dots, c_{\ell_0,\mathfrak{r}}^n, c_{\ell_0,\mathfrak{i}}^n)^T \in \mathbb{R}^{2L},$$

with  $L := \sum_{d=0}^n (\ell_d + 1)$ . Thus,

$$\nabla E^{(n)}(\mathbf{c}_L^{(n)}) = 0 \in \mathbb{R}^{2L}$$

is a system of nonlinear equations for  $\mathbf{c}_L^{(n)} \in \mathbb{R}^{2L}$ .

## A.4 Further definitions

### III–posed problems

We introduce a general classification of mathematical problems into well–posed and ill–posed problems which is due to Hadamard. To give a formal definition of these terms, in [120] the following problem is considered: Let  $A : X \rightarrow Y$  be an operator on the metric spaces  $X$ ,  $Y$ , and let  $g \in Y$  be given. Then find a solution of the operator equation

$$Au = g, \quad u \in X, \quad g \in Y. \quad (\text{P})$$

**Definition A.1.** According to Hadamard, problem (P) is said to be **well–posed** if the following conditions hold:

- (i) For each  $f \in Y$  there exists a unique solution of  $(P)$ ,
- (ii) and the solution is stable under perturbation of the righthand side of  $(P)$ , i.e. the operator  $A^{-1}$  is defined on all of  $Y$  and is continuous.

If any of these condition does not hold, the problem is called **ill-posed**.

In this thesis, the problem is to reconstruct a function  $f : [0, 1]^n \rightarrow \mathbb{R}$ , or the parameters which determine  $f$ , from discrete data samples  $\mathcal{Z}$ , cf. Chapter 4. Obviously, this problem is ill-posed, see [38] for details.

### Definition of VC-dimension and $V_\gamma$ -dimension

Next, we give the formal definitions of the VC-dimension and the  $V_\gamma$ -dimension which have been introduced in Section 4.3 as measures for the complexity, or capacity, of a set of functions  $H$ . The following definitions can be found in [28].

First, let

$$\theta(t) := \begin{cases} 0, & t \leq 0, \\ 1, & t > 0, \end{cases}$$

and let  $V : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be a loss function, like e.g.  $V(f(\mathbf{x}), y) := (f(\mathbf{x}) - y)^2$ . Then, the VC-dimension was first defined for the case of indicator functions and then was extended to real valued functions.

**Definition A.2.** *The VC-dimension of a set  $\{\theta(f(\mathbf{x})) : f \in H\}$  of indicator functions is the maximum number  $d$  of vectors  $\mathbf{x}_1, \dots, \mathbf{x}_d$  that can be separated into two classes in all  $2^d$  ways using functions of the set.*

*If, for any number  $N$ , it is possible to find  $N$  points  $\mathbf{x}_1, \dots, \mathbf{x}_N$  that can be separated in all the  $2^N$  ways, we will say that the VC-dimension of the set is finite.*

This allow for the definition of the VC-dimension of the loss function  $V$  in  $H$ :

**Definition A.3.** *Let  $A \leq V(y, f(\mathbf{x})) \leq B$ ,  $f \in H$ , with  $A$  and  $B < \infty$ . The VC-dimension of the set  $\{V(y, f(\mathbf{x})) : f \in H\}$  (or of  $V$  in  $H$ ) is defined as the VC-dimension of the set of indicator functions  $\{\theta(V(y, f(\mathbf{x})) - \alpha) : \alpha \in (A, B)\}$ .*

A further extension of this definition is given by

**Definition A.4.** *Let  $A \leq V(y, f(\mathbf{x})) \leq B$ ,  $f \in H$ , with  $A$  and  $B < \infty$ . The  $V_\gamma$ -dimension of  $V$  in  $H$  (of the set  $\{V(y, f(\mathbf{x})) : f \in H\}$ ) is defined as the maximum number  $d$  of vectors  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_d, y_d)$  that can be separated into two classes in all  $2^d$  possible ways using rules*

$$\begin{aligned} \text{class 1 if: } & V(y_i, f(\mathbf{x}_i)) \geq s + \gamma \\ \text{class 2 if: } & V(y_i, f(\mathbf{x}_i)) \leq s - \gamma \end{aligned}$$

$n$	$P_e$
2	1 000
3	1 000
4	1 000
5	5 000
6	5 000
7	5 000
8	10 000
9	10 000
10	100 000
11	100 000
12	1 000 000

**Table A.1.** For each dimension  $n$ , the table lists the size  $P_e$  of the test set  $\mathcal{Z}_e$  which is used to calculate the errors  $e_{\mathcal{Z}_e}$  and  $e_{\mathcal{Z}_e}^\infty$ , see Chapter 6. For increasing dimensions the set becomes larger. The choices are based on numerical experiments, i.e. for larger  $P_e$  and fixed  $n$  the values of  $e_{\mathcal{Z}_e}$  and  $e_{\mathcal{Z}_e}^\infty$  did not significantly change.

for  $f \in H$  and some  $s \geq 0$ . If, for any number  $N$ , it is possible to find  $N$  points  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$  that can be separated in all the  $2^N$  ways, we will say that the  $V_\gamma$ -dimension of  $V$  in  $H$  is finite.

Note that for  $\gamma = 0$  this definition becomes the *VC*-dimension from Definition A.3. Intuitively, for  $\gamma > 0$  the “rule” for separating points is more restrictive than the rule in the case  $\gamma = 0$ . It requires that there is a “margin” between the points: points for which  $V(y, f(\mathbf{x}))$  is between  $s - \gamma$  and  $s + \gamma$  are not classified. As a consequence, the  $V_\gamma$ -definition is a decreasing function of  $\gamma$  and in particular it is smaller than the *VC*-dimension.

# Bibliography

- [1] V. ARNOLD, *On functions of three variables*, Dokl. Akad. Nauk SSSR, 114 (1957), pp. 679–681. English translation: American Math. Soc. Transl (2), 28, 51-54, 1963.
- [2] ——, *On the representation of functions of several variables by superpositions of functions of fewer variables*, Mat. Prosveshchenie, 3 (1958), pp. 41–61.
- [3] ——, *On the representation of continuous functions of three variables by superpositions of continuous functions of two variables*, Mat. Sb., 48 (1959), pp. 3–74. English translation: American Math. Soc. Transl. (2), 28, 61-147, 1963.
- [4] N. ARONSZAJN, *Theory of reproducing kernels*, Transactions of the American Mathematical Society, 68 (1950), pp. 337–404.
- [5] A. BARRON, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Transactions on Information Theory, 39 (1993), pp. 930–945.
- [6] V. BEIU, *On Kolmogorov's superpositions and boolean functions*, in SBRN, 1998, pp. 55–60.
- [7] V. BEIU, A. ZAWADSKI, R. ANDONIE, AND S. AUNET, *Using Kolmogorov inspired gates for low power nanoelectronics*, in IWANN, 2005, pp. 438–445.
- [8] R. E. BELLMAN, *Adaptive control processes - A guided tour*, Princeton University Press, Princeton, New Jersey, U.S.A., 1961.
- [9] V. BERN AND A. ZAWADZKI, *On Kolmogorov's superpositions: novel gates and circuits for nanoelectronics?*, vol. 1, 2005, pp. 651–656.
- [10] Y. BODYANSKIY, Y. GORSHKOV, V. KOLODYAZHNIY, AND V. POYEDYNTSEVA, *Neuro-fuzzy Kolmogorov's network*, in ICANN (2), 2005, pp. 1–6.
- [11] Y. BODYANSKIY, V. KOLODYAZHNIY, AND P. OTTO, *Neuro-fuzzy Kolmogorov's network for time series prediction and pattern classification*, in KI, 2005, pp. 191–202.

- [12] J. BRAUN AND M. GRIEBEL, *On a constructive proof of Kolmogorov's superposition theorem*, Constructive Approximation, (2009). Online First.
- [13] R. BRIGOLA, *Fourieranalysis, Distributionen und Anwendungen*, Friedrich Vieweg & Sohn, Braunschweig, 1997.
- [14] H. J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numerica, 13 (2004), pp. 1–123.
- [15] R. CAFLISCH, W. MOROKO, AND A. OWEN, *Valuation of mortgage backed securities using brownian bridges to reduce effective dimension*, J. Comput. Finance, 1 (1997), pp. 27–46.
- [16] K. CIOS, W. PEDRYCZ, AND R. W. SWINIARSKI, *Data mining methods for knowledge discovery*, Kluwer Academic Publishers, Norwell, MA, USA, 1998.
- [17] M. COPPEJANS, *On Kolmogorov's representation of functions of several variables by functions of one variable*, Journal of Econometrics, 123 (2004), pp. 1–31.
- [18] F. CUCKER AND S. SMALE, *On the mathematical foundations of learning*, Bulletin of the AMS, 39 (2001), pp. 1–49.
- [19] ———, *Best choices for regularization parameters in learning theory: On the bias-variance problem*, Foundations of Computational Mathematics, 2 (2002), pp. 413–428.
- [20] G. CYBENKO, *Approximation by superpositions of a sigmoidal function*, Mathematics of Control, Signals, and Systems (MCSS), 2 (1989), pp. 303–314.
- [21] C. DE BOOR, *A Practical Guide to Splines*, Springer Verlag, New York, 1978.
- [22] R. J. P. DE FIGUEIREDO, *Implications and applications of Kolmogorov's superposition theorem*, IEEE Transactions on Automatic Control, AC-25 (1980), pp. 1227–1231.
- [23] L. DEVROYE, L. GYÖRFI, AND G. LUGOSI, *A Probabilistic Theory of Pattern Recognition (Stochastic Modelling and Applied Probability)*, Springer, 1997.
- [24] J. DICK, I. SLOAN, X. WANG, AND H. WOŹNIAKOWSKI, *Liberating the weight*, Journal of Complexity, 20 (2004), pp. 593–623.
- [25] R. DOSS, *Representations of continuous functions of several variables*, American Journal of Mathematics, 98 (1976), pp. 375–383.

- [26] B. EFRON AND C. STEIN, *The jackknife estimate of variance*, Annals of Statist., 9 (1981), pp. 586–596.
- [27] T. EVGENIOU AND M. PONTIL, *On the  $V_\gamma$ -dimension for regression in reproducing kernel Hilbert spaces*, A.i. memo, MIT Artificial Intelligence Lab, (1999).
- [28] T. EVGENIOU, M. PONTIL, AND T. POGGIO, *Regularization networks and support vector machines*, Advances in Computational Mathematics, 13 (2000), pp. 1–50.
- [29] T. EVGENIOUS, T. POGGIO, M. PONTIL, AND A. VERRI, *Regularization and statistical learning theory for data analysis*, Computational Statistics & Data Analysis, 38 (2002), pp. 421–432.
- [30] G. F. FASSHAUER, *Meshfree Approximation Methods with MATLAB*, World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2007.
- [31] C. FEUERSÄNGER AND M. GRIEBEL, *Principal manifold learning by sparse grids*, Computing, 85 (2009).
- [32] T. D. FRANK, *Nonlinear Fokker-Planck Equations: Fundamentals and Applications*, Springer Series in Synergetics, Springer, 2005.
- [33] B. FRIDMAN, *An improvement on the smoothness of the functions in Kolmogorov's theorem on superpositions*, Dokl. Akad. Nauk SSSR, 177 (1967), pp. 1019–1022. English translation: Soviet Math. Dokl. (8), 1550-1553, 1967.
- [34] J. FRIEDMAN AND W. STÜTZLE, *Projection pursuit regression*, J. Amer. Statist. Assoc., 76 (1981), pp. 817–823.
- [35] J. H. FRIEDMAN, *Multivariate adaptive regression splines*, The Annals of Statistics, 19 (1991), pp. 1–141.
- [36] H. L. FRISCH, C. BORZI, G. ORD, J. K. PERCUS, AND G. O. WILLIAMS, *Approximate representation of functions of several variables in terms of function of one variable*, Physical Review Letters, 63 (1989), pp. 927–929.
- [37] K. FUNAHASHI, *On the approximate realization of continuous mappings by neural networks*, Neural Netw., 2 (1989), pp. 183–192.
- [38] J. GARCKE, *Maschinelles Lernen durch Funktionsrekonstruktion mit verallgemeinerten dünnen Gittern*, Doktorarbeit, Institut für Numerische Simulation, Universität Bonn, 2004.

- [39] J. GARNKE, M. GRIEBEL, AND M. THESS, *Data mining with sparse grids*, Computing, 67 (2001), pp. 225–253.
- [40] T. GERSTNER AND M. GRIEBEL, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87.
- [41] F. GIROSI, M. JONES, AND T. POGGIO, *Priors stabilizers and basis functions: From regularization to radial, tensor and additive splines*, Tech. Rep. AIM-1430, 1993.
- [42] F. GIROSI, M. JONES, AND T. POGGIO, *Regularization theory and neural networks architectures*, Neural Computation, 7 (1995), pp. 219–265.
- [43] F. GIROSI AND T. POGGIO, *Representation properties of networks: Kolmogorov's theorem is irrelevant*, Neural Comp., 1 (1989), pp. 465–469.
- [44] A. N. GORBAN, *Approximation of continuous functions of several variables by an arbitrary nonlinear continuous function of one variable, linear functions, and their superpositions*, Appl. Math. Lett., 3 (1998), pp. 45–49.
- [45] M. GRIEBEL, *Sparse grids and related approximation schemes for higher dimensional problems*, in Proceedings of the conference on Foundations of Computational Mathematics (FoCM05), Santander, Spain, 2005.
- [46] P. C. HANSEN, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
- [47] ——, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998.
- [48] T. HASTIE AND R. TIBSHIRANI, *Generalized additive models (with discussion)*, Statist. Sci., 1 (1986), pp. 297–318.
- [49] ——, *Generalized Additive Models*, Monographs on Statistics and Applied Probability 43, Chapman and Hall, 1990.
- [50] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, *The Elements of Statistical Learning*, Springer, 2001.
- [51] S. HAYKIN, *Neural Networks, A Comprehensive Foundation*, Prentice Hall, 1999.
- [52] R. HECHT-NIELSEN, *Counter propagation networks*, Proceedings of the International Conference on Neural Networks II, (1987), pp. 19–32.

- [53] ——, *Kolmogorov's mapping neural network existence theorem*, Proceedings of the International Conference on Neural Networks III, (1987), pp. 11–14.
- [54] R. HECHT-NIELSEN, *Neurocomputing*, Addison-Wesley, 1990.
- [55] T. HEDBERG, *The Kolmogorov superposition theorem, Appendix II to H. S. Shapiro, Topics in Approximation Theory*, Lecture Notes in Math., 187 (1971), pp. 267–275.
- [56] F. HICKERNELL, I. SLOAN, AND G. WASILKOWSKI, *On tractability of weighted integration for certain banach spaces of functions*, Monte Carlo and Quasi-Monte Carlo Methods 2002, (2004), pp. 51–71.
- [57] F. HICKERNELL AND H. WOŹNIAKOWSKI, *Integration and approximation in arbitrary dimensions*, Adv. Comput. Math., 12 (2000), pp. 25–58.
- [58] D. HILBERT, *Mathematical problems*, Bull. Amer. Math. Soc., 8 (1902), pp. 461–462.
- [59] ——, *Über die Gleichung neunten Grades*, Math. Ann. (1927), 243–250, Gesammelte Abhandlungen, Bd. 3, 97 (1935), pp. 393–400.
- [60] K. B. HOWELL, *Principles of Fourier analysis*, Chapman & Hall /CRC, 2001.
- [61] W. HUREWICZ AND H. WALLMAN, *Dimension Theory*, Princeton University Press, Princeton, NJ, 1948.
- [62] B. IGELNIK, *Use of the Kolmogorov's superposition theorem and cubic splines for efficient neural-network modeling*, in KES, 2003, pp. 184–190.
- [63] B. IGELNIK, Y. H. PAO, S. R. LECLAIR, AND C. Y. SHEN, *The ensemble approach to neural-network learning and generalization*, IEEE Trans. on Neural Networks, 10 (1999), pp. 19–30.
- [64] B. IGELNIK AND N. PARIKH, *Kolmogorov's spline network*, IEEE transactions on Neural Networks, 14 (2003), pp. 725–733.
- [65] V. E. ISMAILOV, *On the representation by linear superpositions*, J. Approx. Theory, 151 (2008), pp. 113–125.
- [66] J. P. KAHANE, *Sur le théorème de superposition de Kolmogorov*, J. Approx. Theory, 13 (1975), pp. 229–234.
- [67] H. KATSURA AND D. A. SPRECHER, *Computational aspects of Kolmogorov's superposition theorem*, Neural Networks, 7 (1994), pp. 455–461.

- [68] S. KHAVINSON, *Best Approximation by Linear Superpositions*, vol. 159 of Translations of Mathematical Monographs, AMS, 1997.
- [69] A. N. KOLMOGOROV, *On the representation of continuous functions of many variables by superpositions of continuous functions of one variable and addition*, Doklady Akademii Nauk USSR, 14 (1957), pp. 953–956.
- [70] A. N. KOLMOGOROV AND V. M. TIKHOMIROV,  $\epsilon$ -entropy and  $\epsilon$ -capacity of sets in function spaces, Uspekhi Mat. Nauk, 13 (1959), pp. 3–86. English translation: American Math. Soc. Transl. 17, (2), 277–364, 1961.
- [71] V. KOLODYAZHNIY AND Y. BODYANSKIY, *Fuzzy Kolmogorov's network*, in KES, vol. 3214 of Lecture Notes in Computer Science, Springer, 2004, pp. 764–771.
- [72] V. KOLODYAZHNIY, Y. BODYANSKIY, AND P. OTTO, *Universal approximator employing neo-fuzzy neurons*, in Fuzzy Days, vol. 33 of Advances in Soft Computing, Springer, 2004, pp. 631–640.
- [73] M. KÖPPEN, *On the training of a Kolmogorov network*, ICANN 2002, Lecture Notes In Computer Science, 2415 (2002), pp. 474–479.
- [74] V. KREINOVICH, H. T. NGUYEN, AND D. A. SPRECHER, *Normal forms for fuzzy logic an application of Kolmogorov's theorem*, International Journal of Uncertainty Fuzziness and Knowledge-based Systems, 4 (1996), pp. 331–349.
- [75] V. KURKOVA, *Kolmogorov's theorem is relevant*, Neural Computation, 3 (1991), pp. 617–622.
- [76] ——, *Kolmogorov's theorem and multilayer neural networks*, Neural Networks, 5 (1992), pp. 501–506.
- [77] Y. KWOK, *Mathematical Models of Financial Derivatives*, Springer, 1998.
- [78] P.-E. LENI, Y. D. FOUGEROLLE, AND F. TRUCHETET, *Kolmogorov superposition theorem and its application to multivariate function decompositions and image representation*, 2008, pp. 344–351.
- [79] A. LOPEZ-GOMEZ, S. YOSHIDA, AND K. HIROTA, *Fuzzy functional link network and its application to the representation of the extended Kolmogorov theorem*, International Journal of Fuzzy Systems, 4 (2002), pp. 690–695.
- [80] G. LORENTZ, *Metric entropy, widths, and superpositions of functions*, The American Mathematical Monthly, 69 (1962), pp. 469–485.
- [81] ——, *Approximation of Functions*, Holt, Rinehart & Winston, 1966.

- [82] G. LORENTZ, M. GOLITSCHEK, AND Y. MAKOVoz, *Constructive Approximation*, Springer Verlag, 1996.
- [83] A. MESSIAH, *Quantum Mechanics*, Dover Publications, New York, 2000.
- [84] M. NAKAMURA, R. MINES, AND V. KREINOVICH, *Guaranteed intervals for Kolmogorov's theorem (and their possible relation to neural networks)*, Interval Comput., 3 (1993), pp. 183–199.
- [85] M. NEES, *Approximative versions of Kolmogorov's superposition theorem, proved constructively*, Journal of Computational and Applied Mathematics, 54 (1994), pp. 239–250.
- [86] ——, *Chebyshev approximation by discrete superposition application to neural networks*, Advances in Computational Mathematics, 5 (1996), pp. 137–151.
- [87] R. NERUDA, A. STEDRÝ, AND J. DRKOSOVÁ, *Implementation of Kolmogorov learning algorithm for feedforward neural networks*, in ICCS '01: Proceedings of the International Conference on Computational Science-Part II, London, UK, 2001, Springer-Verlag, pp. 986–995.
- [88] ——, *Kolmogorov learning for feedforward networks*, in Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference, vol. 1, 2001, pp. 77–81.
- [89] R. NERUDA, A. STEDRÝ, AND J. DRKOSOVÁ, *Variants of learning algorithm based on Kolmogorov theorem*, in ICCS '02: Proceedings of the International Conference on Computational Science-Part III, London, UK, 2002, Springer-Verlag, pp. 536–543.
- [90] H. NIEDERREITER, *Random number generation and quasi-Monte Carlo methods*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
- [91] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer Series in Operations Research, Springer, 1999.
- [92] P. A. OSTRAND, *Dimension of metric spaces and Hilbert's problem 13*, Bull. Amer. Math. Soc., 71 (1965), pp. 619–622.
- [93] V. I. PAULSEN, *An introduction to the theory of reproducing kernel Hilbert spaces*, Lecture Notes.
- [94] E. P. D. PEDNAULT, *Transform regression and the Kolmogorov superposition theorem*, in SDM, 2006.

- [95] T. POGGIO AND F. GIROSI, *A theory of networks for approximation and learning*, Laboratory, Massachusetts Institute of Technology, 1140 (1989).
- [96] ———, *Networks for approximation and learning*, in Proceedings of the IEEE, vol. 78, 1990.
- [97] H. RABITZ AND O. ALIS, *General foundations of high-dimensional model representations*, Journal of Mathematical Chemistry, 25 (1999).
- [98] C. REISINGER, *Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben*, Doktorarbeit, Universität Heidelberg, 2003.
- [99] H. RISKEN, *The Fokker-Planck Equation: Methods of Solutions and Applications*, Springer Series in Synergetics, Springer, 2 ed., 1996.
- [100] R. SCHABACK AND H. WENDLAND, *Kernel techniques: From machine learning to meshless methods*, Acta Numerica, 15 (2006), pp. 543–639.
- [101] B. SCHÖLKOPF AND A. J. SMOLA, *From regularization operators to support vector kernels*, in In Advances in Neural information processing systems 10, MIT Press, 1998, pp. 343–349.
- [102] ———, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, USA, 2001.
- [103] C. SCHWAB, *Numerical solution of parabolic problems in high dimensions*, in Proceedings of the 19th GAMM-Seminar, 2003, pp. 1–20.
- [104] I. SLOAN, *QMC integration-beating intractability by weighting the coordinate directions*, Tech. Rep. AMR01/12, University of New South Wales, Applied Mathematics report, 2001.
- [105] I. SLOAN, X. WANG, AND H. WOŹNIAKOWSKI, *Finite-order weights imply tractability of multivariate integration*, Journal of Complexity, 20 (2004), pp. 46–74.
- [106] I. SLOAN AND H. WOŹNIAKOWSKI, *When are quasi-monte carlo algorithms efficient for high dimensional integrals ?*, J. Complexity, 14 (1998), pp. 1–33.
- [107] D. A. SPRECHER, *On the structure of continuous functions of several variables*, Transactions Amer. Math. Soc, 115 (1965), pp. 340–355.
- [108] ———, *On the structure of representations of continuous functions of several variables as finite sums of continuous functions of one variable*, Proceedings of the American Mathematical Society, 17 (1966), pp. 98–105.

- [109] ——, *An improvement in the superposition theorem of Kolmogorov*, Journal of Mathematical Analysis and Applications, 38 (1972), pp. 208–213.
- [110] ——, *A universal mapping for Kolmogorov's superposition theorem*, Neural Networks, 6 (1993), pp. 1089–1094.
- [111] ——, *A numerical implementation of Kolmogorov's superpositions*, Neural Networks, 9 (1996), pp. 765–772.
- [112] ——, *A numerical implementation of Kolmogorov's superpositions II*, Neural Networks, 10 (1997), pp. 447–457.
- [113] D. A. SPRECHER AND S. DRAGHICI, *Space filling curves and Kolmogorov superposition-based neural networks*, Neural Networks, 15 (2002), pp. 57–67.
- [114] Y. STERNFELD, *Uniformly separating families of functions*, Israel Journal of Mathematics, 29 (1978), pp. 61–91.
- [115] ——, *Superpositions of continuous functions*, Journal of Approximation Theory, 25 (1979), pp. 360–368.
- [116] ——, *Dimension, superposition of functions and separation of points, in compact metric spaces*, Israel Journal of Mathematics, 50 (1985).
- [117] C. STONE, *Additive regression and other nonparametric models*, Ann. Statist., 13 (1985), pp. 689–705.
- [118] A. N. TIKHONOV, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Dokl., 4 (1963), pp. 1035–1038.
- [119] A. N. TIKHONOV AND V. A. ARSENIN, *Solutions of ill-posed problems*, Winston, Washington D.C., 1977.
- [120] A. N. TIKHONOV, A. V. GONCHARSKY, V. V. STEPANOV, AND A. G. YAGOLA, *Numerical Methods for the solution of ill-posed problems*, Kluwer Academic Publishers, 1995.
- [121] S. TRÖMEL, C. SIMMER, J. BRAUN, T. GERSTNER, AND M. GRIEBEL, *Towards the use of integral radar volume descriptors for quantitative areal precipitation estimation - results from pseudo-radar observations.*, Journal of Atmospheric and Oceanic Technology, 26 (2009), pp. 1798–1813.
- [122] V. N. VAPNIK, *Estimation of Dependences Based on Empirical Data*, Springer Verlag, 1982.
- [123] ——, *The Nature of Statistical Learning Theory*, Springer Verlag, 1995.

- [124] ——, *Statistical Learning Theory*, Springer Verlag, 1999.
- [125] A. VITUSHKIN, *Proof of the existence of analytic functions of several variables not representable by linear superpositions of continuously differentiable functions of fewer variables*, Dokl. Akad. Nauk SSSR, 156 (1964), pp. 1258–1261. English translation: Soviet Math. Dokl. (5), 793–796, 1964.
- [126] A. VITUSHKIN, *On Hilbert’s thirteenth problem and related questions*, Russian Math. Surveys, 51 (2004), pp. 11–25.
- [127] G. WAHBA, *Spline Models for Observational Data*, vol. 59, Series in Applied Mathematics, 1990.
- [128] X. WANG AND I. SLOAN, *Why are high-dimensional finance problems often of low effective dimension?*, SIAM J. Sci. Comput., 27 (2005), pp. 159–183.
- [129] G. WASILKOWSKI AND H. WOŹNIAKOWSKI, *Explicit cost bounds of algorithms for multivariate tensor product problems*, J. Complexity, 11 (1995), pp. 1–56.
- [130] ——, *Weighted tensor product algorithms for linear multivariate problems*, J. Complexity, 15 (1999), pp. 402–447.
- [131] G. WASILKOWSKI AND H. WOŹNIAKOWSKI, *Finite-order weights imply tractability of linear multivariate problems*, Journal of Approximation Theory, 130 (2004), pp. 57–77.
- [132] H. WENDLAND, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2005.
- [133] P. WOJTASZCYK, *Banach spaces for analysts*, vol. 25 of Cambridge Stud. Adv. Math., Cambridge University Press, 1991.
- [134] Y. YAM, H. T. NGUYEN, AND V. KREINOVICH, *Multi-resolution techniques in the rules-based intelligent control systems: A universal approximation result*, in Proceedings of the 14th IEEE International Symposium on Intelligent Control/Intelligent Systems and Semiotics ISIC/ISAS’99, 1999, pp. 213–218.
- [135] R.-X. YUE AND F.-J. HICKERNELL, *Designs for smoothing spline ANOVA models*, Metrika. International Journal for Theoretical and Applied Statistics, 55 (2002), pp. 161–176.
- [136] Z.-J. ZENG AND A. KEANE, *Approximation capabilities of hierarchical fuzzy systems*, IEEE Transactions on Fuzzy Systems, 13 (2005).
- [137] *GNU Multiple Precision arithmetic library*. <http://gmplib.org/>.

- [138] *GNU Scientific Library*. <http://www.gnu.org/software/gsl/>.
- [139] *Portable, Extensible Toolkit for Scientific computation*. <http://www.mcs.anl.gov/petsc/petsc-as/>.