Felix Bartel

Least Squares in Sampling Complexity and Statistical Learning

Felix Bartel

Least Squares in Sampling Complexity and Statistical Learning

TECHNISCHE UNIVERSITÄT
CHEMNITZ

Universitätsverlag Chemnitz
2024

Impressum

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über `https://www.dnb.de` abrufbar.

Die vorliegende Arbeit wurde von der Fakultät für Mathematik der Technischen Universität Chemnitz als Dissertation zur Erlangung des akademischen Grades Dr. rer. nat. genehmigt.

Gutachter:
Prof. Dr. Daniel Potts
Prof. Dr. Sergei Pereverzyev
Prof. Dr. Dirk Nuyens

Tag der Einreichung: 12.10.2023
Tag der Verteidigung: 07.12.2023

Universitätsverlag Chemnitz
`https://www.tu-chemnitz.de/ub/univerlag`

Print on Demand
Gedruckt auf FSC-zertifiziertem Papier

ISBN 978-3-96100-204-7
DOI: 10.51382/978-3-96100-204-7
URN: urn:nbn:de:bsz:ch1-qucosa2-887943

Titelgrafik: Felix Bartel

Contents

1 Introduction 7

2 Least squares approximation 19

2.1 History 20

2.2 The method 21

2.3 Implementation 23

3 Reproducing kernel Hilbert spaces (RKHS) 25

3.1 Basic concepts of RKHSs 26

3.2 Constructing RKHSs from a given Mercer representation . . 29

3.3 Interlude on spectral theory of compact operators 33

3.4 Weak Mercer representation for RKHSs compactly embedded into L_2 36

3.5 Examples in one dimension 39

3.5.1 Sobolev spaces on the torus $\mathbb{T}$ 40

3.5.2 Sobolev spaces on the interval $[0, 1]$ 42

3.5.3 Jacobi polynomials on the interval $[-1, 1]$ 57

3.6 Extension to higher dimensions 61

3.6.1 Isotropic Sobolev spaces on the torus $\mathbb{T}^d$ 61

3.6.2 Sobolev spaces with dominating mixed smoothness . 65

3.6.3 Truncated ANOVA decomposition 66

4 Concentration inequalities 69

4.1 Concentration inequalities for random vectors 69

4.2 Concentration inequalities for the spectral norm of (infinite) matrices 76

5 Subsampling of finite frames 79

5.1 Random weighted subsampling of finite frames 80

5.2 Deterministic weighted subsampling of finite frames 82

5.3 Non-weighted subsampling of finite frames 90

5.4 Numerical experiments 96

6 L_2-Marcinkiewicz-Zygmund (MZ) inequalities 105

6.1 Random L_2-MZ inequalities 108

6.2 Examples of deterministic L_2-MZ inequalities 110

6.3 Subsampling of L_2-MZ inequalities 114

7 Least squares in the worst-case setting **119**
7.1 Sampling recovery in spaces of finite measure 120
7.2 Sampling recovery in RKHSs 122
7.2.1 Optimal sampling complexity with unstructured points 122
7.2.2 Sampling recovery with L_2-MZ points 128
7.2.3 Sampling recovery with subsampled L_2-MZ points . 131
7.3 Application on the torus $\mathbb{T}^d$ with rank-1 lattices 137
7.4 Comparison to the kernel method 143

8 Least squares in statistical learning **147**
8.1 Basic concepts in statistical learning 148
8.2 Error guarantees for least squares approximation 150
8.3 Application on the unit interval $[0, 1]$ 157
8.3.1 Analysis 157
8.3.2 Numerical Experiments 159
8.4 Numerical experiment on the unit cube $[0, 1]^d$ 162
8.5 Comparison to other methods 164

9 Cross-validation **167**
9.1 Concentration of the cross-validation score 168
9.2 Fast cross-validation for least squares approximation 178
9.3 Numerical experiments 184
9.3.1 Numerical experiment on the torus $\mathbb{T}$ 184
9.3.2 Numerical experiment on the unit cube $[0, 1]^d$ 186

10 Outlook **189**

List of Symbols **191**

Bibliography **193**

Chapter 1

Introduction

Data gathering is a constant in human history with examples like the use of tally marks to keep track of live stock, the documentation of the position of celestial bodies, or equipping everyday items with sensors in an internet-of-things fashion. Mathematically speaking, the recording of such observations resemble discrete measurements of a function. With the evergrowing supply of data, the tabular form is overwhelming for a human to interpret in its raw form. The field of data science enters here to simplify given data in order to answer specific questions about their analysis. At the core lies the conversion of discrete measurements into a function. This can be as simple as connecting dots in a graph, but as soon as the scenario is a little more advanced, tools from approximation theory and numerical analysis should be deployed.

In this thesis we have an in-depth look into one approach to tackle this function approximation problem: the least squares approximation method, which dates back to around 1800. It approximates functions based on discrete and possibly noisy function evaluations. Despite its age, it is present in every numerical toolbox, implemented in embedded systems as well as high-level computers, and is used countless times on a daily basis. The popularity of the least squares approximation method gave rise to lots of interesting questions with many being the goal of current research. More than 200 years after the discovery of the algorithm there are still over 250 new publications about least squares approximation listed in the online bibliographic database MathSciNet on a yearly basis.

Our goal is to give a short introduction to this topic. This allows us to present the current state of the theory, where we contribute new results in several different areas aiming to advance the theoretical validation of least squares approximation being an all around viable method which can be tailored to a broad range of applications. We confirm our theoretical findings with numerical experiments which are reproducible and publicly available at `https://github.com/felixbartel/dissertation`. Next, we present an outline in a chapter-based form as we touch distinctive areas of research.

Outline of the thesis

Chapter 2: Least squares approximation. At the core of this thesis is the least squares approximation method. For points $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$, values $\boldsymbol{y} = (y_1, \dots, y_n)^\mathsf{T}$, and a function space V it is defined by

$$S_V^{\boldsymbol{X}} \boldsymbol{y} := \arg\min_{g \in V} \sum_{i=1}^{n} |y_i - g(\boldsymbol{x}^i)|^2 \,.$$

Looking into the history of least squares approximation, there is a fiery exchange of letters about the origins of the method itself, which we will briefly present. We will touch upon some predecessors of least squares approximation, address its prevalence over the years, and its impact today. Based on an introductory example about polynomial approximation to get the idea, we discuss how to elevate the concept to match today's requirements and present the method as we will use it throughout the thesis. Further, we comment on some details of the implementation and computational complexity of the underlying algorithm.

Chapter 3: Reproducing kernel Hilbert spaces (RKHS). As a general function may differ arbitrarily in every single point and in our problem setting we start from a discrete set of function evaluations, which is either chosen as in Chapter 7 or given as in Chapter 8, the problem of approximation from samples cannot be approached without some sort of restriction. We model the prior assumptions by requiring that the target function belongs to a class of functions describing the smoothness thereof. In particular, we will shortly introduce the theory of reproducing kernel Hilbert spaces (RKHSs), which supply a decisive set of tools under minimal assumptions in order to define a concept of smoothness. As the name suggests, every RKHS is coupled with a kernel function. We first construct the RKHS from a given series representation of the kernel, more specifically from a Mercer representation. As examples of this sort may be academical, we gather tools from the spectral theory of compact operators to obtain the reverse direction. With that we show that RKHSs, which are compactly embedded into the space of square-integrable functions L_2, have a kernel with a weak Mercer representation. This will, in turn, yield a basis for all common RKHSs which simultaneously is a basis of L_2 up to normalization and usable in least squares approximation.

On this note, we will give examples which will return in the latter experiments. Starting with the one-dimensional setting, we will introduce unweighted Sobolev spaces on the torus $\mathbb{T}$ and unit interval $[0,1]$ motivated by the solution of an ordinary differential equation. We pose a basis for the non-periodic Sobolev space of smoothness two $H^2([0,1])$ in Theorem 3.27. In theory, this basis was known before but not used in practice due to its numerical instability, cf. [AIN12, Section 3]. We propose an approximation consisting of cosine and exponential terms with negative argument, which is numerically stable. We prove its accuracy in Theorem 3.28 and put it to test in various numerical experiments, cf. Sections 8.3.2, 8.4, and 9.3.2. This is a practical extension of the popular half-period cosine basis, cf. [WW09, IN08, Adc10b, AH11, DNP14, SNC16, CKNS16, KMNN21], and is useful for the approximation of functions with traditional Sobolev smoothness $s \geq 2$ yielding better accuracy for smaller polynomial degree. Further, we will introduce weighted Sobolev spaces with polynomial basis on the unit interval. We elevate these spaces to higher dimensions presenting isotropic Sobolev spaces, Sobolev spaces with dominating mixed smoothness, as well as the concept of analysis of variance (ANOVA). Each being more restrictive and more applicable in high-dimensional approximation.

Chapter 4: Concentration inequalities. Since randomness gained popularity by significantly pushing forward results in compressive sensing, the techniques spread over to other research areas with numerical analysis being no exception. It often seems to be an indispensable tool to show improved bounds by abandoning the deterministic component and certainty of results. Where results stated in expectation give a first idea to what is possible, in the current state of research one is interested in bounds holding with high probability. Even though there exist realizations of the involved random variables such that the statement would not hold, knowing that it fails, e.g. only in one out of $1\,000$ cases is often sufficient and can be fine tuned on demand. Bounds of this from are known as concentration results with the famous Markov inequality being a basic example. There, the fail probability or tail decreases linearly whereas an exponentially decaying tail is preferable. In this chapter we present a small collection of concentration inequalities of this sort, which are key tools in the proofs of the later approximation results. Next to Bernstein's inequality, we prove a complex version of the Hanson-Wright inequality for quadratic forms Theorem 4.2, and pose McDiarmid's inequality for general vector valued functions. Further, we present bounds for the spectral norm of possibly infinite-dimensional matrices.

Chapter 5: Subsampling of finite frames. The goal in function approximation is to achieve a small error in some norm defined on the whole domain, i.e., a quantity where every single point matters. The discrete set of points, in which we have given function evaluations for the sampling problem, is a subset thereof. Therefore, the task can be understood as a subsampling procedure while trying to keep the inherent information. Abstracting this concept to basic linear algebra, we formulate the subsampling question in terms of finite frames, i.e., a set of vectors $\boldsymbol{y}^1, \dots, \boldsymbol{y}^M \in \mathbb{C}^m$ such that for $0 < A \le B < \infty$ it holds

$$A\|\boldsymbol{a}\|^2 \le \sum_{i=1}^{M} |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \le B\|\boldsymbol{a}\|^2 \quad \text{for all} \quad \boldsymbol{a} \in \mathbb{C}^m \,,$$

where M is the frame size and m the frame dimension. Here, we are interested in optimal subframes $(\boldsymbol{y}^i)_{i\in J}$ in the sense of having only a small number $|J|$ of frame elements in comparison to the dimension m of the frame, where $m \le |J|$ is a natural limit. With the random tools form Chapter 4 we show in Theorem 5.1 that a random weighted subframe of size $|J| \sim m \log m$, drawn with respect to a special distribution, keeps the condition of the frame intact. The remaining logarithmic gap can be closed using deterministic subsampling techniques related to the recent solution of the famous Kadison-Singer problem to achieve $|J| \sim m$. However, this is a pure existence result and not constructive, cf. [MSS15]. We present the `BSS`-algorithm (named after J. D. Batson, D. A. Spielman, and N. Srivastava in [BSS09]), a constructive approach with polynomial runtime bringing the frame size $|J|$ arbitrarily close to the frame dimension m in Theorem 5.3.

In order to apply these techniques to the sampling problem the weights coming from the subsampling present a hurdle. We modify the density of the random subsampling such that we loose the weights in Theorem 5.7 with logarithmic oversampling. Further, by constructing an extension of the frame in Theorem 5.9, we apply the `BSS`-algorithms in such a way, that we obtain an unweighted subset of the original frame with linear oversampling whilst saving the lower frame bound. In particular, for a target oversampling factor $b > 1 + \frac{1}{m}$, we construct a set of indices $J \subseteq \{1, \dots, M\}$ in Theorem 5.8 such that $|J| \le \lceil bm \rceil$ and

$$\frac{1}{M} \sum_{i=1}^{M} |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \le 89 \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^3} \frac{1}{m} \sum_{i\in J} |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \quad \text{for all} \quad \boldsymbol{a} \in \mathbb{C}^m \,.$$

Consequently, if the original vectors have a lower frame bound, the subset

$(y_i)_{i\in J}$ does as well. We round this section up by testing the proposed algorithms with numerical examples.

Chapter 6: L_2-Marcinkiewicz-Zygmund (MZ) inequalities. In this section we discuss the equivalence of continuous L_2-norms with discrete norms based on a finite number of function evaluations. This equivalence, known as L_2-Marcinkiewicz-Zygmund (MZ) inequalities, can only hold on a finite-dimensional function space where it is a crucial tool to show L_2-error bounds when working with discrete samples. A set of points $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$ and weights $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n)$ fulfill an L_2-MZ inequality with constants $0 < A \le B < \infty$ for a finite-dimensional function space V, if

$$A\|f\|_{L_2}^2 \le \sum_{i=1}^{n} \omega_i |f(\boldsymbol{x}^i)|^2 \le B\|f\|_{L_2}^2 \quad \text{for all} \quad f \in V\,.$$

In many specific settings, there are deterministic examples of L_2-MZ inequalities with an inherent structure, which makes fast algorithms for computations with these points applicable, as we will see in Section 6.2. Exact quadrature points also fall into this category, cf. Theorem 6.3. Their existence was thouroughly studied, see e.g. [Tem18, KKLT22]. We will present a generic way in Theorem 6.4 to construct points fulfilling these connections based on a universal random construction with tools from Chapter 4 such that the number of required points n is logarithmicly more than the dimension of the underlying function space $\dim(V)$. This gives a vast zoo of available starting points to use in least squares approximation.

We emphasize on the equivalence between frames, L_2-MZ inequalities, and the condition of the least squares matrix in Theorem 6.2. Switching to the frame characterization, we are able to apply the subsampling techniques from Chapter 5 in order to reduce the number of points to merely linear oversampling $n \sim \dim(V)$ whilst keeping the approximation properties. In the end, we use these points in least squares approximation, where the good conditioning, resembled in the condition of the least squares matrix, cf. Theorem 6.1, yields good error bounds and limits the necessary number of iterations for a good solution, cf. Theorem 2.3.

Chapter 7: Least squares in the worst-case setting. In this chapter we have all tools at hand in order to approach the main task of proving error bounds for function approximation. We investigate the worst-case setting, namely, we seek a set of points which is usable to approximate not only one function but all elements of a class of functions. For a function class F, this is quantified by the **(linear) sampling width**

$$g_n(F, L_2(D, \varrho_T)) \coloneqq \inf_{\substack{\boldsymbol{x}^1,\dots,\boldsymbol{x}^{n-1} \in D \\ \varphi_1,\dots,\varphi_{n-1} \in L_2}} \sup_{\|f\|_F \le 1} \Big\| f - \sum_{i=1}^{n-1} f(\boldsymbol{x}^i)\varphi_i \Big\|_{L_2},$$

which model the best set of points and linear algorithm for approximating every function $f \in F$. By loosing the restriction to function evaluations and allowing measurements from arbitrary linear functionals, we obtain a benchmark called **linear width**

$$a_m(F, L_2(D, \varrho_T)) \coloneqq \inf_{\substack{\ell_1,\dots,\ell_{m-1}\colon F \to \mathbb{C} \\ \varphi_1,\dots,\varphi_{m-1} \in L_2}} \sup_{\|f\|_F \le 1} \Big\| f - \sum_{k=1}^{m-1} \ell_k(f)\varphi_i \Big\|_{L_2}.$$

Showing error bounds in terms of the linear width for least squares approximation yields an upper bound on the sampling width which guarantees the sharpness thereof. In the RKHS setting we exploit freedom in the way of subsampling L_2-MZ inequalities to show a technique to obtain sharp error bounds up to a logarithm. Our approach uses the polynomial runtime `BSS`-algorithm from Chapter 5. We prove the following error bound in the worst-case setting for RKHSs $H(K)$ in Theorem 7.8:

$$g^2_{\lceil bm \rceil}(H(K), L_2) \le C \frac{(b+1)^2}{(b-1)^3} \frac{\log m}{m} \sum_{k=m}^{\infty} a^2_m(H(K), L_2).$$

These results are optimal up to a single logarithm, which improves on earlier results in [MU21, Tem21, LT22] by being constructive and implementable. A year prior of the publication of this thesis, tight error bounds were proven, showing that function evaluations are as powerful as measurements from arbitrary linear functionals for separable RKHS, i.e., it holds $g_{cn}(H(K), L_2)^2 \le \frac{1}{n}\sum_{k\ge n} a_k(H(K), L_2)^2$, cf. [DKU23]. However, this relies on highly nonconstructive subsampling techniques using the Kadison-Singer Theorem.

Further, we will give basic error bounds in Theorem 7.10 for function approximation with points from general L_2-MZ inequalities which are worse in general but the inherent structure of the point sets may be combinable with fast algorithms. We propose a subsampling of L_2-MZ inequalities which exploits some freedom in the way of subsampling to focus on good approximation properties in the worst-case setting for RKHS. In that way, we obtain the near-optimal approximation rates as above while keeping the inherent structure from the initial L_2-MZ inequality. This is done with a random approach with logarithmic oversampling in Theorem 7.13 and with a two-step procedure by further applying the `BSS`-algorithm in Theorem 7.15 to have merely linear oversampling.

We apply this general theory for the well-understood rank-1 lattices in Theorem 7.16, where the subsampling makes it possible to regain the good approximation properties whilst making it possible to utilize fast Fourier algorithms for a fast implementation of the involved matrix-vector product. We confirm its applicability by conducting numerical experiments.

Leaving the RKHS setting we prove error bounds for least squares approximation in the worst-case setting for arbitrary (not necessarily Hilbert) function spaces with finite measure in Theorem 7.3. Here we use the **Kolmogorov width** in $\ell_\infty(D)$

$$d_m(F, \ell_\infty(D)) := \inf_{\substack{V \subseteq F \\ \dim(V) = m-1}} \sup_{\|f\|_F \le 1} \inf_{g \in V} \Big\| f - g \Big\|_\infty ,$$

which includes non-linear approximations in contrast to the linear width. Note, that these coincide when the error is measured in a Hilbert spaces norm. With the optimal L_2-MZ inequalities available from Chapter 6, this yields a relation between the sampling width and the Kolmogorov width in the space of bounded functions $\ell_\infty(D)$ in Theorem 7.4:

$$g_{\lceil bm \rceil}(F, L_2) \le C \Big(\frac{b+1}{b-1} \Big)^{3/2} d_m(F, \ell_\infty(D)) .$$

This improves on earlier results in [Tem21, LT22] by loosing restrictions on the involved spaces and lowering the smallest oversampling factor b to be arbitrarily close to one.

Chapter 8: Least squares in statistical learning. With statistical learning we enter a framework in machine learning. Instead of having a look at worst-case errors as in Chapter 7, we investigate individual function approximation. The model is based on randomness in the points and the function evaluations, which includes the scenario of noise and makes the probabilistic tools from Chapter 4 vital. As a benchmark we use the error of the projection from the ansatz space of the least squares approximation. After introducing the necessary notion, we show in Theorem 8.3 that the least squares approximation has the same error as the projection up to a multiplicative constant provided noise-free samples and logarithmic oversampling. Including noise, our bounds in Theorems 8.4 and 8.5 resemble the classical over- and underfitting behavior one has to balance to achieve the smallest error. More precisely, in the regime of logarithmic oversampling $m \lesssim n \log n$ with $m = \dim(V)$ the number of ansatz functions and $n = |\boldsymbol{X}|$ the number of points, we bound the error by the best possible approximation in V plus a term linearly growing in the size of the ansatz space $m = \dim(V)$ due to noise, i.e.,

$$\|f_q - S_V^{\boldsymbol{X}} \boldsymbol{y}\|_{L_2}^2 \le C_1 \|f_q - P_V f_q\|_{L_2}^2 + C_2 \frac{m}{n} \,,$$

where f_q is the regression function and $P_V f_q$ the L_2-projection onto V. This improves on earlier result by being statements holding with high probability, which were previously only known in expectation or with a coarser bound, cf. [Bar02, CDL13, MNvST14, CM17, KUV21, HNP22, LPU23].

Further, we consider the covariate-shift setting, namely, we consider a source measure ϱ_S, with respect to which we draw the samples, and a target measure ϱ_T for measuring the error to be different. The relevance of this setting is given by, e.g. algorithms learning from artificial data with the goal of performing good in a real world scenario. We investigate different effects by comparing different combinations of measures and ansatz functions on the unit interval. The analytical results are confirmed by numerical experiments. Finally, we conduct experiments on the five-dimensional cube $[0,1]^5$ to show the applicability of our theory.

Chapter 9: Cross-validation. In Chapter 8 we observe under- and overfitting behavior which is common in function approximation from noisy samples and rises the question of parameter choice naturally. In least squares approximation the reason lies in the choice of which and how many ansatz functions are used. To know the best ansatz functions with certainty, we would need to compute the error of the approximation itself. But doing this requires the knowledge of the target function making the problem obsolete. Instead, we have to work with what we are given, namely, the samples at hand. We have a look at such a purely data-driven method named leave-one-out cross-validation. More precisely, we introduce the importance weighted cross-validation score $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}} \boldsymbol{y})$ in Theorem 9.1, which is suitable also for the covariate-shift setting. Here, one computes the approximation based on the data with single samples omitted. Doing this for every single sample, it constructs an estimator for the error itself. It is widely used in practice but has two fundamental challenges: (1) the naive way to implement the cross-validation score leads to a very expensive algorithm as one has to set up as many approximations as there are given points and (2) conceptually, the cross-validation score makes sense but theoretical validation is little.

To address point (1), in Theorems 9.13 and 9.14 we extended a result for the fast computation in [TW96] from the torus and the point grid to arbitrary domains with an exact L_2-MZ inequality. This allows to reduce the cost of computing the cross-validation score to the cost of computing the least squares approximation itself. Since these fast computation formulas only hold for exact L_2-MZ inequalities, we introduced the approximated cross-validation score in Theorem 9.16. This allows to use the same fast algorithms and we show bounds on the involved error in Theorem 9.20.

To address point (2), in Theorem 9.2, we combine techniques from [BE02] and [BH22] in order to show that the cross-validation score is estimating the error with respect to the target distribution when the weights are chosen from samples of the Radon-Nikodym derivative $\beta = \frac{\mathrm{d}\varrho_T}{\mathrm{d}\varrho_S}$, i.e.,

$$\mathbb{E}_{\boldsymbol{X},\boldsymbol{y}}\Big(\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})\Big) = \mathbb{E}_{\boldsymbol{X}_{-1},\boldsymbol{y}_{-1}}\Big(\|S_V^{\boldsymbol{X}_{-1}}\boldsymbol{y}_{-1} - f_q\|^2_{L_2(D,\varrho_T)}\Big) + \sigma_q^2\,,$$

where X_{-1} and $\boldsymbol{y}_{-1}$ denotes the omission of the first sample. To not only know what the importance weighted cross-validation score is estimating but to rather have a qualitative statement about how good it can estimate the L_2-error, we show the concentration of cross-validation around its expectation in Theorem 9.10. For that we extended a robustness concept of approximation algorithms from [BE02] and combine it with an extension of the McDiarmid concentration inequality to show a guarantee for importance weighted cross-validation with least squares approximation. This yields a guarantee to use cross-validation as an a posteriori parameter choice strategy in Theorem 9.11:

$$\|f_q - T_K S_V^{\boldsymbol{X}_{-1}}\|^2_{L_2(D,\varrho_T)} \le \mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma_q^2 + C\frac{N^{3/2}(V)}{\sqrt{n}}\,.$$

We conclude this chapter by confirming our theory by conducting a simple experiment on the one-dimensional torus $\mathbb{T}$ as well as resuming to the numerical experiment on the five-dimensional unit cube $[0,1]^5$ from Chapter 8 and will see, that the proposed fast cross-validation score is viable in the numerical experiments as the theory suggests.

List of publications

Parts of the content of this thesis is already published in the following articles:

- [BHP20]: F. Bartel, R. Hielscher, and D. Potts. Fast Cross-validation in Harmonic Approximation. *Appl. Comput. Harmon. Anal., 49(2):415-437*, 2020.

- [BPS22]: F. Bartel, D. Potts, and M. Schmischke. Grouped Transformations in High-Dimensional Explainable ANOVA Approximation. *SIAM J. Sci. Comput., 44(3):A1606-A1631*, 2022.

- [BH22]: F. Bartel and R. Hielscher. Concentration inequalities for cross-validation in scattered data approximation. *J. Approx. Theory, 277:Paper No. 105715, 17*, 2022.

- [BSU23]: F. Bartel, M. Schäfer, and T. Ullrich. Constructive subsampling of finite frames with applications in optimal function recovery. *Appl. Comput. Harmon. Anal., 65:209-248*, 2022.

- [BKPU23]: F. Bartel, L. Kämmerer, D. Potts, and T. Ullrich. On the reconstruction of functions from values at subsampled quadrature points. *Math. Comp., published electronically*, 2023.

- [Bar23]: F. Bartel. Stability and error guarantees for least squares approximation with noisy samples. *SMAI J. Comput. Math., 9:95-120*, 2023.

- [BT23]: F. Bartel and F. Taubert. Nonlinear Approximation with subsampled Rank-1 Lattices. *Fourteenth International Conference on Sampling Theory and Applications*, 2023.

- [NBUL23]: F. Bartel, K. Lüttgen, N. Nagel, and T. Ullrich. Efficient recovery of non-periodic multivariate functions from few scattered samples. *Fourteenth International Conference on Sampling Theory and Applications*, 2023.

Acknowledgments

In the process of creating this thesis I enjoyed the help of many people. First and foremost, I sincerely thank my outstanding supervisor Prof. Daniel Potts for countless mathematical discussions which always sparked enthusiasm and promoted our mathematical advances. He also guided and ensured for the means to attend plenty of conferences, give talks in seminars, and even two research stays in the course of being a PhD student. A huge contribution to that have also Prof. Ralf Hielscher and Prof. Tino Ullrich. I appreciate the countles hours we spend in front of blackboards or elsewhere from which stem the papers being the basis of this thesis. On that note, I also want to thank Prof. Otmar Scherzer, Prof. Peter Binev, and Prof. Vladimir Temlyakov hosting me for my research stays at the University of Vienna and the University of South Carolina totaling twelve months.

Furthermore, over the course of my studies, I had the honor to collaborate with my wonderful colleagues Prof. Ralf Hielscher, Dr. Lutz Kämmerer, Kai Lüttgen, Nicolas Nagel, Prof. Daniel Potts, Dr. Martin Schäfer, Dr. Michael Schmischke, Fabian Taubert, and Prof. Tino Ullrich. All other colleagues from the analysis floor also helped covering many little questions and created an excellent working climate for many on- and off-topic discussions. In particular, my office neighbours Dr. Michael Quellmalz and Kseniya Akhalaya had to endure most of them.

Next, I want to thank the eager readers of my papers giving hints in order to improve them by showing possible placements of my theorems with examples being Prof. Sergei Prereverzyev and Dr. Werner Zellinger introducing the covariate shift setting to me and giving references to general regularization theory playing an important role in Chapters 8 and 9, Prof. Albert Cohen proposing a vast list of references for individual function approximation for a good read and proper placement of my results, or Dr. Matthieu Dolbeault for simplifying an aspect of the extension used for subsampling in Chapter 5. Here, I also appreciate Prof. Dirk Nuyens and Prof. Sergei Pereverzyev for agreeing to review this dissertation giving insightful comments and suggestions.

Over the course of my studies I stayed in many places and many flatmates and friends who had to tolerate my mathematical fascination on some days and slight frustration on others. I appreciate them very much rooting for me regardless of their mathematical background. Last but not least, a huge thanks to my family putting me in the right tracks making all of this possible in the first place and my girlfriend Patricia enduring me in particular on the final stretch.

Chapter 2

Least squares approximation

Least squares is a method for solving overdetermined systems of equations, i.e., problems with more conditions than degrees of freedom. We use it by its original purpose, namely function approximation. We begin with an example. Imagine we have given data $(x_i, y_i) \in [0,1] \times \mathbb{R}$ for $i = 1, \ldots, n$ and search for a quadratic polynomial approximating the data

$$g(x) = ax^2 + bx + c$$

which is parameterized by $a, b, c \in \mathbb{R}$. We depicted the setting in Figure 2.1.

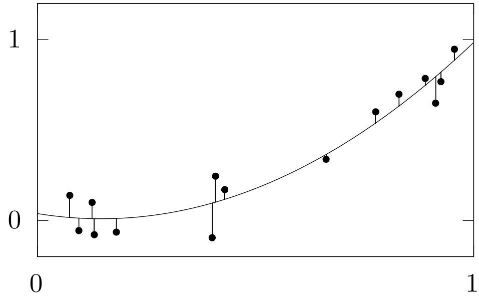


Figure 2.1: Toy example fitting a quadratic polynomial using least squares approximation.

The idea of least squares approximation is to find a, b, and c such that the corresponding the polynomial minimizes the sum of squared residuals

$$\sum_{i=1}^{n} |g(x_i) - y_i|^2 .$$

Because the considered residuals $g(x_i) - y_i$ are linear in the unknowns a, b, c the method is then called linear least squares approximation.

In this chapter we give some historic remarks on the discovery of the least squares approximation method, introduce it in the required generality, and give some insights into the implementation and runtime of the method.

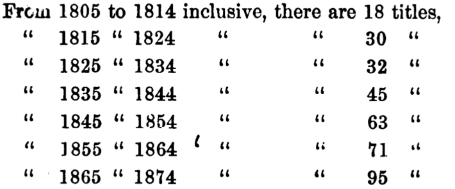
Figure 2.2: **Left:** Only known depiction of A. Legendre (1752-1833) (image source: https://commons.wikimedia.org/wiki/File:Legendre.jpg).
Middle: portrait of C. F. Gauss (1777-1855) (image source: https://commons.wikimedia.org/wiki/File:Carl_Friedrich_Gauss_1840_by_Jensen.jpg).
Right: portrait of R. Adrain (1775-1843) (image source: https://commons.wikimedia.org/wiki/File:Robert_Adrain,_1775_-_1843.jpg).

2.1 History

The least squares approximation method is more than 200 years old with predecessors ranging even 300 years back. The first official occurrence of the method is a publication by A. Legendre where the method was used for analyzing data concerning the shape of the earth, cf. [Leg05]. The same data was used by R. J. Boscovich 1757 and P. Laplace in 1799 with the least absolute deviations method, where one minimizes the absolute value of residuals instead of their squares. Independently, 1808 R. Adrain discovered the least squares approximation method, cf. [Adr08]. Raising some tension, 1808 C. F. Gauss published a book claiming to have discovered the least squares approximation method as early as 1795 in his mathematical diary for calculating the orbits of celestial bodies, cf. [Gau11]. This resulted in an exchange of letters with the intend to determine the righteous inventor of the least squares approximation method, where a collection can be found in [Pla72, Sti81].

Today, it is agreed that A. Legendre officially discovered and published the method at first and C. F. Gauss is co-credit as he contributed significant theoretical advances. He showed, when the observations come from an exponential family, the least squares approximation and maximum likelihood estimator coincide, cf. Theorem 2.2.

A detailed list of 408 related papers collected by M. Merriman, dating as early as 1722 until 1876, can be found in [Mer77]. Note, that this is not a complete list as the author did not include literature in the Russian or Hungarian language. There, 22 titles dated earlier than the official discovery in 1805 which can be viewed as predecessors and 354 afterwards, which are counted in 10 year intervals in the table which we copied from [Mer77].



From	1805	to	1814	inclusive,	there are	18	titles,
"	1815	"	1824	"	"	30	"
"	1825	"	1834	"	"	32	"
"	1835	"	1844	"	"	45	"
"	1845	"	1854	"	"	63	"
"	1855	"	1864	"	"	71	"
"	1865	"	1874	"	"	95	"

This emphasizes the rapidly growing significance of the least squares approximation method. As of today, least squares approximation is taught in every mathematical course, implemented in every numerical linear algebra software package, and is the standard tool in data fitting with applications in all places.

2.2 The method

Now we formulate the least squares approximation method in a general context as in [Bjö96, Chapter 8]. For a domain $D \neq \emptyset$ and $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, we consider $n \in \mathbb{N}$ data tuples $(\boldsymbol{x}^i, y_i) \in D \times \mathbb{K}$, $i = 1, \ldots, n$, for which we seek an approximation $g\colon D \to \mathbb{K}$. For an $m - 1 \leq n$ dimensional function space $V = \operatorname{span}\{\eta_1, \ldots, \eta_{m-1}\} \subseteq \mathbb{K}^D$, we make the ansatz as a linear combination

$$g(x) = \sum_{k=1}^{m-1} \hat{g}_k \eta_k(x)\,,$$

where we have to determine the coefficients $\hat{g}_k \in \mathbb{K}$. In general we cannot guarantee interpolation $g(\boldsymbol{x}^i) = y_i$ in every point. Instead, the least squares approximation method minimizes the weighted sum of squared residuals

$$\sum_{i=1}^{n} \omega_i |g(\boldsymbol{x}^i) - y_i|^2 \to \min\,,$$

with $\omega_1, \ldots, \omega_n \geq 0$. The result is then the **weighted least squares approximation** $S_V^{\boldsymbol{X}} \boldsymbol{y} = S_V^{\boldsymbol{X}}(\omega_1, \ldots, \omega_n)\boldsymbol{y}\colon \mathbb{K}^n \to \mathbb{K}^D$, where $\boldsymbol{X} = \{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n\}$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathsf{T}}$. When all weights are equal $\omega_1 = \cdots = \omega_n$ we speak of **plain least squares approximation**. Since the ansatz is linear the considered least squares approximation is linear. Thus, we may formulate the optimization problem in terms of matrices: minimize $\hat{\boldsymbol{g}} \mapsto \|\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 = (\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y})^* \boldsymbol{W} (\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y})$ with

$$\begin{aligned} \boldsymbol{L} &= \begin{pmatrix} \eta_1(\boldsymbol{x}^1) & \dots & \eta_{m-1}(\boldsymbol{x}^1) \\ \vdots & \ddots & \vdots \\ \eta_1(\boldsymbol{x}^n) & \dots & \eta_{m-1}(\boldsymbol{x}^n) \end{pmatrix} \in \mathbb{K}^{n\times(m-1)}, \\ \boldsymbol{W} &= \begin{pmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_n \end{pmatrix} \in [0,\infty)^{n\times n}, \end{aligned} \tag{2.1}$$

$\hat{\boldsymbol{g}} = (\hat{g}_1, \dots, \hat{g}_{m-1})^\mathsf{T} \in \mathbb{K}^{m-1}$, and $\boldsymbol{y} = (y_1, \dots, y_n)^\mathsf{T} \in \mathbb{K}^n$.

Lemma 2.1. *Let $\boldsymbol{L} \in \mathbb{K}^{n\times m-1}$, $\boldsymbol{W} \in [0,\infty)^{n\times n}$, and $\boldsymbol{y} \in \mathbb{K}^n$ be as above and let $\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L}$ be invertible. Further, let $\hat{\boldsymbol{g}}^\star$ be the minimizer of the least squares functional $\|\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2$. Then $\hat{\boldsymbol{g}}^\star$ is uniquely determined by*

$$\hat{\boldsymbol{g}}^\star = (\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{y}\,.$$

Proof. The assumption implies the positive definiteness of $\boldsymbol{W}$ and, therefore, the strong convexity of the given optimization problem. Thus, by computing stationary points we find the unique minimizer. For that we compute the root of the gradient of the least squares functional

$$\nabla_{\hat{\boldsymbol{g}}}(\|\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2) = 2\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L}\hat{\boldsymbol{g}} - 2\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{y} \overset{!}{=} \boldsymbol{0}\,.$$

Since $\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L}$ is positive definite, we obtain the explicit formula for $\hat{\boldsymbol{g}}^\star$. ■

So far, we introduced the least squares approximation method without any guarantee for its quality. In Chapters 7 and 8 we will show error guarantees for function approximation in different settings. For now, we state that the maximum likelihood estimator for data with normal (Gaussian) noise coincides with the least squares estimator which was first observed by C. F. Gauss in [Gau11].

Lemma 2.2. *Let $V = \operatorname{span}\{\eta_1, \dots, \eta_{m-1}\} \subseteq \mathbb{K}^D$, $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq D$, and $\sigma^2 \geq 0$. Further, let $Y_i \sim \mathcal{N}(\sum_{k=1}^{m-1} \hat{g}_k \eta_k(\boldsymbol{x}^i), \sigma^2)$ for $i = 1, \dots, n$ be random variables, where $\mathcal{N}$ denotes the normal distribution with variance σ^2. Or equivalently, $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{L}\hat{\boldsymbol{g}}, \sigma^2\boldsymbol{I})$. Then the maximum likelihood estimator coincides with the least squares approximation $S_V^{\boldsymbol{X}}(\boldsymbol{L}\hat{\boldsymbol{g}})$.*

Proof. Since Y_i are i.i.d., the likelihood function evaluates to

$$\begin{aligned}\mathcal{L}(\hat{\boldsymbol{g}}|\boldsymbol{y}) = \mathbb{P}(\boldsymbol{y}|\hat{\boldsymbol{g}}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{|y_i - [\boldsymbol{L}\hat{\boldsymbol{g}}]_i|^2}{2\sigma^2}\Big) \\ &= (2\pi\sigma^2)^{-n/2} \exp\Big(- \frac{1}{2\sigma^2} \sum_{i=1}^{n} |y_i - [\boldsymbol{L}\hat{\boldsymbol{g}}]_i|^2\Big) .\end{aligned}$$

Using that $x \mapsto \exp(-x)$ is monotone decreasing, we obtain

$$\arg\max_{\hat{\boldsymbol{g}}\in\mathbb{K}^{m-1}} \mathcal{L}(\hat{\boldsymbol{g}}|\boldsymbol{y}) = \arg\min_{\hat{\boldsymbol{g}}\in\mathbb{K}^{m-1}} \sum_{i=1}^{n} |y_i - [\boldsymbol{L}\hat{\boldsymbol{g}}]_i|^2 ,$$

which is the defining equation of the least squares estimator. ■

2.3 Implementation

With Theorem 2.2 we covered first theoretical properties of the least squares approximation method which will be continued in Chapters 7 and 8. This reasons for the applicability of the method in practice. For the implementation, one could utilize the explicit solution formula from Theorem 2.1. But this makes use of a matrix inverse, which is unfeasible to compute when the problem is large.

Instead, we utilize Krylov subspace iteration methods which were developed by the naval architect A. N. Krylov (1863-1945) and were elected as one of the top 10 algorithms of the 20th century by the *IEEE Computing in Science & Engineering Magazine* [DS00]. Methods of this type search the solution in **Krylov subspaces**

$$\mathcal{K}_r(\boldsymbol{A}, \boldsymbol{b}) \coloneqq \operatorname{span}\{\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, \dots, \boldsymbol{A}^{r-1}\boldsymbol{b}\} ,$$

where $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, $\boldsymbol{b} \in \mathbb{C}^n$, and $r \in \mathbb{N}$. A famous example is the conjugated gradient method from M. R. Hestenes and E. Stiefel [HS52]. Analytically equivalent, but tailored to the least squares problem in regards to numerical stability is the `LSQR` algorithm by C. C. Paige and M. A. Saunders [PS82].

Using $r = n$ iterations, these methods exhaust the full search space and find an exact solution. In general this is not necessary and one may stop earlier. Since only matrix-vector products are utilized this results in a runtime of $\mathcal{O}(rn^2)$, where r is the number of iterations. This can be improved even further, when the matrix-vector product is computable in a fast manner by using, e.g. sparsity in the matrix structure or fast Fourier algorithms, which are on the list of top 10 algorithms of the 20th century too, cf. [DS00]. This then reduces the solution of the least squares problem to $\mathcal{O}(r \cdot \operatorname{nnz}(\boldsymbol{A}))$ with $\operatorname{nnz}(\boldsymbol{A})$ the number of non-zero entries of $\boldsymbol{A}$ or $\mathcal{O}(rn \log n)$ when Fourier algorithms are applicable. Regardless of the matrix structure, the runtime depends linear on the number of iterations r. The following lemma gives a bound on the relative error after r iterations when the extremal singular values $\sigma_{\min}(\boldsymbol{W}^{1/2}\boldsymbol{L}) = \lambda_{\min}(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})$ and $\sigma_{\max}(\boldsymbol{W}^{1/2}\boldsymbol{L}) = \lambda_{\max}(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})$ are well-behaved.

Lemma 2.3. *Let $\boldsymbol{W}^{1/2}\boldsymbol{L} \in \mathbb{K}^{(m-1)\times n}$ with singular values $1/2 \le \sigma_{\min}^2 \le \sigma_{\max}^2 \le 3/2$, $\boldsymbol{y} \in \mathbb{K}^n$, and $\hat{\boldsymbol{g}}^\star = \arg\min_{\hat{\boldsymbol{g}}\in\mathbb{K}^{m-1}} \|\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 \in \mathbb{K}^{m-1}$. The residual $\hat{\boldsymbol{g}}^{(r)}$ of the* `LSQR` *algorithm in the r-th step then satisfies*

$$\|\hat{\boldsymbol{g}}^{(r)} - \hat{\boldsymbol{g}}^\star\|_2 \le 3 \cdot 2^{1-r} \|\hat{\boldsymbol{g}}^\star\|_2 \,.$$

Proof. Applying the `LSQR` algorithm is analytically equivalent to applying the conjugate gradient method to the system of equations $\boldsymbol{L}^*\boldsymbol{L}\hat{\boldsymbol{g}} = \boldsymbol{L}^*\boldsymbol{y}$. It is left to use the error estimate for conjugate gradient method from [Gre97, Theorem 3.1.1]. ■

In Chapter 6, we show settings and conditions for the matrix $\boldsymbol{W}^{1/2}\boldsymbol{L}$ to be well-conditioned as assumed above. Computing in double precision, Theorem 2.3 guarantees that the relative error $\|\hat{\boldsymbol{g}}(r) - \hat{\boldsymbol{g}}^\star\|_2/\|\hat{\boldsymbol{g}}^\star\|_2$ is smaller than machine epsilon $\varepsilon = 10^{-16}$ using $r = 56$ iterations. In our numerical experiments we set the maximal number of iterations in the range of 10 to 20, which seems sufficient.

This makes the least squares approximation method fast and we turn our attention to the task of finding and preparing tools to prove error bounds for it.

Chapter 3

Reproducing kernel Hilbert spaces (RKHS)

As we work with discrete function evaluations and functions $f\colon D \to \mathbb{K}$, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, may be very arbitrary, we need some sort of prior. In this chapter we will have a look at a broad range of function spaces, which allows us to introduce minimal restrictions to show error bounds for least squares approximation later on. The concept in question are reproducing kernel Hilbert spaces $H(K)$, which are obtained by requiring the continuity of function evaluations in a Hilbert function space. This condition is rather simple but has far reaching consequences. For instance the existence and correspondence to kernels, i.e., a function of two variables $K(x,y)$, with the "reproducing property"

$$f(x) = \langle f, K(\cdot, x)\rangle_{H(K)} \,.$$

The theory of reproducing kernel Hilbert spaces was researched from two different angles in the beginning of the 20th century:

- One group considered a given kernel and studied it in itself, applying it to various fields like integral equations, and only using the corresponding function class as a tool of research. Some representatives are J. Mercer [Mer09, Mer11], E. H. Moore [Moo16], S. Bochner [Boc22], A. Weil [Wei40], or R. Godement [God48]. In 1935 E. H. Moore established a first link by showing that for kernels with certain properties there is a Hilbert space in which the kernel has the reproducing property, cf. [Moo35, Moo39].

- The second group had a look at function classes coming primarily from the solution of partial differential equations. Computing the corresponding reproducing kernel as a tool was used by e.g. S. Zaremba [Zar07, Zar08], G. Szegő [Sze21], S. Bergmann [Ber22], N. Aronszajn [Aro35], or S. Bergman and M. Schiffer [BS47b, BS47a, BS48]. In 1939 N. Aronszajn showed that for a class of functions there is a kernel with a reproducing property, cf. [Aro44] .

It was 1950 when a first systematic study was done by N. Aronszajn in order to link the two research directions, cf. [Aro50]. Later overviews include H. Meschkowski [Mes62] or L. Schwartz [Sch64]. More modern approaches can be found in [BTA04] by A. Berlinet and C. Thomas-Agnan, [Wen05a] by H. Wendland, or in [SC08] by I. Steinwart and A. Christmann, where we orient ourselves on the latter.

In particular, we will start by introducing the RKHSs itself along with some basic concepts in Section 3.1. Then we focus on the construction of basis functions from a so-called Mercer-representation of a kernel in Section 3.2. After introducing auxiliary functional analysis tools in Section 3.3, we use them in Section 3.4 to show the existence of a weak Mercer-representation of RKHSs compactly embedded into the space of square integrable functions L_2. This, in turn, yields basis functions for the respective RKHS. Finally, we give one- and multidimensional examples in Sections 3.5 and 3.6, respectively.

3.1 Basic concepts of RKHSs

In this section we introduce the RKHSs, show selected properties, and repeat on the connection of RKHSs and kernels. We start with their definitions.

Definition 3.1. *Let H be a $\mathbb{K}$-Hilbert function space over a domain $D \neq \emptyset$.*

(i) We say H is a ***reproducing kernel Hilbert space (RKHS)*** *when the Dirac functional*

$$\delta_x \colon H \to \mathbb{K},\ f \mapsto f(x)$$

is continuous for all $x \in D$.

(ii) We say $K \colon D \times D \to \mathbb{K}$ is a ***reproducing kernel*** *for H, if*

$$f(x) = \langle f, K(\cdot, x)\rangle_H$$

for all $f \in H$ and $x \in D$.

Due to the continuity of the sampling δ_x in an RKHS, norm convergence implies pointwise convergence, i.e., for $(f_n)_{n\in\mathbb{N}} \subseteq H$ with $f_n \to f$ in H we have

$$|f_n(x) - f(x)| = |\delta_x(f_n - f)| \leq \|\delta_x\|_{H'} \|f_n - f\|_H \to 0\,,$$

where $\|\varphi\|_{H'} = \sup_{f\in H} |\varphi(f)|$ is the norm in the dual space H' of H. Thus, the norm in an RKHS is stronger than the pointwise convergence.

Next, we define symmetric positive definite kernels K and show their one-to-one correspondence to RKHSs.

Definition 3.2. *Let $K \colon D \times D \to \mathbb{K}$.*

*(i) We say K is **symmetric** when $K(x,y) = \overline{K(y,x)}$ for all $x, y \in D$.*

*(ii) We say K is **positive definite** when*

$$\sum_{i,j=1}^{n} \alpha_i \overline{\alpha_j} K(x_j, x_i) \geq 0$$

for all $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$ and $x_1, \ldots, x_n \in D$.

Note, that the reproducing property implies the symmetry and positive definiteness of the kernel:

Lemma 3.3. *A reproducing kernel K is symmetric and positive definite.*

Proof. Since $K(x,y) = \langle K(\cdot, y), K(\cdot, x) \rangle_H$ the symmetry is inherited from the scalar product. The positive definiteness follows from

$$\sum_{i,j=1}^{n} \alpha_i \overline{\alpha_j} K(x_j, x_i) = \Big\langle \sum_{i=1}^{n} \alpha_i K(\cdot, x_i), \sum_{j=1}^{n} \alpha_j K(\cdot, x_j) \Big\rangle_H \geq 0\,. \qquad \blacksquare$$

The next theorem states that every RKHS has a unique reproducing kernel, which by the above lemma is symmetric and positive definite.

Theorem 3.4. *Let H be an RKHS over D. Then*

$$K \colon D \times D \to \mathbb{K}, \quad (x,y) \mapsto \langle \delta_x, \delta_y \rangle_{H'}$$

is the unique reproducing kernel of H.

A proof can be found in [SC08, Theorem 4.20]. As it is good to familiarize ourselves with the reproducing concepts, we state it here as well.

Proof. **Step 1.** Showing the reproducing property. By the Riesz representer theorem there exists $\phi_x \in H$ such that $\delta_x(f) = \langle f, \phi_x \rangle_H$ for all $f \in H$ and $x \in D$. For $x, y \in D$, we obtain

$$K(x,y) = \langle \delta_x, \delta_y \rangle_{H'} = \langle \phi_x, \phi_y \rangle_H = \phi_y(x)\,,$$

and, thus,

$$f(y) = \delta_y(f) = \langle f, \phi_y \rangle_H = \langle f, K(\cdot, y) \rangle_H\,.$$

Step 2. Showing the uniqueness. Assume we have two reproducing kernels K and K'. Then

$$\langle f, K(\cdot, x) - K'(\cdot, x) \rangle_H = f(x) - f(x) = 0$$

for all $f \in H$. In particular for $f = K(\cdot, x) - K'(\cdot, x)$ we obtain $\|K(\cdot, x) - K'(\cdot, x)\|_H = 0$ for all $x \in D$, i.e., $K = K'$. $\blacksquare$

The direction of showing that every symmetric, positive definite kernel has a corresponding RKHS is known as Moore-Aronszajn Theorem.

Theorem 3.5 (Moore-Aronszajn)**.** *Let $K\colon D \times D \to \mathbb{K}$ be a positive definite symmetric function. Then there exists a unique Hilbert function space $H(K)$ with K as reproducing kernel.*

The proof is available in [BTA04, Theorem 3] or [Wen05a, Theorems 10.10 and 10.11] and is constructive.

Now we have established the one-to-one connection of RKHSs and symmetric, positive definite kernels. This justifies the notation $H(K)$ for the RKHS corresponding to the kernel K known from the literature, which we adapt. Further, various problems for a function space $H(K)$ become transparent from the kernel perspective and vice versa. The following two lemmata are on the construction of new RKHSs and the connection of different kernel properties to properties of the functions in the corresponding RKHS.

Lemma 3.6. *Let $H_i(K_i)$ be RKHSs on D_i with reproducing kernels K_i for $i = 1, 2$. Then*

(i) for $D_1 = D_2$ the kernel $K = K_1 + K_2\colon D_1 \times D_1 \to \mathbb{K}$ is the reproducing kernel of the space $H(K) = H_1(K_1) \oplus H_2(K_2) = \{f\colon D_1 \to \mathbb{K} : f = f_1 + f_2, f_1 \in H_1(K_1), f_2 \in H_2(K_2)\}$.

(ii) the kernel $K = K_1 \otimes K_2\colon (D_1 \times D_2) \times (D_1 \times D_2) \to \mathbb{K}$ is the reproducing kernel of the space $H(K) = H_1(K_1) \otimes H_2(K_2) = \{f\colon D_1 \times D_2 \to \mathbb{K} : f = f_1 \otimes f_2, f_1 \in H_1(K_1), f_2 \in H_2(K_2)\}$.

Proof. Theorems 5 and 13 of [BTA04]. ■

Lemma 3.7. *Let $H(K)$ be an RKHS on D with reproducing kernel K.*

(i) Then K is bounded if and only if every $f \in H(K)$ is bounded.

(ii) Let (D, Σ) be a measurable space. Then $K(\cdot, x)\colon D \to \mathbb{K}$ is measurable for all $x \in D$ if and only if every $f \in H(K)$ is measurable.

(iii) Let D be a topological space. Then $K(\cdot, x)\colon D \to \mathbb{K}$ is continuous and bounded for $x \in D$ if and only if every $f \in H(K)$ is continuous and bounded.

(iv) Let $D \subseteq \mathbb{R}^d$ be an open subset and $m \geq 0$. If

$$\frac{\partial^{\|\boldsymbol{\alpha}\|_1}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} \frac{\partial^{\|\boldsymbol{\alpha}\|_1}}{\partial \boldsymbol{y}^{\boldsymbol{\alpha}}} K(\boldsymbol{x}, \boldsymbol{y}) = \frac{\partial^{\|\boldsymbol{\alpha}\|_1}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \frac{\partial^{\|\boldsymbol{\alpha}\|_1}}{\partial y_1^{\alpha_1} \cdots y_d^{\alpha_d}} K(\boldsymbol{x}, \boldsymbol{y})$$

exists and is continuous for all multi-indices $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)^\mathsf{T} \in \mathbb{N}_0^d$ *with* $\|\boldsymbol{\alpha}\|_1 \leq m$*, then* $\frac{\partial^{\|\boldsymbol{\alpha}\|_1}}{\partial x^{\boldsymbol{\alpha}}} f(\boldsymbol{x})$ *exists and is continuous for all* $f \in H(K)$ *and all multi-indices* $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ *with* $|\boldsymbol{\alpha}| \leq m$*.*

Proof. The statements are found in Lemmata 4.23, 4.24, 4.28, and Corollary 4.36 of [SC08]. ■

3.2 Constructing RKHSs from a given Mercer representation

A **Mercer representation** of a kernel K is an expansion in terms of an orthonormal system $\{[\eta_1]_\sim, [\eta_2]_\sim, \dots\} \subseteq L_2$ of the space of square integrable functions L_2, cf. Theorem 3.9. In particular, we then construct

$$K(x, y) := \sum_{k=1}^{\infty} \sigma_k^2 \overline{\eta_k(y)} \eta_k(x)$$

for a sequence $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ which defines an RKHS $H(K)$, cf. Theorem 3.11 or [SC08, Formula (3)]. Given such a representation yields insights to approximation of functions $f \in H(K)$ when the error is measured in L_2 which we cover in this section. Further, we will answer the question of the best possible finite rank approximation of the embedding $I_K \colon H(K) \to L_2$. This is an indicator on how good finite discretization can perform which will be a benchmark when we work with finitely many samples. The quantity of interest is modeled by the approximation numbers a_m of an operator, cf. [Pie87, Definition 2.3.1].

Definition 3.8. *For* H_1*,* H_2 *Hilbert spaces, and* $T \in \mathcal{L}(H_1, H_2)$ *an operator, we define the* m*-th* ***approximation number*** $a_m(T)$ *via*

$$\begin{aligned} a_m(T) &:= \inf\{\|T - A\|_{\mathcal{L}(H_1,H_2)} : \operatorname{rank}(A) < m\} \\ &= \inf_{\substack{L_1,\dots,L_{m-1} \in H_1' \\ \varphi_1,\dots,\varphi_{m-1} \in H_2}} \sup_{\|f\|_{H_1} \leq 1} \Big\| Tf - \sum_{k=1}^{m-1} L_k(f)\varphi_k \Big\|_{H_2} . \end{aligned}$$

In our case H_1 will be the RKHS $H(K)$ and H_2 the space of square integrable functions L_2. At the end of this section we specify $L_1, \dots, L_{m-1}$ and $\varphi_1, \dots, \varphi_{m-1}$ realizing the infimum.

Strictly speaking L_2 is not a function space but rather a Hilbert space of equivalence classes. Because this brings some problems, we pay special attention to this in this section starting with the definition of L_2.

Definition 3.9. *Let (D, Σ, ϱ_T) be a measure space,*

$$\mathcal{L}_2 = \mathcal{L}_2(D, \varrho_T) := \Big\{ f\colon D \to \mathbb{K} : \|f\|_{\mathcal{L}_2}^2 = \int_D |f|^2 \,\mathrm{d}\varrho_T < \infty \Big\},$$

$$\text{and} \quad \mathcal{N} = \mathcal{N}(D, \varrho_T) := \{ f\colon D \to \mathbb{K} : f = 0 \ \varrho_T\text{-almost everywhere} \} .$$

The ***Lebesgue space of square-integrable functions*** *is the quotient space $L_2 = L_2(D, \varrho_T) := \mathcal{L}_2(D, \varrho_T)/\mathcal{N}$ consisting of equivalence classes $[f]_\sim := \{g \in \mathcal{L}_2(D, \varrho_T) : \varrho_T(\{f \neq g\}) = 0\}$ with the norm $\|[f]_\sim\|_{L_2} = \|f\|_{\mathcal{L}_2}$.*

Often the notation is abused by writing $f \in L_2$ or $\|f\|_{L_2}$ for functions $f\colon D \to \mathbb{K}$ for convenience of readability. We do the same everywhere but this chapter, where we have a close look at the consequences of the ϱ_T-null sets.

The first intricacy is the embedding operator itself, as it has to map functions to equivalence classes. The substitute in this case is the operator

$$I_K\colon H(K) \to L_2, \quad f \mapsto [f]_\sim \,, \tag{3.1}$$

which does the same up to ϱ_T-null sets and is well-defined for kernels with finite Frobenius-norm $\|K\|_F^2 := \int_D \int_D K(x,y) \,\mathrm{d}\varrho_T(x) \,\mathrm{d}\varrho_T(y) < \infty$.

The embedding operator is connected with an integral operator, which will be a useful tool for our analysis and is introduced in the next lemma.

Lemma 3.10. *Let K be a kernel with finite Frobenius norm $\|K\|_F < \infty$ and T_K be the following Fredholm integral operator of first kind*

$$T_K\colon L_2 \to L_2, \quad [f]_\sim \mapsto \Big[\int_D K(\cdot, x) f(x) \,\mathrm{d}\varrho_T(x) \Big]_\sim \,, \tag{3.2}$$

I_K the inclusion operator (3.1)*, and $S_K = I_K^*$ its adjoint. Then the following diagram commutes:*

$$\begin{tikzcd}
L_2(D, \varrho_T) \arrow[rr, "T_K"] \arrow[dr, "S_K"'] & & L_2(D, \varrho_T) \\
& H(K) \arrow[ur, "I_K"'] &
\end{tikzcd} .$$

Proof. With the finite Frobenius-norm of the kernel, we have that the Fredholm integral operator is well-defined. Further, we have $\langle S_K [f]_\sim, g \rangle_{H(K)} =$

$\langle [f]_\sim, I_K g\rangle_{L_2} = \int_D f\overline{g}\,\mathrm{d}\varrho_T$. In particular, for $g = K(\cdot, y)$

$$(S_K[f]_\sim)(y) = \langle S_K[f]_\sim, K(\cdot, y)\rangle_{H(K)} = \int_D K(y,x)f(x)\,\mathrm{d}\varrho_T(x)\,.$$

Thus, $T_K = I_K S_K$. ■

Now we use the integral operator T_K to show that the Mercer representation does define an RKHS.

Lemma 3.11. *We consider a kernel with a given* ***Mercer representation***

$$K(x,y) := \sum_{k=1}^{\infty} \overline{e_k(y)} e_k(x)\,,$$

such that for numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ and $e_k := \sigma_k \eta_k$ we have an orthonormal system $\{[\eta_1]_\sim, [\eta_2]_\sim, \dots\} \subseteq L_2$. Then K defines an RKHS $H(K)$ with the orthonormal basis $\{e_1, e_2, \dots\}$.

Proof. By definition the kernel K is symmetric and positive definite, which implies the unique existence of an RKHS $H(K)$ with K as reproducing kernel by the Moore-Aronzajn Theorem 3.5. By the constructive nature of the proof, we know that e_k is a basis of $H(K)$. It remains to show the orthonormality.

For the integral operator T_K from (3.2) we have

$$\begin{aligned} T_K[f]_\sim &= \Big[\int_D \sum_{k=1}^{\infty} \overline{e_k(x)} e_k f(x)\,\mathrm{d}\varrho_T(x)\Big]_\sim \\ &= \sum_{k=1}^{\infty} \int_D f(x)\overline{e_k(x)}\,\mathrm{d}\varrho_T(x)[e_k]_\sim \\ &= \sum_{k=1}^{\infty} \sigma_k^2 \langle f, [\eta_k]_\sim\rangle_{L_2} [\eta_k]_\sim\,. \end{aligned}$$

Thus, σ_k^2 is an eigenvalue with eigenfunction $[\eta_k]_\sim$. With

$$S_K[\eta_k]_\sim = \int_D \sum_{\ell=1}^{\infty} \overline{e_\ell(x)} e_\ell \eta_k(x)\,\mathrm{d}\varrho(x) = \sum_{\ell=1}^{\infty} \langle \eta_k, e_\ell\rangle_{L_2} e_\ell = \sigma_k e_k$$

we obtain the orthonormality of $\{e_k\}_{k=1}^{\infty}$:

$$\langle e_k, e_\ell\rangle_{H(K)} = \frac{\langle S_K[\eta_k]_\sim, S_K[\eta_\ell]_\sim\rangle_{H(K)}}{\sigma_k\sigma_\ell} = \frac{\langle T_K[\eta_k]_\sim, [\eta_\ell]_\sim\rangle_{L_2}}{\sigma_k\sigma_\ell} = \delta_{k\ell}\,.$$

■

Having an RKHS from a Mercer representation of K, we next quantify the approximation numbers a_m from Theorem 3.8 giving insights to the discretization properties of $H(K)$.

Lemma 3.12. *Let the assumptions from Theorem 3.11 hold. Then*

$$a_m(I_K\colon H(K) \to L_2) = \sup_{\|f\|_{H(K)}\le 1} \Big\|[f]_\sim - P_m f\Big\| = \sigma_m\,,$$

where $P_m f = \sum_{k=1}^{m-1} \langle f, e_k\rangle_{H(K)} [e_k]_\sim$ *is the projection.*

Proof. Since $[e_k]_\sim = \sigma_k [\eta_k]_\sim$, we have

$$\sup_{\|f\|_{H(K)}\le 1} \Big\|f - P_m f\Big\| = \sup_{\|f\|_{H(K)}\le 1} \Big(\sum_{k=m}^{\infty} \sigma_k^2 |\langle f, e_k\rangle_{H(K)}|^2\Big)^{1/2} = \sigma_m\,,$$

which shows the worst-case error for the projection. It remains to show that the projection is the minimizer over all $L_k \in H'$ and $\varphi_k \in H$ as in the definition of the approximation number a_m:

$$\sup_{\|f\|_H \le 1} \Big\|[f]_\sim - \sum_{k=1}^{m-1} L_k(f)\varphi_k\Big\|_{L_2}\,.$$

Step 1. Showing that the minimum is attained for $\varphi_k = [e_k]_\sim$, $k = 1, \ldots, m-1$. Assuming $e_\ell \notin \operatorname{span}\{\varphi_1, \ldots, \varphi_{m-1}\}$ for some $1 \le \ell \le m-1$, we obtain

$$\sup_{\|f\|_{H(K)}\le 1} \Big\|[f]_\sim - \sum_{k=1}^{m-1} L_k(f)\varphi_k\Big\|_{L_2} \ge \|[e_\ell]_\sim\|_{L_2} = \sigma_\ell\,.$$

Since we have the monotonicity $\sigma_1 \ge \sigma_2 \ge \ldots$, this is bigger or equal σ_m yielding a contradiction.

Step 2. Showing that the minimum is attained for $L_k(f) = \langle f, e_k\rangle_{H(K)}$. With the L_2-orthogonality of $[e_k]_\sim$, we obtain

$$\begin{aligned}&\Big\|[f]_\sim - \sum_{k=1}^{m-1} L_k(f)[e_k]_\sim\Big\|_{L_2} \\ &\quad = \Big\|\sum_{k=1}^{m-1} (L_k(f) - \langle f, e_k\rangle_{H(K)})[e_k]_\sim\Big\|_{L_2} + \Big\|\sum_{k=m}^{\infty} \langle f, e_k\rangle_{H(K)}[e_k]_\sim\Big\|_{L_2}\,,\end{aligned}$$

where L_k does not influence the latter summand and the projection eliminates the first summand. ■

The above lemma shows that the functions $[e_k]_\sim$ corresponding to the largest σ_k give the optimal finite rank approximation with error σ_m in case we have given a Mercer representation. This will motivate to use these functions constructing the ansatz for the least squares approximation.

Having constructed an RKHS from a given Mercer representation the natural question arises whether the reverse is also true, i.e., if every kernel of an RKHS has a Mercer representation. To answer this question, we first need some tools from spectral theory, which we will cover in the next section.

3.3 Interlude on spectral theory of compact operators

In this section we state some general results from functional analysis in order to apply them to RKHSs later on.

The first result is the spectral theorem, which is often stated for self-adjoint or normal compact operators $T \in \mathcal{L}(H_1)$ in a $\mathbb{K}$-Hilbert space for $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, respectively. We use a general version for compact operators $T \in \mathcal{L}(H_1, H_2)$ mapping from one Hilbert space H_1 into another H_2.

Theorem 3.13 (general spectral theorem)**.** *Let H_1, H_2 be Hilbert spaces and $T \in \mathcal{L}(H_1, H_2)$ compact. Then there exist non-negative numbers $\sigma_1 \geq \sigma_2 \geq \dots$ and orthonormal systems $\{e_1, e_2, \dots\} \subseteq H_1$ and $\{\eta_1, \eta_2, \dots\} \subseteq H_2$ such that*

$$\begin{aligned} Tx &= \sum_{k=1}^{\infty} \sigma_k \langle x, e_k \rangle_{H_1} \eta_k \quad \textit{for all} \quad x \in H_1\,, \\ H_1 &= \ker(T) \oplus_{H_1} \overline{\operatorname{span}\{e_1, e_2, \dots\}}\,, \\ \textit{and} \qquad H_2 &= \ker(T^*) \oplus_{H_2} \overline{\operatorname{span}\{\eta_1, \eta_2, \dots\}}\,. \end{aligned}$$

We call $\sigma_1, \sigma_2, \dots$ the **singular numbers**, $e_1, e_2, \dots$ and $\eta_1, \eta_2, \dots$ the **left-** and **right-singular functions** of T, respectively.

Proof. Since $T^*T \colon H_1 \to H_1$ is compact and self-adjoint, we apply the common spectral theorem, cf. [Wer00, Theorem VI.3.2]. We obtain a zero sequence $\sigma_1^2, \sigma_2^2, \dots \geq 0$ (since T^*T is positive) and an orthonormal system $\{e_1, e_2, \dots\} \subseteq H_1$ such that

$$T^*Tx = \sum_{k=1}^{\infty} \sigma_k^2 \langle x, e_k \rangle_{H_1} e_k \quad \text{for all} \quad x \in H_1$$

and $H_1 = \ker(T^*T) \oplus_{H_1} \overline{\operatorname{span}\{e_1, e_2, \dots\}}$.

With $\langle T^*Ty, y\rangle_{H_1} = \langle Ty, Ty\rangle_{H_2} = \|Ty\|_{H_2}$, we obtain $\ker(T^*T) = \ker(T)$, which shows the second equation. Doing the same for TT^*, we obtain the third equation. It remains to show the spectral representation.

The functions $\eta_k := Te_k/\sqrt{\sigma_k^2}$ form an orthonormal system in H_2 since

$$\langle Te_k, Te_\ell\rangle_{H_2} = \langle T^*Te_k, e_\ell\rangle_{H_1} = \sigma_k^2\langle e_k, e_\ell\rangle_{H_1} = \sigma_k^2\delta_{k\ell}\,.$$

For $x = y + \sum_{k=1}^{\infty}\langle x, e_k\rangle_{H_1} e_k$ with $y \in \ker(T)$, we obtain

$$Tx = Ty + \sum_{k=1}^{\infty}\langle x, e_k\rangle_{H_1} Te_k = \sum_{k=1}^{\infty}\sqrt{\sigma_k^2}\langle x, e_k\rangle_{H_1}\eta_k\,. \qquad \blacksquare$$

Note, that the spectral theorem is the infinite-dimensional generalization of the singular value decomposition of matrices. For Banach spaces there are many generalizations for singular values of an operator, which all coincide in the Hilbert case. For an overview we refer to [Pie87]. In our case, the approximation numbers a_m from Theorem 3.8 are closest to our interest.

Lemma 3.14 ([Pie87, Proposition 2.11.6])**.** *Let H_1, H_2 be Hilbert spaces, and $T \in \mathcal{L}(H_1, H_2)$ compact. Then the singular numbers $\sigma_1, \sigma_2, \ldots$ of T and its approximation numbers coincide, i.e., $a_m(T) = \sigma_m$.*

This shows immediately that the truncated singular value decomposition is the best finite rank approximation of T. Using approximation numbers, we introduce the Schatten-von Neumann operators, cf. [Pie87, Definition 2.11.15].

Definition 3.15. *For H_1 and H_2 Hilbert spaces, we define the* ***Schatten-von Neumann classes*** *for $0 < p$ by*

$$\mathcal{S}_p(H_1, H_2) := \begin{cases} \{T \in \mathcal{L}(H_1, H_2) : \sum_{m=1}^{\infty} a_m(T)^p < \infty\} & : 0 < p < \infty\,, \\ \mathcal{L}(H_1, H_2) & : p = \infty\,. \end{cases}$$

In particular, if $T \in \mathcal{S}_1(H_1, H_2)$ we say T is ***nuclear*** *and if $T \in \mathcal{S}_2(H_1, H_2)$ we say T is* ***Hilbert-Schmidt****.*

The Schatten-von Neumann classes generalize the trace-class operators and are compact for $p < \infty$ as well.

Lemma 3.16. *If $T \in \mathcal{S}_p(H_1, H_2)$ for $p < \infty$, then T is compact.*

Proof. $T \in \mathcal{S}_p(H_1, H_2)$ for $p < \infty$ implies that $a_k(T) \to 0$ for $k \to \infty$. Thus, there is a sequence of finite rank operators converging to T. The compactness of T follows from [Con90, Theorem 4.4]. $\blacksquare$

Lemma 3.17. *Let H_1, H_2 be Hilbert spaces, $T \in \mathcal{L}(H_1, H_2)$ a linear operator, and $\{e_k\}_{k\in I} \subseteq H_1$ an orthonormal system in H_1 for a possible uncountable index set I. If $\sum_{k\in I} \|Te_k\|_{H_2}^2 < \infty$ then $\sum_{k\in I} \|Te_k\|_{H_2}^2 = \sum_{k=1}^{\infty} a_k(T)^2$. In particular, $T \in \mathcal{S}_2(H_1, H_2)$.*

Proof. **Step 1.** We show $J := \{k \in I : \|Te_k\|_{H_2}^2 > 0\}$ is at most countable. For $A_n := \{k \in I : \|Te_k\|_{H_2}^2 \geq 1/n\}$, we have

$$\frac{|A_n|}{n} = \sum_{k\in A_n} \frac{1}{n} \leq \sum_{k\in A_n} \|Te_k\|_{H_2}^2 \leq \sum_{k\in I} \|Te_k\|_{H_2}^2 < \infty .$$

Thus $J = \bigcup_{n=1}^{\infty} A_n$ is at most countable.

Step 2. Showing the compactness of T. Without loss of generality let $J = \mathbb{N}$. For an element $x = \sum_{k\in I} \langle x, e_k\rangle_{H_1} e_k \in H_1$ we define $T_{m-1}x := \sum_{k=1}^{m-1} \langle x, e_k\rangle_{H_1} Te_k$. Using triangle inequality, Chauchy-Schwarz inequality, and Bessel's inequality, we obtain

$$\begin{aligned}\|(T - T_{m-1})x\|_{H_2} &\leq \sum_{k\in I\setminus\{1,\ldots,m-1\}} |\langle x, e_k\rangle_{H_1}| \cdot \|Te_k\|_{H_2} \\ &\leq \sqrt{\sum_{k\in I\setminus\{1,\ldots,m-1\}} |\langle x, e_k\rangle_{H_1}|^2} \sqrt{\sum_{k\in I\setminus\{1,\ldots,m-1\}} \|Te_k\|_{H_2}^2} \\ &\leq \|x\|_{H_1} \sqrt{\sum_{k=m}^{\infty} \|Te_k\|_{H_2}^2} .\end{aligned}$$

Thus, $\|T - T_{m-1}\|_{\mathcal{L}(H_1,H_2)} \leq \sqrt{\sum_{k=m}^{\infty} \|Te_k\|_{H_2}^2} \to 0$, i.e., T_{m-1} is a sequence of finite rank operators converging to T. By [Con90, Theorem 4.4], this implies the compactness of T.

Step 3. Using the spectral Theorem 3.13 and Theorem 3.14, we have $Tx = \sum_{m=1}^{\infty} a_m(T)\langle x, f_m\rangle_{H_1} \eta_m$ for $\{f_1, f_2, \ldots\}$ and $\{\eta_1, \eta_2, \ldots\}$ orthonormal bases in H_1 and H_2, respectively. In particular,

$$\|Te_k\|_{H_2}^2 = \Big\| \sum_{m=1}^{\infty} a_m(T)\langle e_k, f_m\rangle_{H_1} \eta_m \Big\|_{H_2}^2 = \sum_{m=1}^{\infty} a_m(T)^2 |\langle e_k, f_m\rangle_{H_1}|^2$$

and by Fubini's theorem

$$\begin{aligned}\sum_{k\in I}\|Te_k\|_{H_2}^2 &= \sum_{m=1}^{\infty} a_m(T)^2 \sum_{k\in I} |\langle e_k, f_m\rangle_{H_1}|^2\\ &= \sum_{m=1}^{\infty} a_m(T)^2 \|f_m\|_{H_1}^2 = \sum_{m=1}^{\infty} a_m(T)^2 \,. \end{aligned} \qquad \blacksquare$$

We will use the tools presented in this section to construct bases of RKHSs giving direct insights into the approximation of functions by discretization. It goes without saying, that this theory has many more applications.

3.4 Weak Mercer representation for RKHSs compactly embedded into L_2

In Section 3.2 we constructed an RKHS from a given Mercer representation. Now, we show that the finite trace condition (3.4) ensures the compactness of the embedding into L_2, giving us a weak Mercer representation (3.5). This in turn allows us to use the theorems on approximation from Section 3.2. In particular, this allows us to construct a system of functions $\{e_1, e_2, \dots\} \subseteq H(K)$ which is simultaneously orthogonal in $H(K)$ and L_2. Being the solution of well-studied Fredholm integral equations (3.2) of first kind, this gives access to a pool of ansatz functions for approximation.

For $\{f_k\}_{k\in I}$ an orthonormal basis in $H(K)$ with an possibly uncountable index set I, we have

$$K(\cdot,x) = \sum_{k\in I}\langle K(\cdot,x), f_k\rangle_{H(K)} f_k = \sum_{k\in I}\overline{f_k(x)} f_k, \tag{3.3}$$

which holds even pointwise. In general our constructed $\{e_1, e_2, \dots\}$ are not a basis of $H(K)$ and $K(x,y) = \sum_{k=1}^{\infty}\overline{e_k(y)}e_k(x)$ does not hold pointwise in both components. But requiring separability of $H(K)$ we will ensure the above representation almost everywhere, which is sufficient for our purposes. In order to do that, we start with defining properties of embeddings and the finite trace condition of kernels.

Definition 3.18. *Let (D, Σ, ϱ_T) be a measure space and $H(K)$ an RKHS with measurable reproducing kernel $K\colon D\times D \to \mathbb{K}$. We say $H(K)$ is continuously (compactly) embedded into L_2 when I_K from* (3.1) *is continuous (compact).*

Lemma 3.19. *Let (D, Σ, ϱ_T) be a measure space and $H(K)$ an RKHS with measurable reproducing kernel $K\colon D \times D \to \mathbb{K}$ with* ***finite trace****, i.e.,*

$$\operatorname{trace}(K) := \int_D K(x,x)\,\mathrm{d}\varrho_T(x) < \infty\,. \tag{3.4}$$

Then I_K from (3.1)*, $S_K := I_K^*$ are Hilbert-Schmidt and $T_K = I_K S_K$ from* (3.2) *is nuclear. In particular, $H(K)$ is compactly embedded into L_2.*

Proof. **Step 1.** Showing the continuity of the embedding I_K. Using the Cauchy-Schwarz inequality and $\|K(\cdot,x)\|_{H(K)}^2 = K(x,x)$, we obtain

$$\begin{aligned}\|f\|_{L_2}^2 &= \int_D |\langle f, K(\cdot,x)\rangle_{H(K)}|^2\,\mathrm{d}\varrho_T(x)\\ &\le \int_D \|f\|_{H(K)}^2 \|K(\cdot,x)\|_{H(K)}^2\,\mathrm{d}\varrho_T(x)\\ &= \|f\|_{H(K)}^2\operatorname{trace}(K)\,.\end{aligned}$$

Since the trace is finite by assumption, this gives the continuity of the embedding.

Step 2. Showing the compactness of the embedding I_K. Using Fubini's theorem, Bessel's inequality, and $\|K(\cdot,x)\|_{H(K)}^2 = K(x,x)$, we obtain

$$\begin{aligned}\sum_{k\in I}\|I_K e_k\|_{L_2}^2 &= \sum_{k\in I}\int_D |e_k(x)|^2\,\mathrm{d}\varrho_T(x)\\ &= \int_D \sum_{k\in I}|\langle e_k, K(\cdot,x)\rangle_{H(K)}|^2\,\mathrm{d}\varrho_T(x)\\ &\le \int_D \|K(\cdot,x)\|_{H(K)}^2\,\mathrm{d}\varrho_T(x)\\ &= \int_D K(x,x)\,\mathrm{d}\varrho_T(x) < \infty\,.\end{aligned}$$

Thus by Theorem 3.17 we have that I_K is Hilbert-Schmidt. Since $a_k(I_K) = a_k(S_K)$, S_K is Hilbert-Schmidt too. By Theorem 3.16 I_K and S_K are compact.

Step 3. T_K being nuclear follows from I_K and S_K being Hilbert-Schmidt by applying [Pie87, Theorem 2.3.13]. ■

Having the compactness of the embedding operator we will use the spectral Theorem 3.13 in order to construct basis of the RKHS and L_2.

Theorem 3.20. *Let (D, Σ, ϱ_T) be a measure space and $H(K)$ an RKHS with measurable reproducing kernel $K\colon D \times D \to \mathbb{K}$ with finite trace* (3.4)*.*

Then there are non-negative numbers $\sigma_1 \geq \sigma_2 \geq \dots$ and orthonormal systems $\{e_1, e_2, \dots\} \subseteq H(K)$ and $\{[\eta_1]_\sim, [\eta_2]_\sim, \dots\} \subseteq L_2$ with $\sigma_k[\eta_k]_\sim = [e_k]_\sim$ such that

$$H(K) = \ker(I_K) \oplus_{H(K)} \overline{\operatorname{span}\{e_1, e_2, \dots\}}$$

and

$$L_2 = \ker(S_K) \oplus_{L_2} \overline{\operatorname{span}\{[\eta_1]_\sim, [\eta_2]_\sim, \dots\}}\,.$$

Moreover, for I_K from (3.1)*, $S_K := I_K^*$, and T_K from* (3.2) *we have*

$$T_K = \sum_{k=1}^{\infty} \sigma_k^2 \langle \cdot, [\eta_k]_\sim \rangle_{L_2} [\eta_k]_\sim\,, \quad S_K I_K = \sum_{k=1}^{\infty} \sigma_k^2 \langle \cdot, e_k \rangle_{H(K)} e_k\,,$$

and

$$I_K = \sum_{k=1}^{\infty} \sigma_k \langle \cdot, e_k \rangle_{H(K)} [\eta_k]_\sim\,.$$

Proof. We apply Theorem 3.10 and Theorem 3.13 to I_K. ■

When I_K is injective, we have $\ker(I_K) = \{0\}$. Then, by Theorem 3.20, the functions $e_1, e_2, \dots$ form an orthonormal system in $H(K)$ and we have a pointwise Mercer representation (3.3). This is ensured by bounded and continuous kernels K (Mercer kernel), cf. [SC08, Theorem 4.49] or [BTA04, Theorem 40].

In our application we do not need the pointwise Mercer representation. Instead, assuming $H(K)$ separable and $\operatorname{trace}(K) < \infty$ gives us the representation almost everywhere as stated in the following lemma.

Lemma 3.21. *Let the assumptions from Theorem 3.20 hold and let $H(K)$ be separable. Then we have $\operatorname{trace}(K) = \sum_{k=1}^{\infty} \sigma_k^2$ and the weak Mercer representation, i.e., for ϱ_T-almost all $x \in D$*

$$K(\cdot, x) = \sum_{k=1}^{\infty} \overline{e_k(x)} e_k\,. \tag{3.5}$$

Proof. Let $\{d_k\}_{k \in I}$ be an orthonormal basis of $\ker(I_K)$ for an index set I and N_k the support of d_k. Since $H(K)$ is separable, I is at most countable and we use the countable additivity of ϱ_T

$$\varrho_T\Big(\bigcup_{k \in I} N_k\Big) \leq \sum_{k \in I} \varrho_T(N_k) = 0\,.$$

By (3.3) we obtain

$$K(\cdot,x)=\sum_{k=1}^{\infty}\overline{e_k(x)}e_k+\sum_{k\in I}\overline{d_k(x)}d_k\,,$$

where the latter sum vanishes ϱ_T-almost everywhere.

By Parseval's equality, we have ϱ_T-almost everywhere

$$K(x,x)=\|K(\cdot,x)\|_{H(K)}^2=\Big\|\sum_{k=1}^{\infty}\overline{e_k(x)}e_k\Big\|_{H(K)}^2=\sum_{k=1}^{\infty}|e_k(x)|^2\,.$$

Thus,

$$\operatorname{trace}(K)=\int_D K(x,x)\,\mathrm{d}\varrho_T(x)=\int_D\sum_{k=1}^{\infty}|e_k(x)|^2\,\mathrm{d}\varrho_T(x)\,.$$

Using Fubini's theorem the latter evaluates to $\sum_{k=1}^{\infty}\sigma_k^2$. ■

Note, without separability we have $\operatorname{trace}(K)\le\sum_{k=1}^{\infty}\sigma_k^2$ (via Bessel's inequality) and the above weak Mercer representation holds on $D\setminus N_x$ where $\varrho(N_x)=0$ and N_x depends on $x\in D$, cf. [SS12, Corollary 3.2]. This shows further, that the finite trace condition (3.4) is natural in sampling theory. In fact, it was shown that the sampling problem can be arbitrarily bad otherwise, cf. [HNV08, Theorem 1].

With that we have a base ground of priors for functions suitable in function approximation and reasoned for minimal necessary conditions, i.e., having a separable RKHS with finite trace.

3.5 Examples in one dimension

In this section we give examples for RKHSs in dimension $d=1$. An important class of functions are Sobolev spaces. They were originally introduced in 1950 by S. L. Sobolev in order to give a natural smoothness characterization for solving partial differential equations, cf. [Sob50]. It turned out that the smoothness in terms of existence of continuous derivatives is too harsh in this context. Instead, derivatives in a suitable weak sense are considered with respect to an L_p norm. As our goal is to approximate functions in the L_2 norm, Sobolev spaces fit our application as well.

For domains we focus on the torus $D=\mathbb{T}$ and the unit interval $D=[0,1]$. We will take the approach from a given Mercer representation following

Section 3.2 and also from a given Sobolev space where we apply the techniques from Section 3.4 giving different perspectives and generalizations.

An in-depth overview on Sobolev and related functions spaces is available in the book series by H. Triebel starting with [Tri10]. For a collection of RKHSs and corresponding kernels we refer to [BTA04, Chapter 7].

3.5.1 Sobolev spaces on the torus $\mathbb{T}$

To begin with, we use the one-dimensional torus for the domain $D = \mathbb{T} = \mathbb{R}_{/\mathbb{Z}}$. As the analysis is easy, this is a good starting point. The periodic Sobolev spaces on $\mathbb{T}$ arise naturally from a family of differential equations of the form

$$y^{(2s)} = (-1)^s \frac{1-\sigma^2}{\sigma^2} y, \quad y \in L_2(\mathbb{T}) \tag{3.6}$$

for $s \in \mathbb{N}$ and $\sigma^2 \leq 1$. They have the well-known solutions

$$y_r(x) = \exp\Big(\omega_{2s}^r \sqrt[2s]{\frac{1-\sigma^2}{\sigma^2}} x\Big), \quad r = 0, \dots, 2s-1 \qquad \text{for even } s \tag{3.7}$$

$$\text{and} \quad y_r(x) = \exp\Big(\omega_{4s}^r \sqrt[2s]{\frac{1-\sigma^2}{\sigma^2}} x\Big), \quad r = 1, 3, \dots, 4s-1 \quad \text{for odd } s\,, \tag{3.8}$$

where $\omega_N = \exp(2\pi \mathrm{i}/N)$ is the N-th root of unity. The torus enforces the periodicity of the solutions and, thus, we end up with

$$y_{1,2}(x) = \exp\Big(\pm \sqrt[2s]{\frac{1-\sigma^2}{\sigma^2}} \mathrm{i}x\Big)\,.$$

Next, we consider the corresponding weak formulation to use integration by parts. I.e., for a test function $\varphi \in C_0^\infty(\mathbb{T})$ (infinitely differentiable and vanishing at the border) we are having a look at

$$\begin{aligned} & \langle y^{(2s)}, \varphi \rangle_{L_2} = \Big\langle (-1)^s \frac{1-\sigma^2}{\sigma^2} y, \varphi \Big\rangle_{L_2} \\ \Leftrightarrow \qquad & \langle y, \varphi \rangle_{L_2} = \sigma^2(\langle y, \varphi \rangle_{L_2} + (-1)^s \langle y^{(2s)}, \varphi \rangle_{L_2}) \\ & \qquad = \sigma^2(\langle y, \varphi \rangle_{L_2} + \langle y^{(s)}, \varphi^{(s)} \rangle_{L_2})\,, \end{aligned}$$

where we used the fact, that we are able to transfer the derivatives in the inner product via integration by parts:

$$\langle y', \varphi' \rangle_{L_2} = y'(x)\overline{\varphi(x)}|_{x=0}^1 - \int_{\mathbb{T}} y'' \overline{\varphi} \,\mathrm{d}\varrho_T = -\langle y'', \varphi \rangle_{L_2}\,.$$

The right-hand side of the above equation defines an inner product of the form $\langle y, \varphi\rangle_{H^s} := \langle y, \varphi\rangle_{L_2} + \langle y^{(s)}, \varphi^{(s)}\rangle_{L_2}$ with a corresponding norm and Hilbert space

$$H^s(\mathbb{T}) := \left\{ f \in L_2(\mathbb{T}) : \|f\|^2_{H^s} = \|f\|^2_{L_2} + \|f^{(s)}\|^2_{L_2} < \infty \right\},$$

which we call **Sobolev space** of smoothness s on the torus $\mathbb{T}$. Considering the embedding $I_K\colon H^s(\mathbb{T}) \to L_2(\mathbb{T})$ from (3.1) and its adjoint $S_K = I_K^*$, we have

$$\sigma^2 \langle y, \varphi\rangle_{H^s} = \langle y, \varphi\rangle_{L_2} = \langle S_K I_K y, \varphi\rangle_{H^s}\,.$$

Thus, every eigenfunction of $S_K I_K$ is a weak solution of (3.6). Since eigenfunctions of a symmetric operator corresponding to different eigenvalues have to be orthogonal, we obtain eigenpairs forming a basis in $H^s(\mathbb{T})$:

Theorem 3.22. *The Sobolev space $H^s(\mathbb{T})$ with $s \in \mathbb{N}$ has the orthonormal basis $\{e_1, e_2, \dots\} = \{\sigma_1\eta_1, \sigma_2\eta_2, \dots\}$ where*

$$\sigma_k^2 = (1 + (2\pi\lfloor k/2\rfloor)^{2s})^{-1} \quad \textit{and} \quad \eta_k = \exp(2\pi \mathrm{i}(-1)^k \lfloor k/2\rfloor \cdot)\,.$$

Further $\{\eta_k\}_{k=1}^\infty$ is an orthonormal basis in $L_2(\mathbb{T})$.

Proof. **Step 1.** Showing $\{e_k\}_{k=1}^\infty$ is a basis $H^s(\mathbb{T})$. We consider the embedding operator $I_K\colon H^s(\mathbb{T}) \to L_2(\mathbb{T}), f \mapsto [f]_\sim$ and its adjoint $S_K = I_K^*$. The operator $S_K I_K\colon H^s(\mathbb{T}) \to H^s(\mathbb{T})$ is surjective, since for $e \in H^s(\mathbb{T}) \setminus \{0\}$ we have

$$\langle S_K I_K e, e\rangle_{H^s} = \langle I_K e, I_K e\rangle_{L_2} = \|e\|^2_{L_2} \neq 0\,.$$

Thus, computing all eigenpairs of $S_K I_K$ we obtain a basis of $H^s(\mathbb{T})$.

Step 2. With the preceding discussion it is left to find the orthonormal eigenfunctions. The functions $\{\eta_k\}_{k=1}^\infty$ are the well-known Fourier orthonormal basis of L_2. The functions $\{e_k\}_{k=1}^\infty$ inherit the orthonormality using the eigen property:

$$\langle \eta_k, \eta_\ell\rangle_{L_2} = \langle S_K I_K \eta_k, \eta_\ell\rangle_{H^s} = \sigma_k^2 \langle \eta_k, \eta_\ell\rangle_{H^s} = \langle e_k, e_\ell\rangle_{H^s}\,. \qquad \blacksquare$$

So far we defined the Sobolev smoothness for integer smoothness s. This concept extends as follows for arbitrary $s \geq 0$.

Theorem 3.23. *Let $s \geq 0$ and*

$$K_s(x, y) := \sum_{k=1}^\infty \frac{\exp(2\pi \mathrm{i}(-1)^k \lfloor k/2\rfloor (x - y))}{1 + (2\pi\lfloor k/2\rfloor)^{2s}}\,.$$

Then K_s has finite trace for $s > 1/2$. Furthermore, if $s \in \mathbb{N}$ the corresponding RKHS $H(K_s)$ coincides with $H^s(\mathbb{T})$.

Proof. For σ_k and η_k as in Theorem 3.22 the given kernel K_s is of the form $K_s(x,y) = \sum_{k=1}^{\infty} \sigma_k^2 \overline{\eta_k(y)} \eta_k(x)$. The finite trace statement for $s > 1/2$ follows immediately from the $L_2(\mathbb{T})$-orthonomality of η_k.

With the given Mercer representation we apply Theorem 3.11 and obtain $H(K_s)$ with orthonormal basis $\{\sigma_k \eta_k\}_{k=1}^{\infty}$. For $s \in \mathbb{N}$ this orthonormal basis coincides with the orthonormal basis $\{e_k\}_{k=1}^{\infty}$ of $H^s(\mathbb{T})$ from Theorem 3.22 and, thus, $H(K_s) = H^s(\mathbb{T})$. ■

The coefficients $\langle f, \exp(2\pi \mathrm{i}(-1)^k \lfloor k/2 \rfloor \cdot)\rangle_{L_2}$ are called **Fourier coefficients**, which are the same independent of the smoothness s of the Sobolev space. As a result we obtain a characterization of the Sobolev smoothness in terms of the decay of the Fourier coefficients:

Corollary 3.24. *Let $s \geq 0$. Then $f \in H^s(\mathbb{T})$ if and only if*

$$\sum_{k=1}^{\infty} k^{2s} \left| \left\langle f, \exp\left(2\pi \mathrm{i}(-1)^k \left\lfloor \frac{k}{2} \right\rfloor \cdot \right) \right\rangle_{L_2} \right|^2 < \infty .$$

Proof. By Theorem 3.11 and Theorem 3.23 we have the orthonormal basis $\{e_1, e_2, \dots\} = \{\sigma_1 \eta_1, \sigma_2 \eta_2, \dots\}$ in $H^s(\mathbb{T})$ where $\{\eta_1, \eta_2, \dots\}$ form an orthonormal basis in $L_2(\mathbb{T})$ with $\sigma_k = (1 + (2\pi \lfloor k/2 \rfloor)^{2s})^{-1/2}$ and $\eta_k = \exp(2\pi \mathrm{i}(-1)^k \lfloor k/2 \rfloor \cdot)$. By Parseval's equality, we have $f \in H^s(\mathbb{T})$ if and only if

$$\|f\|_{H^s}^2 = \sum_{k=1}^{\infty} |\langle f, e_k \rangle_{H^s}|^2 < \infty .$$

We know that σ_k^2 and e_k is an eigenpair of $S_K I_K$ with $I_K \colon H^s(\mathbb{T}) \to L_2(\mathbb{T}), f \mapsto [f]_\sim$ and its adjoint $S_K = I_K^*$. Thus,

$$\begin{aligned} \sum_{k=1}^{\infty} |\langle f, e_k \rangle_{H^s}|^2 &= \sum_{k=1}^{\infty} |\langle f, \sigma_k^{-2} S_K I_K e_k \rangle_{H^s}|^2 \\ &= \sum_{k=1}^{\infty} \sigma_k^{-2} |\langle [f]_\sim, [\eta_k]_\sim \rangle_{L_2(\mathbb{T})}|^2 , \end{aligned}$$

where $e_k = \sigma_k \eta_k$ was used in the last equality. We obtain the assertion by plugging in $\eta_k = \exp(2\pi \mathrm{i}(-1)^k \lfloor k/2 \rfloor \cdot)$. ■

3.5.2 Sobolev spaces on the interval $[0,1]$

Another example domain we consider is the unit interval $D = [0,1]$. It is often needed in applications. Even though it only differs from the torus $\mathbb{T}$ at the border, it yields to some interesting effects in the analysis.

Similar to the torus $\mathbb{T}$, where we gave a basis for the Sobolev spaces $H^s(\mathbb{T})$ in Theorem 3.22, we do the same here in Theorems 3.26 and 3.27 for smoothness $s = 1, 2$. With the approach from Section 3.5.1, the basis was independent of the smoothness and the spaces could be characterized by the decay of the Fourier coefficients of a function, cf. Theorem 3.24. Using the same approach on the unit interval, the basis is not independent of the smoothness and has to be computed for different smoothnesses separately. We use the function $f(x) = x$ as a counter example. It has infinite smoothness, i.e., $f \in H^s([0,1])$ for any $s \in \mathbb{N}$, but the decay of the coefficients with respect to the $H^1([0,1])$-basis is bounded by $3/2$, cf. Theorem 3.34.

We start by considering the same differential operator as in Section 3.5.1, namely

$$y^{(2s)} = (-1)^s \frac{1-\sigma^2}{\sigma^2} y \tag{3.9}$$

To work with the weak formulation we will use again a integration by parts trick, which now involves boundary terms:

Lemma 3.25. *For $f, g \in L_2([0,1])$ with $2s$ weak derivatives, it holds*

$$\begin{aligned}&\langle f^{(s)}, g^{(s)} \rangle_{L_2}\\ &\quad = \sum_{l=0}^{s-1} (-1)^l [f^{(s-1+l)}(x) g^{(s-1-l)}(x)]_{x=0}^1 + (-1)^s \langle f^{(2s)}, g \rangle_{L_2} .\end{aligned}$$

Proof. We proof the statement via induction. The base case $s = 0$ is clear. For the induction step, $s \to s+1$ we have by integration by parts

$$\langle f^{(s+1)}, g^{(s+1)} \rangle_{L_2} = [f^{(s)}(x) g^{(s)}(x)]_{x=0}^1 - \langle f^{(s+2)}, g^{(s)} \rangle_{L_2} .$$

Using the induction hypothesis for s, we obtain

$$\begin{aligned}&\langle f^{(s+1)}, g^{(s+1)} \rangle_{L_2}\\ &\quad = [f^{(s)}(x) g^{(s)}(x)]_{x=0}^1 - \Big(\sum_{l=0}^{s-1} (-1)^l [f^{(s+1+l)}(x) g^{(s-1-l)}(x)]_{x=0}^1\\ &\qquad + (-1)^s \langle f^{(2s+2)}, g \rangle_{L_2} \Big) .\end{aligned}$$

It is left to shift the sum to obtain the assertion

$$\begin{aligned}
&\langle f^{(s+1)}, g^{(s+1)}\rangle_{L_2} \\
&\quad = [f^{(s)}(x)g^{(s)}(x)]_{x=0}^{1} - \sum_{l=1}^{s}(-1)^l [f^{(s+l)}(x)g^{(s-l)}(x)]_{x=0}^{1} \\
&\qquad + (-1)^{s+1}\langle f^{(2s+2)}, g\rangle_{L_2} \\
&\quad = \sum_{l=0}^{s}(-1)^l [f^{((s+1)-1+l)}(x)g^{((s+1)-1-l)}(x)]_{x=0}^{1} \\
&\qquad + (-1)^{s+1}\langle f^{(2(s+1))}, g\rangle_{L_2}\,. \qquad \blacksquare
\end{aligned}$$

With that, we have a look at the weak formulation of (3.9). For $\varphi \in C_0^\infty([0,1])$ we obtain

$$\begin{aligned}
& \langle y^{(2s)}, \varphi\rangle_{L_2} = \left\langle (-1)^s \frac{1-\sigma^2}{\sigma^2} y, \varphi \right\rangle_{L_2} \\
\Leftrightarrow \quad & \langle y, \varphi\rangle_{L_2} = \sigma^2(\langle y, \varphi\rangle_{L_2} + (-1)^s \langle y^{(2s)}, \varphi\rangle_{L_2}) \\
& \qquad = \sigma^2(\langle y, \varphi\rangle_{L_2} + \langle y^{(s)}, \varphi^{(s)}\rangle_{L_2}) \\
& \qquad\quad - \sigma^s \sum_{l=0}^{s-1}(-1)^l [y^{(s-1+l)}(x)\varphi^{(s-1-l)}(x)]_{x=0}^{1}\,.
\end{aligned}$$

Where the boundary terms vanish since φ vanishes at the border. The above right-hand side defines an inner product $\langle y, \varphi\rangle_{H^s} := \langle y, \varphi\rangle_{L_2} + \langle y^{(s)}, \varphi^{(s)}\rangle_{L_2}$ with a corresponding norm and Hilbert space

$$H^s([0,1]) := \left\{ f \in L_2([0,1]) : \|f\|_{H^s}^2 = \|f\|_{L_2}^2 + \|f^{(s)}\|_{L_2}^2 < \infty \right\},$$

which we call **Sobolev space** of smoothness s on the interval $[0,1]$. It is also known as unanchored Sobolev space as it does not fix boundary conditions, cf. [NW08, Section A.2.1] for $s = 1$ or [NS23, Definition 1].

In the quest to compute a basis for H^s, we have the $2s$ solutions as in (3.7) for the differential equation. By the Sobolev embedding Theorem [AF03, Theorem 4.12] we have for $s \in \mathbb{N}$ and $\varepsilon > 0$ that $H^{s+1/2+\varepsilon}([0,1]) \hookrightarrow C^s([0,1])$ (space of s times continuously differentiable functions). Thus, boundary conditions on the first $s-1$ derivatives would also alter the function space. These are known as anchored Sobolev spaces and are smaller, cf. [NW08, Section A.2.2] or [KSWW09, Example 4.2]. Boundary conditions on higher derivatives do not affect the resulting space as they only act on a set of measure zero. We will use the $2s$ conditions

$$y^{(s)}(0) = y^{(s)}(1) = \cdots = y^{(2s-1)}(0) = y^{(2s-1)}(1) = 0\,.$$

Next, we compute the eigenbases for $H^1([0,1])$ and $H^2([0,1])$ similar to what we did for $H^s(\mathbb{T})$ in Theorem 3.22.

Theorem 3.26. *The Sobolev space $H^1([0,1])$ has the orthonormal basis $\{e_k\}_{k=1}^{\infty} = \{\sigma_k\eta_k\}_{k=1}^{\infty}$ where*

$$\sigma_k^2 = \frac{1}{1+\pi^2(k-1)^2} \quad \textit{and} \quad \eta_k(x) = \begin{cases} 1 & : k = 1 \\ \sqrt{2}\cos(\pi(k-1)x) & : k \geq 2\,. \end{cases}$$

Further $\{\eta_k\}_{k=1}^{\infty}$ is an orthonormal system in $L_2([0,1])$.

Proof. **Step 1.** Showing $\{e_k\}_{k=1}^{\infty}$ is a basis of $H^s([0,1])$. With the above discussion, the eigenfunctions e_k certainly fulfill the following differential equation

$$e_k = \frac{\sigma_k^2}{1-\sigma_k^2}e_k'' \quad \text{with} \quad e_k'(0) = e_k'(1) = 0\,.$$

The proposed functions are exactly the ones fulfilling the above.

Step 2. Showing the orthonormality. The $L_2([0,1])$-orthonormality of η_k is easy to verify. The $H^1([0,1])$-orthonormality of e_k follows analogously to Theorem 3.22. ■

The $H^1([0,1])$ basis above was already considered in [WW09, Lemma 4.1] with the same proof technique, in [IN08] as a modified Fourier expansion, and in [SNC16] in the context of samples along tent-transformed rank-1 lattices, and further in [Adc10b, AH11, DNP14, CKNS16, KMNN21]. The natural question arises of Sobolev spaces on the unit interval for higher smoothness. The following $H^2([0,1])$ basis was already posed in [AIN12, Section 3], where higher-order Sobolev-spaces are found as well.

Theorem 3.27. *The Sobolev space $H^2([0,1])$ has the orthonormal basis $\{e_k\}_{k=1}^{\infty} = \{\sigma_k\eta_k\}_{k=1}^{\infty}$ where*

$$\sigma_1 = \sigma_2 = 1, \quad \eta_1(x) = 1\,, \quad \textit{and} \quad \eta_2(x) = 2\sqrt{3}x - \sqrt{3}$$

and for $k \geq 3$, $\sigma_k^2 = \frac{1}{1+t_k^4}$ with $t_k > 0$ the solutions of $\cosh(t_k)\cos(t_k) = 1$ *($t_k \approx \frac{2k-3}{2}\pi$, cf. Theorem 3.29) and*

$$\eta_k(x) = \cosh(t_kx)+\cos(t_kx)-\frac{\cosh(t_k)-\cos(t_k)}{\sinh(t_k)-\sin(t_k)}(\sinh(t_kx)+\sin(t_kx))\,.$$

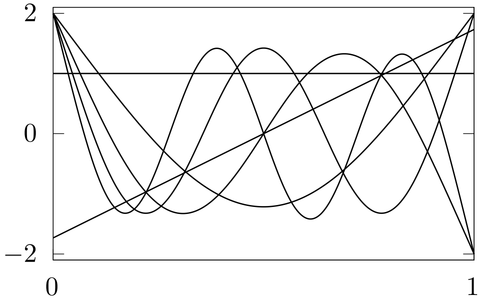


Figure 3.1: First six basis functions of $H^2([0,1])$.

Further, $\{\eta_k\}_{k=1}^{\infty}$ *is an orthonormal system in* $L_2([0,1])$ *and*

$$\|\eta_k\|_\infty \le \begin{cases} 1 & \textit{for } k = 1 \\ \sqrt{3} & \textit{for } k = 2 \\ \sqrt{6} & \textit{for } k \ge 3\,. \end{cases}$$

The kernel for $H^1([0,1])$ and $H^2([0,1])$ itself is given by the series representation according to Theorem 3.11. Before proving the above, we want to discuss the basis in terms of its approximation properties and numerical stability. The singular values σ_k for $H^2([0,1])$ decay quadratically in contrast to linearly for $H^1([0,1])$, giving better approximation properties, cf. Theorem 3.12. The first $H^2([0,1])$ basis functions are depicted in Figure 3.1.

However, as cosh and sinh both grow exponentially, the representation of the $H^2([0,1])$ basis in Theorem 3.27 is prone to cancellations and, therefore, numerical unstable. In the next theorem we pose an approximation which is numerically stable and is published in [Bar23, Theorem 4.3].

Theorem 3.28. *For* $0 < t_3 < t_4 < \dots$ *fulfilling* $\cosh(t_k)\cos(t_k) = 1$, *let* η_k *be as in Theorem 3.27. Further, for* $k \ge 3$, *let* $\tilde{t}_k = \pi(2k-1)/2$ *and*

$$\begin{aligned}\tilde{\eta}_k(x) &= \sqrt{2}\cos\Big(\tilde{t}_k x + \pi/4\Big) \\ &\quad + \mathbb{1}_{[0,1/2]}(x)\exp\Big(-\tilde{t}_k x\Big) + \mathbb{1}_{[1/2,1]}(x)(-1)^k\exp\Big(-\tilde{t}_k(1-x)\Big)\,.\end{aligned}$$

Then $|\eta_k(x) - \tilde{\eta}_k(x)| \le \varepsilon$ *for* $k \ge \frac{2}{\pi}\log(18/\varepsilon) + 2$. *In particular, the approximation* $\tilde{\eta}_k$ *is exact up to machine precision* $\varepsilon = 10^{-16}$ *for* $k \ge 28$.

In the following we proof Theorems 3.27 and 3.28. We suggest to skip to page 56 in a first reading because of the technical nature of the proofs.

Proof of the first part of Theorem 3.27. **Step 1.** Showing $\{\eta_k\}_{k=1}^{\infty}$ is a basis for $H^s([0,1])$. Analogously to Theorem 3.26, for σ_k^2, η_k an eigenpair of $S_K I_K$, we obtain the following differential equation

$$\eta_k = \frac{\sigma_k^2}{1-\sigma_k^2}\eta_k^{(4)} \quad \text{with} \quad \eta_k^{(2)}(0) = \eta_k^{(2)}(1) = \eta_k^{(3)}(0) = \eta_k^{(3)}(1) = 0\,.$$

Now we distinguish three cases for the value of σ_k^2:

First case. $\sigma_k^2 = 1$. The ansatz function becomes

$$\eta_k(x) = A + Bx + Cx^2 + Dx^3.$$

From the conditions $\eta_k^{(2)}(0) = \eta_k^{(3)}(0) = 0$ we obtain $D = C = 0$. The two remaining degrees of freedom are restricted by demanding $L_2([0,1])$-orthonormality. By simple calculus we obtain the proposed eigenfunctions η_1 and η_2.

Second case. $\sigma_k^2 > 1$. Set $t_k := \sqrt[4]{(\sigma_k^2-1)/(\sigma_k^2)}$. The ansatz becomes

$$\begin{aligned}\eta_k(x) = {} & A\cosh(t_k x)\cos(t_k x) + B\cosh(t_k x)\sin(t_k x)\\ & + C\sinh(t_k x)\cos(t_k x) + D\sinh(t_k x)\sin(t_k x).\end{aligned}$$

The conditions $\eta_k^{(2)}(0) = \eta_k^{(3)}(0) = 0$ transform to $D = 0$ and $B = C$. The two remaining degrees of freedom are fixed by the conditions $\eta_k^{(2)}(1) = \eta_k^{(3)}(1) = 0$ which, in matrix form, look as follows

$$\begin{bmatrix} -\sinh(t_k)\sin(t_k) & \sinh(t_k)\cos(t_k) - \cosh(t_k)\sin(t_k) \\ -\sinh(t_k)\cos(t_k) - \cosh(t_k)\sin(t_k) & -2\sinh(t_k)\sin(t_k) \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \mathbf{0}\,.$$

For a non-trivial solution we need this matrix to be non-regular. To achieve that we have a look at the roots of its determinant:

$$2\sinh^2(t_k)\sin^2(t_k) + \sinh^2(t_k)\cos^2(t_k) - \cosh^2(t_k)\sin^2(t_k) \overset{!}{=} 0.$$

Using $\sin^2(t_k) + \cos^2(t_k) = \cosh^2(t_k) - \sinh^2(t_k) = 1$ we have

$$\sinh^2(t_k) - \sin^2(t_k) = \frac{1}{2}\cosh(2t_k) + \frac{1}{2}\cos(2t_k) - 1 \overset{!}{=} 0$$

which is only fulfilled for $t_k = 0$, or equivalently, $\sigma_k^2 = 1$. Hence, there are no eigenvalues bigger than 1.

Third case. $\sigma_k^2/(1-\sigma_k^2) > 0 \Leftrightarrow \sigma_k^2 < 1$. Introducing the numbers $t_k := \sqrt[4]{(1-\sigma_k^2)/\sigma_k^2}$, we use the ansatz

$$\eta_k(x) = A\cos(t_k x) + B\sin(t_k x) + C\cosh(t_k x) + D\sinh(t_k x).$$

The conditions $\eta_k^{(2)}(0) = \eta_k^{(3)}(0) = 0$ transform to $A = C$ and $B = D$, respectively. The conditions $\eta_k^{(2)}(1) = \eta_k^{(3)}(1) = 0$ can be put into a system of equations:

$$\begin{bmatrix} \cosh(t_k) - \cos(t_k) & \sinh(t_k) - \sin(t_k) \\ \sinh(t_k) + \sin(t_k) & \cosh(t_k) - \cos(t_k) \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \mathbf{0}$$

or, by using $\cosh^2(t_k) - \sinh^2(t_k) = \cos^2(t_k) + \sin^2(t_k) = 1$, equivalently

$$\begin{bmatrix} \cosh(t_k) - \cos(t_k) & \sinh(t_k) - \sin(t_k) \\ 0 & 1 - \cosh(t_k)\cos(t_k) \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \mathbf{0}.$$

For non-trivial solutions we need non-regularity of that matrix which transforms to the condition $\cosh(t_k)\cos(t_k) = 1$. With the leftover degree of freedom we choose

$$A = C = 1 \quad \text{and} \quad B = D = -\frac{\cosh(t_k) - \cos(t_k)}{\sinh(t_k) - \sin(t_k)}$$

and obtain η_k for $k \geq 3$ as proposed in the theorem.

Step 2. Showing the orthonormality. From the eigendecomposition in Step 1 follows the compactness of $S_K I_K$, cf. [Con90, Theorem 4.4]. By the spectral Theorem 3.13 we know that the eigenspaces of different eigenvalues are orthogonal. For $\sigma_1 = \sigma_2 = 1$ we have orthonormality by construction. All other eigenvalues are different and, thus, the corresponding eigenfunctions η_k are orthogonal. If we show the $L_2([0,1])$-normalization of η_k we obtain $L_2([0,1])$-orthonormality and the $H^2([0,1])$-orthonormality of e_k follows analogously to Theorem 3.22.

For the $L_2([0,1])$-norm we obtain

$$\begin{aligned}
\int_0^1 |\eta_k|^2 \,\mathrm{d}x &= \int_0^1 (\cosh(t_k x) + \cos(t_k x))^2 \,\mathrm{d}x \\
&\quad + B^2 \int_0^1 (\sinh(t_k x) + \sin(t_k x))^2 \,\mathrm{d}x \\
&\quad + 2B \int_0^1 (\cosh(t_k x) + \cos(t_k x))(\sinh(t_k x) + \sin(t_k x)) \,\mathrm{d}x \\
&= 1 + \frac{\sin(2t_k) + \sinh(2t_k) + 4\cos(t_k)\sinh(t_k) + 4\sin(t_k)\cosh(t_k)}{4t_k} \\
&\quad + B^2 \frac{-\sin(2t_k) + \sinh(2t_k) - 4\cos(t_k)\sinh(t_k) + 4\sin(t_k)\cosh(t_k)}{4t_k} \\
&\quad + 2B \frac{(\sin(t_k) + \sinh(t_k))^2}{2t_k} \\
&= 1 + \frac{1+B^2}{4t_k}(\sinh(2t_k) + 4\sin(t_k)\cosh(t_k)) \\
&\quad + \frac{1-B^2}{4t_k}(\sin(2t_k) + 4\cos(t_k)\sinh(t_k)) + B\frac{(\sin(t_k) + \sinh(t_k))^2}{t_k} \\
&= 1 + \frac{1+B^2}{2t_k}\cosh(t_k)(\sinh(t_k) + 2\sin(t_k)) \\
&\quad + \frac{1-B^2}{2t_k}\cos(t_k)(\sin(t_k) + 2\sinh(t_k)) + B\frac{(\sin(t_k) + \sinh(t_k))^2}{t_k} \,.
\end{aligned}$$

Using $\cos(t_k)\cosh(t_k) = 1$, we have

$$1 + B^2 = 2\frac{\sinh(t_k)}{\sinh(t_k) - \sin(t_k)} \quad \text{and} \quad 1 - B^2 = -2\frac{\sin(t_k)}{\sinh(t_k) - \sin(t_k)} \,. \tag{3.10}$$

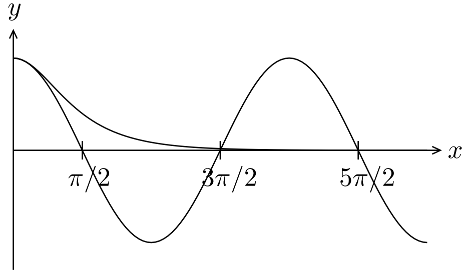


Figure 3.2: $\cos(t)$ and $1/\cosh(t)$

Thus,

$$\begin{aligned}\int_0^1 |\eta_k|^2 \,\mathrm{d}x = 1 + \frac{1}{t_k(\sinh(t_k) - \sin(t_k))}\Big(&\\ \sinh(t_k)\cosh(t_k)(\sinh(t_k) + 2\sin(t_k))&\\ - \sin(t_k)\cos(t_k)(\sin(t_k) + 2\sinh(t_k))&\\ - (\cosh(t_k) - \cos(t_k))(\sin(t_k) + \sinh(t_k))^2\Big)&\\ = 1 + \frac{\cos(t_k)\sinh^2(t_k) - \cosh(t_k)\sin^2(t_k)}{t_k(\sinh(t_k) - \sin(t_k))}&\\ = 1 + \frac{\cos(t_k)\cosh^2(t_k) - \cos(t_k) - \cosh(t_k) + \cosh(t_k)\cos^2(t_k)}{t_k(\sinh(t_k) - \sin(t_k))}&,\end{aligned}$$

where $\cos^2(t_k) + \sin^2(t_k) = \cosh^2(t_k) - \sinh^2(t_k) = 1$ was used in the last equality. Using $\cosh(t_k)\cos(t_k) = 1$, the latter summand evaluates to zero and we have proven the $L_2([0,1])$-normality. ■

Lemma 3.29. *For* $0 < t_3 < t_4 < \ldots$ *fulfilling* $\cosh(t_k)\cos(t_k) = 1$ *and* $\tilde{t}_k = \frac{2k-3}{2}\pi$*, we have*

$$\frac{3}{2}\pi < t_3 \quad \textit{and} \quad \left|\tilde{t}_k - t_k\right| \le \varepsilon$$

for $k \ge \frac{1}{\pi}\log(\pi/\varepsilon) + 2$*. In particular* $|\tilde{t}_k - t_k| \le \pi\exp(-2\pi)$ *for all* $k \ge 3$.

Proof. Since $0 < 1/\cosh(t) < 1$ for $t > 0$ and the oscillating behavior of

$\cos(t)$, as depicted in Figure 3.2, we obtain

$$t_k \in \begin{cases} \left(\frac{2k-3}{2}\pi, \frac{2k-2}{2}\pi\right) & \text{for } k \text{ even} \\ \left(\frac{2k-4}{2}\pi, \frac{2k-3}{2}\pi\right) & \text{for } k \text{ odd}\,. \end{cases}$$

In particular, $\frac{3}{2}\pi < t_3$. Furthermore, for even k and $t \in \left(\frac{2k-3}{2}\pi, \frac{2k-2}{2}\pi\right)$ we have

$$\frac{1}{\cosh(t)} \le 2\exp(-t) \le 2\exp\Big(-\frac{2k-3}{2}\pi\Big) \quad \text{and} \quad \cos(t) \ge \frac{t - \frac{2k-3}{2}\pi}{\pi/2}.$$

The function bounds intersect for a value larger than t_k, which we use to refine the interval:

$$t_k \in \Big(\tilde{t}_k, \tilde{t}_k + \pi\exp\Big(-\frac{2k-3}{2}\pi\Big)\Big).$$

Similarly, for odd k and $t \in \left(\frac{2k-4}{2}\pi, \frac{2k-3}{2}\pi\right)$ we obtain

$$t_k \in \Big(\tilde{t}_k - \pi\exp\Big(-\frac{2k-4}{2}\pi\Big), \tilde{t}_k\Big).$$

Thus, for $k \ge 3$ we have $\left|\tilde{t}_k - t_k\right| \le \pi\exp(-(k-2)\pi)$, which is smaller than ε for $k \ge \log(\pi/\varepsilon)/\pi + 2$. ∎

Lemma 3.30. *For* $0 < t_3 < t_4 < \dots$ *fulfilling* $\cosh(t_k)\cos(t_k) = 1$, *we have that* $\eta^{\mathrm{I}}_{t_k}$ *defined by*

$$\eta^{\mathrm{I}}_{t_k}(x) := \cosh(t_k x) - \frac{\cosh(t_k) - \cos(t_k)}{\sinh(t_k) - \sin(t_k)}\sinh(t_k x) \tag{3.11}$$

is convex and non-negative for all odd k and monotone for all even k.

Proof. **Step 1.** We distinguish for different values of $B = B(t) := (\cosh(t) - \cos(t))/(\sinh(t) - \sin(t))$. For $B < 1$ we have

$$\eta^{\mathrm{I}}_t(x) = \cosh(tx) - B\sinh(tx) \ge \cosh(tx) - \sinh(tx) \ge 0$$

and by the same argument $(\eta^{\mathrm{I}}_t(x))^{(2)} = t^2\eta^{\mathrm{I}}_t(x) \ge 0$ for all $x \ge 0$. Thus, $\eta^{\mathrm{I}}_t(x)$ is convex and non-negative.

For $B > 1$ we obtain

$$(\eta^{\mathrm{I}}_t)' = t(\sinh(tx) - B\cosh(tx)) \le t(\sinh(tx) - \cosh(tx)) \le 0$$

for all $x \geq 0$. Thus, η_t^{I} is monotone.

Step 2. It is left to show for which k's $B(t_k)$ attains a value smaller or bigger than one:

$$\begin{aligned} B(t_k) \lesseqgtr 1 \quad &\Leftrightarrow \quad \cosh(t_k) - \cos(t_k) \lesseqgtr \sinh(t_k) - \sin(t_k) \\ &\Leftrightarrow \quad \exp(-t_k) - \sqrt{2}\cos(t_k + \pi/4) \lesseqgtr 0. \end{aligned}$$

We will show that $\exp(-t_k) - \sqrt{2}\cos(t_k + \pi/4)$ has the same sign as $(-1)^k$ and, thus, are finished. We do this by estimating their difference by a quantity smaller than one. With $\tilde{t}_k = \frac{2k-3}{2}\pi$ we obtain

$$\begin{aligned} &|\exp(-t_k) - \sqrt{2}\cos(t_k + \pi/4) - (-1)^k| \\ &\quad = |\exp(-t_k) - \sqrt{2}\cos(t_k + \pi/4) + \sqrt{2}\cos(\tilde{t}_k + \pi/4)|. \end{aligned}$$

Using that cos is Lipschitz-continuous with constant 1 and Theorem 3.29 we estimate the above by

$$\begin{aligned} &|\exp(-t_k) - \sqrt{2}\cos(t_k + \pi/4) - (-1)^k| \leq |\exp(-t_k)| + \sqrt{2}|t_k - \tilde{t}_k| \\ &\quad \leq \exp(-3/2\pi) + \sqrt{2}\pi\exp(-2\pi), \end{aligned}$$

which is certainly smaller than one. ■

Lemma 3.31. *For* $0 < t_3 < t_4 < \dots$ *fulfilling* $\cosh(t_k)\cos(t_k) = 1$, *we have that* $\eta_{t_k}^{\mathrm{I}}$ *defined in* (3.11) *is even with respect to the axis* $x = 1/2$ *for all odd* k *and vice versa.*

Proof. **Step 1.** We will show that $\eta_{t_k}^{\mathrm{I}}$ has any symmetry around $x = 1/2$. We shift the function and split it into an odd and an even part. For $B = B(t) = (\cosh(t) - \cos(t))/(\sinh(t) - \sin(t))$, we obtain

$$\begin{aligned} \eta_t^{\mathrm{I}}(x + 1/2) &= \cosh(tx + t/2) - B\sinh(tx + t/2) \\ &= \underbrace{(\cosh(t/2) - B\sinh(t/2))}_{=:\alpha}\cosh(tx) \\ &\quad + \underbrace{(\sinh(t/2) - B\cosh(t/2))}_{=:\beta}\sinh(tx). \end{aligned}$$

Multiplying the two factors α and β in front of $\cosh(tx)$ and $\sinh(tx)$, we obtain

$$\begin{aligned}\alpha\cdot\beta &= -B\cosh^2(t/2) - B\sinh^2(t/2) + (1+B^2)\cosh(t/2)\sinh(t/2)\\ &= -B\frac{\cosh(t)-1}{2} - B\frac{\cosh(t)+1}{2} + (1+B^2)\frac{\sinh(t)}{2}\\ &= -B\cosh(t) + (1+B^2)\frac{\sinh(t)}{2}\,.\end{aligned}$$

Using (3.10), $\cosh(t)\cos(t) = 1$, and $1 = \cosh^2(t) - \sinh^2(t)$ this evaluates to

$$\alpha\cdot\beta = -\frac{\cosh^2(t)-1}{\sinh(t)-\sin(t)} + \frac{\sinh^2(t)}{\sinh(t)-\sin(t)} = 0\,.$$

Since we are not dealing with the zero function, either α or β is zero. Thus, $x \mapsto \eta_t^{\mathrm{I}}(x+1/2)$ obeys a symmetry.

Step 2. It remains to specify the kind of symmetry. By Theorem 3.30 we have that η_t^{I} is convex for odd k. Since a convex non-constant function cannot be odd it has to be even. Also by Theorem 3.30 we have that η_t^{I} is monotone for even k. Since a monotone non-zero function cannot be even it has to be odd. ■

Proof of the second part of Theorem 3.27. The cases $k = 1, 2$ are clear. For $k \geq 3$ we split the function into η_t^{I} defined in (3.11) and

$$\eta_t^{\mathrm{II}}(x) := \cos(tx) - \frac{\cosh(t)-\cos(t)}{\sinh(t)-\sin(t)}\sin(tx).$$

We will show that each of these is bounded by $1.01\sqrt{2}$ and, thus, obtain the assertion.

Step 1. In order to bound $\eta_{t_k}^{\mathrm{I}}$ we firstly have a look at the boundary points $x \in \{0,1\}$. With Theorem 3.31 we obtain

$$\eta_{t_k}^{\mathrm{I}}(0) = \left|\eta_{t_k}^{\mathrm{I}}(1)\right| = 1. \tag{3.12}$$

By Theorem 3.30 $\eta_{t_k}^{\mathrm{I}}$ is either non-negative and convex or monotone and, thus, cannot exceed its values on the boundary.

Step 2. In order to bound $\eta_{t_k}^{\mathrm{II}}$ we define

$$B := \frac{\cosh(t_k)-\cos(t_k)}{\sinh(t_k)-\sin(t_k)} \quad \text{and} \quad \vartheta = \arg(1+Bi).$$

Next, we use the exponential definition of sine and cosine and the polar representation of complex numbers to obtain

$$\begin{aligned}\eta^{\mathrm{II}}_{t_k}(x) &= \cos(t_k x) - B\sin(t_k x)\\ &= \frac{\exp(\mathrm{i}t_k x) + \exp(-\mathrm{i}t_k x)}{2} + Bi\frac{\exp(\mathrm{i}t_k x) - \exp(-\mathrm{i}t_k x)}{2}\\ &= \frac{(1+Bi)\exp(\mathrm{i}t_k x) + (1-Bi)\exp(-\mathrm{i}t_k x)}{2}\\ &= \sqrt{1+B^2}\,\frac{\exp(\mathrm{i}(t_k x + \vartheta)) + \exp(-\mathrm{i}(t_k x + \vartheta))}{2}\\ &= \sqrt{1+B^2}\cos(t_k x + \vartheta)\,.\end{aligned}$$

Thus, by (3.10)

$$|\eta^{\mathrm{II}}_{t_k}(x)| \le \sqrt{1+B^2} = \sqrt{\frac{2}{1-\sin(t_k)/\sinh(t_k)}} \le \sqrt{\frac{2}{1-1/\sinh(t_k)}} \tag{3.13}$$

From Theorem 3.29 we use $t_k \ge 3/2\pi$ in combination with the monotonicity in (3.13) we have $|\eta^{\mathrm{II}}_{t_k}(x)| \le 1.01\sqrt{2}$. ■

Lemma 3.32. *For $t \ge \max\{2\log(4/\varepsilon), 3/2\pi\}$ we have for $x \in [0,1/2]$*

$$\Big|\Big(1 - \frac{\cosh(t)-\cos(t)}{\sinh(t)-\sin(t)}\Big)\sinh(tx)\Big| \le \varepsilon\,.$$

Proof. We use $\cosh(t) - \sinh(t) = \exp(-t)$ and $\cos(t) - \sin(t) = \sqrt{2}\cos(t + \pi/4)$ to estimate

$$\begin{aligned}&\Big|\Big(1 - \frac{\cosh(t)-\cos(t)}{\sinh(t)-\sin(t)}\Big)\sinh(tx)\Big|\\ &\quad= |\sqrt{2}\cos(t+\pi/4) - \exp(-t)|\Big|\frac{\sinh(tx)}{\sinh(t)-\sin(t)}\Big|.\end{aligned}$$

Since we have $x \le 1/2$, sinh strictly monotone growing, and $t \ge 3/2\pi$ by Theorem 3.29, we further estimate

$$\begin{aligned}\Big|\Big(1 - \frac{\cosh(t)-\cos(t)}{\sinh(t)-\sin(t)}\Big)\sinh(tx)\Big| &\le 2\Big|\frac{\sinh(t/2)}{\sinh(t)-\sin(t)}\Big|\\ &= 2\Big|\frac{1}{2\cosh(t/2)}\,\frac{1}{1-\sin(t)/\sinh(t)}\Big|\,.\end{aligned}$$

Using $1 - \sin(t)/\sinh(t) > 1/2$ for $t > 3/2\pi$, we obtain

$$\left|\Big(1 - \frac{\cosh(t) - \cos(t)}{\sinh(t) - \sin(t)}\Big) \sinh(tx)\right| \leq \frac{2}{\cosh(t/2)} \leq \frac{4}{\exp(t/2)}\,,$$

which is smaller than ε for $t \geq 2\log(4/\varepsilon)$. ■

Lemma 3.33. *For* $0 < t_3 < t_4 < \ldots$ *fulfilling* $\cosh(t_k)\cos(t_k) = 1$, *we have*

$$\left|\eta^{\mathrm{II}}_{t_k}(x) - \sqrt{2}\cos(t_k x + \pi/4)\right| \leq \varepsilon \quad \textit{for} \quad x \in [0,1]$$

and

$$\left|\eta^{\mathrm{I}}_{t_k}(x) - \exp(-tx)\right| \leq \varepsilon \quad \textit{for} \quad x \in [0,1/2]$$

for $k \geq \frac{2}{\pi}\log(4/\varepsilon) + 2$.

Proof. **Step 1.** For the first inequality we use

$$\sqrt{2}\cos(tx + \pi/4) = \cos(tx) - \sin(tx)$$

to obtain

$$\left|\eta^{\mathrm{II}}_t - \sqrt{2}\cos(tx + \pi/4)\right| = \left|\Big(1 - \frac{\cosh(t) - \cos(t)}{\sinh(t) - \sin(t)}\Big) \sin(tx)\right|$$

which is smaller than ε for $t > \max\{2\log(4/\varepsilon), 3/2\pi\}$ by Theorem 3.32.

The second inequality follows analogously from $\exp(-tx) = \cosh(tx) - \sinh(tx)$ and Theorem 3.32.

Step 2. It is left to show the condition $t \geq \max\{2\log(4/\varepsilon), 3/2\pi\}$ from Step 1. By Theorem 3.29 we have $t_k \geq 3/2\pi$. Further, by assumption, we have

$$k \geq \frac{2}{\pi}\log\Big(\frac{4}{\varepsilon}\Big) + 2 \geq \frac{2}{\pi}\log\Big(\frac{4}{\varepsilon}\Big) + \exp(-2\pi) + \frac{3}{2}\,.$$

Thus,

$$2\log\Big(\frac{4}{\varepsilon}\Big) \leq \frac{2k-3}{2}\pi - \pi\exp(-2\pi) \leq t_k$$

where the last inequality follows from Theorem 3.29. ■

Proof of Theorem 3.28. Because of the symmetry shown in Theorem 3.31 we assume without loss of generality $x \in [0, 1/2]$. Then

$$\begin{aligned}|\eta_k(x) - \tilde{\eta}_k(x)| \leq{}& \left|\eta^{\mathrm{I}}_k(x) - \exp(-t_k x)\right| + \left|\eta^{\mathrm{II}}_k(x) - \sqrt{2}\cos(t_k x + \pi/4)\right| \\ &+ \left|\exp(-t_k x) - \exp(-\tilde{t}_k x)\right| + \sqrt{2}\left|\cos(t_k x + \pi/4) - \cos(\tilde{t}_k x + \pi/4)\right|.\end{aligned}$$

By Theorem 3.33, the first two summands are each smaller than $\varepsilon/4$ each for $k > \frac{2}{\pi}\log(16/\varepsilon)+2$. We estimate the two latter summands as follows.

Since cos is Lipschitz continuous with constant one we have

$$\sqrt{2}\Big|\cos(t_k x+\pi/4)-\cos\left(\tilde{t}_k x+\frac{\pi}{4}\right)\Big| \le \sqrt{2}\Big|t_k-\tilde{t}_k\Big|$$

which, by Theorem 3.29 is smaller or equal than $\varepsilon/4$ for $k \ge \frac{1}{\pi}\log(18/\varepsilon)+2$.

Since exp is Lipschitz continuous with constant 1 on $(-\infty,0)$, we have

$$\Big|\exp(-t_k x)-\exp(-\tilde{t}_k x)\Big| \le \Big|t_k-\tilde{t}_k\Big|$$

which, by Theorem 3.29 is smaller or equal than $\varepsilon/4$ for $k \ge \frac{1}{\pi}\log(16/\varepsilon)+2$.

Overall, we obtain $|\eta_k(x)-\tilde{\eta}_k(x)| < 4\frac{\varepsilon}{4}=\varepsilon$ for

$$\begin{aligned} k &\ge \frac{2}{\pi}\log(18/\varepsilon)+2 \\ &\ge \max\left\{\frac{2}{\pi}\log(16/\varepsilon)+2,\ \frac{1}{\pi}\log(18/\varepsilon)+2,\ \frac{1}{\pi}\log(16/\varepsilon)+2\right\}. \quad ■ \end{aligned}$$

As the bases for $H^1([0,1])$ and $H^2([0,1])$ from Theorem 3.26 and Theorem 3.27 differ, a characterization by the decay of the coefficients as in Theorem 3.24 is not apparent. It turns out to not be possible in this case: We consider the function $x \mapsto x$, which is in $H^s([0,1])$ for $s \in \mathbb{N}$. A basis for all $H^s([0,1])$ is in particular a basis in $H^1([0,1])$, which we already know from Theorem 3.26. The characterization by the coefficients with respect to the $H^1([0,1])$ basis yields a smaller function space as the following lemma shows.

Lemma 3.34. *Let $s \ge 0$ and $\tilde{H}^s([0,1])$ be the RKHS constructed from a kernel according to Theorem 3.11 with*

$$\sigma_k^2(s)=\frac{1}{1+\pi^2(k-1)^{2s}} \quad \textit{and} \quad \eta_k(x)=\begin{cases}1 & : k=1\\ \sqrt{2}\cos(\pi(k-1)x) & : k\ge 2\,.\end{cases}$$

Then $x \mapsto x \in \tilde{H}^s([0,1])$ if and only if $s<3/2$.

Proof. We have

$$\begin{aligned}\langle x,\sqrt{2}\cos(\pi)(k-1)x\rangle_{L_2} &= -\sqrt{2}\int_0^1 \frac{\sin(\pi(k-1)x)}{\pi(k-1)}\,\mathrm{d}x \\ &= \frac{\sqrt{2}((-1)^{k+1}-1)}{\pi^2(k-1)^2}\,.\end{aligned}$$

Thus,

$$\begin{aligned}\|x\|_{\tilde{H}^s} &= \sum_{k=1}^{\infty} |\langle x, e_k\rangle_{\tilde{H}^s}|^2 = \sum_{k=1}^{\infty} \sigma_k^2 |\langle x, \eta_k\rangle_{L_2}|^2 \\ &= \frac{1}{2} + \sum_{k=2}^{\infty} \frac{1+\pi^2 k^{2s}}{\pi^4 (k-1)^4} 2((-1)^{k+1}-1)^2\end{aligned}$$

which is finite if and only if $s < 3/2$. ■

This shows on the one hand the lack of a universal basis for $H^s([0,1])$ for all smoothnesses $s \in \mathbb{N}$. On the other hand we see that defining function spaces using the decay of the cosine coefficients yields a different smoothness concept, where even the analytic function $x \mapsto x$ is excluded in most of them. Nevertheless, with the convenience of having simple cosine functions for a basis makes them interesting from a computational perspective, cf. [CKNS16, IKP18, GSY19].

3.5.3 Jacobi polynomials on the interval $[-1, 1]$

Now we consider algebraic polynomials on the interval $[-1,1]$ where special attention has to be drawn to effects near the border. A detailed overview can be found in the survey article [MX15] by F. Marcellán and Y. Xu or in the book [JMN21] from P. Junghanns, G. Mastroianni, and I. Notarangelo.

For $\alpha, \beta > -1$ and $v^{\alpha,\beta}(x) = (1-x)^\alpha (1+x)^\beta$ we consider the differential equation

$$v^{-\alpha,-\beta}(x) \frac{\mathrm{d}}{\mathrm{d}x}\Big(v^{\alpha+1,\beta+1}(x)\frac{\mathrm{d}}{\mathrm{d}x} y(x)\Big) = -k(k+\alpha+\beta+1)y(x)\,.$$

The solutions are the so-called **Jacobi polynomials** $p_k^{\alpha,\beta}$, which are orthogonal with respect to $\langle f, g\rangle_{\alpha,\beta} = \int_{-1}^{1} f(x)\overline{g(x)} v^{\alpha,\beta}(x)\,\mathrm{d}x$, cf. [Sze75, Theorem 4.2.1]. In case $\alpha = \beta$, they are called **Gegenbauer polynomials**, where we pay special attention to the following two examples:

- With $\alpha = \beta = 0$ (unweighted Lebesgue measure $v^{0,0} \equiv 1$) we obtain the **Legendre polynomials** $\{P_k\}_{k=0}^{\infty}$ with

  $$P_k(x) = \frac{1}{2^k k!} \frac{\mathrm{d}^k}{\mathrm{d}x^k} (x^2-1)^k\,,$$

  which are normalized such that $P_k(1) = 1$ and $\|P_k\|_{L_2([-1,1])}^2 = 2/(2k+1)$.

- For $\alpha = \beta = -1/2$ (Chebyshev measure $v^{-1/2,-1/2}(x) = (1-x^2)^{-1/2}$) we obtain the **Chebyshev polynomials** $\{T_k\}_{k=0}^{\infty}$ with

  $$T_k(x) = \cos(k \arccos(x))$$

  which are normalized such that $\|T_0\|^2_{L_2([-1,1],(1-x^2)^{-1/2})} = \pi$ and $\|T_k\|^2_{L_2([-1,1],(1-x^2)^{-1/2})} = \pi/2$.

Using Theorem 3.11, we define RKHSs from them with a certain decay of the coefficients $\langle f, p_k^{\alpha,\beta}\rangle_{L_2([-1,1],v^{\alpha,\beta})}$

$$\begin{aligned} &L_2^s([-1,1],v^{\alpha,\beta}) \\ &\quad = \left\{ f \in L_2([-1,1],v^{\alpha,\beta}) : \sum_{k=0}^{\infty} (k+1)^{2s} |\langle f, p_k^{\alpha,\beta}\rangle_{\alpha,\beta}|^2 < \infty \right\}, \end{aligned}$$

which can be found in [JMN21, (2.4.1)] and are called **Sobolev-type subspace** $L_2^s([-1,1],v^{\alpha,\beta})$ of $L_2([-1,1],v^{\alpha,\beta})$. As in Section 3.5.1, we are interested in the connection of this coefficient decay and smoothness properties of the function f.

Lemma 3.35. *Let $s \in \mathbb{N}$. Then $f \in L_2^s([-1,1],v^{\alpha,\beta})$ if and only if*

$$x \mapsto f^{(r)}(x)(1-x^2)^{r/2} \in L_2([-1,1],v^{\alpha,\beta})$$

for $r = 0,\dots,s$.

Proof. [BHS92, Conclusion 2.3] or [JMN21, Lemma 2.4.7]. ■

Theorem 3.35 gives a characterization, where the weak existence of the derivatives is coupled with weights depending on the order of the derivative. Comparing these spaces to the ones from Section 3.5.2, we choose $\alpha = \beta = 0$. The difference is the additional weight $(1-x^2)^{r/2}$ in Theorem 3.35, which allows the functions to grow at the border. Thus, up to the scaling of the interval, $H^s([0,1])$ is smaller than $L_2^s([-1,1],v^{0,0})$. This comes with a drawback when approximating functions from samples. In order to see the price we pay we have to introduce the **Christoffel function** and its supremum. We do this for a general domain $D \subseteq \mathbb{R}^d$, as it will be useful later on as well:

$$N(V,x) = \sum_{k=1}^{m-1} |\eta_k(x)|^2 \quad \text{and} \quad N(V) := \sup_{x\in D} N(V,x)\,, \tag{3.14}$$

with $V = \{\eta_1, \ldots, \eta_{m-1}\}$ an orthonormal system. We know by the Riesz representer Theorem that the norm of an element from the dual space is equal to the norm of its representer. Thus,

$$N(V,x) = \sup_{\|\hat{\boldsymbol{f}}\|_2^2 \le 1} \sum_{k=1}^{m-1} \hat{f}_k \eta_k(x) = \sup_{f \in \operatorname{span}\{\eta_1,\ldots,\eta_{m-1}\}} \frac{f(x)}{\|f\|_{L_2}}, \qquad (3.15)$$

which shows that the Christoffel function is independent of the chosen orthonormal basis.

Via Hölder's inequality we obtain

$$\|f\|_\infty = \Big| \sum_{k=1}^{m-1} \hat{f}_k \eta_k(x) \Big| \le \sqrt{N(V)} \|f\|_{L_2}.$$

Thus, for a small Christoffel function, the function evaluations $f(x)$ do not blow up when the L_2 norm is small. The smallest value is of the order m as $\int_D N(V,\cdot)\, \mathrm{d}\varrho_T = m$ and by the mean value Theorem there exists $x \in D$ where this value is attained.

We have seen that the $H^1([0,1])$ and $H^2([0,1])$ bases from Section 3.5.2 are bounded orthonormal systems (BOS), i.e., there exists $B > 0$ such that $\|\eta_k\|_\infty \le B$ for all $k \in \mathbb{N}$. In this case we obtain the optimal order of the Christoffel function $N(V) \le Bm$.

In contrast to the Legendre polynomials, where we have

$$N(V) = \sum_{k=0}^{m-1} \frac{|P_k(1)|^2}{\|P_k\|_{L_2}} = \sum_{k=0}^{m-1} \frac{2k+1}{2} = \frac{m + m(m-1)}{2},$$

which grows quadratically. This also affects the needed number of samples for a well-conditioned least squares matrix, cf. Theorem 6.4. As a consequence the results in Chapters 7 and 8 hold for a larger number of basis functions of the spaces $H^1([0,1])$ and $H^2([0,1])$ compared to their polynomial counterpart.

In the following lemma we relate unweighted smoothness to the decay of the coefficients in the special case for Legendre and Chebyshev coefficients.

Lemma 3.36. *Let $f, \ldots, f^{(s-1)} \colon [-1,1] \to \mathbb{C}$ be absolute continuous and $f^{(s)}$ of bounded variation $V < \infty$. Then we have for the Legendre coefficients*

$$\begin{aligned}\Big| \Big\langle f, \frac{P_k}{\|P_k\|_{L_2}} \Big\rangle_{L_2([-1,1],v^{0,0})} \Big| &\le \frac{2V}{\sqrt{\pi(2k+1)(k-s)}} \prod_{\ell=1}^{s} \frac{1}{k-\ell+1/2} \\ &\lesssim k^{-(s+1)}\end{aligned}$$

and for the Chebyshev coefficients

$$\left|\left\langle f, \frac{T_k}{\|T_k\|_{L_2}}\right\rangle_{L_2([-1,1],v^{-1/2,-1/2})}\right| \le \sqrt{\frac{2}{\pi}} V \prod_{\ell=0}^{s} \frac{1}{k-\ell} \lesssim k^{-(s+1)} .$$

In particular, we obtain for the unweighted norm $f \in L_2^{s+1/2}([-1,1],v^{0,0})$ and for the Chebyshev norm $f \in L_2^{s+1/2}([-1,1],v^{-1/2,-1/2})$.

Proof. The estimate for the Legendre coefficients was shown in [Wan23, Theorem 3.5] and for the Chebyshev coefficients in [Tre13, Theorem 7.1] or [PPST18, Theorem 6.16]. ■

The reverse direction in Theorem 3.36 is not know to us. Further, at first sight it seems like we received half an order of convergence for free. This is not the case as the bounded variation condition is stronger as the following lemma shows in the unweighted case.

Lemma 3.37. *Let X^s be all functions $f\colon [-1,1] \to \mathbb{C}$ satisfying the assumptions of Theorem 3.36, i.e., we have $f, \dots, f^{(s-1)}\colon [-1,1] \to \mathbb{C}$ are absolute continuous and $f^{(s)}$ is of bounded variation.*

Then $f \in H^{s+1/2-\varepsilon}([-1,1],v^{0,0})$ for all $\varepsilon > 0$, where the Sobolev space for non-integer smoothness is defined in the usual way, cf. [DNPV12].

Proof. We need to define the Besov space $B^s_{p,q}$ for $p = 1$, $q = \infty$, and integer smoothness s

$$B^s_{1,\infty} := \left\{ f \in L_1 : \sup_{h \neq 0} \frac{\|\Delta_h^2 f^{(s-1)}\|_{L_1}}{|h|} < \infty \right\}$$

with the finite difference $(\Delta_h f)(x) := f(x+h) - f(x)$ and $\Delta_h^2 = \Delta_h \circ \Delta_h$, cf. [Tri92, Section 1.2.5].

For $f \in X^s$ the derivative $f^{(s)}$ is of bounded variation. Thus, also the finite difference $\Delta_h^2 f$ is of bounded variation. In particular, $f^{(s)} \in L_1$ and, therefore, $f \in B^{s+1}_{1,\infty}$. By [Tri92, (2.3.2/23)], we further have $B^{s+1}_{1,\infty} \hookrightarrow B^{s+1-\varepsilon}_{1,1}$ for any $\varepsilon > 0$. Thus,

$$X^s \hookrightarrow B^{s+1}_{1,\infty} \hookrightarrow B^{s+1-\varepsilon}_{1,1} \hookrightarrow H^{s+1/2-\varepsilon} ,$$

where the third embedding follows from the Sobolev inequality, cf. [Tri10, (2.7.1/1)]. ■

Concluding this section, we have seen that the polynomial spaces are well-studied and in the special case of $L_2^s([-1,1],v^{0,0})$ for $s = 1,2$ can be compared to $H^1([0,1])$ and $H^2([0,1])$ from Section 3.5.2. In particular, we have seen that $L_2^s([-1,1],v^{0,0})$ is bigger than $H^s([0,1])$ but does not have a bounded orthonormal basis, which affects analysis of functions in the uniform norm and, consequently, the approximation from samples.

3.6 Extension to higher dimensions

In Section 3.5 we gave one-dimensional examples of RKHSs. Many phenomena and insights are covered by them. But already problems in the three-dimensional physical space require an extension of this concept. In the heyday of machine learning with superfluous amounts of data from the internet of things high-dimensional problems become more important. Here, the curse of dimensionality is ubiquitous, i.e., the amount of data needed grows exponentially when the dimension increases.

We introduce three different concepts to extend the one-dimensional function spaces to higher dimensions:

- isotropic Sobolev spaces,
- Sobolev spaces with dominating mixed smoothness, and
- truncated analysis of variance (ANOVA) decomposition.

They increase in computational applicability for higher dimensions and decrease in expressiveness, i.e., the size of the corresponding function space.

3.6.1 Isotropic Sobolev spaces on the torus $\mathbb{T}^d$

When coming up with smoothness concepts for a function $f\colon D\to\mathbb{K}$, $D\subseteq\mathbb{R}^d$ in multiple dimensions $d>1$ the straight-forward approach is to require the existence and continuity of all partial derivatives of first order. Functions of this type are denoted by $C^1(D)$. For $s\in\mathbb{N}_0$, this concepts recursively extends to **s-times continuously differentiable functions**

$$C^s(D) := \left\{ f\colon D\to\mathbb{K}^d : D^{\boldsymbol{\alpha}} f \in C(D) \text{ for all } \boldsymbol{\alpha}\in\mathbb{N}_0^d \text{ with } \|\boldsymbol{\alpha}\|_1\le s \right\},$$

where $D^{\boldsymbol{\alpha}} f := \partial^{\|\boldsymbol{\alpha}\|_1} f/(\partial x_1^{\alpha_1}\dots\partial x_d^{\alpha_d})$. To regain the Hilbert structure the canonical approach is to demand square-integrability instead of continuity of the same derivatives. This leads to the **isotropric Sobolev spaces of smoothness** $s\in\mathbb{N}_0$

$$H^s(D) := \left\{ f\in L_2(D) : \|f\|_{H^s}^2 = \sum_{\|\boldsymbol{\alpha}\|_1\le s} \|D^{\boldsymbol{\alpha}} f\|_{L_2}^2 < \infty \right\}.$$

A first central question is whether the one-dimensional orthogonal basis functions η_k keep their properties when extending them to higher dimensions via the canonical tensor product ansatz

$$\eta_{\boldsymbol{k}}(\boldsymbol{x}) = \eta_{(k_1,\dots,k_d)}((x_1,\dots,x_d)^\mathsf{T}) = \prod_{j=1}^{d} \eta_{k_j}(x_j)\,.$$

Already for the interval $D = [0,1]$ considered in Section 3.5.2 this question is non-trivial as it was commented in [IN09, Section 2] or [Adc10a, Section 3.5].

As we merely want to introduce the core idea, we focus on the domain D being the d-dimensional torus $\mathbb{T}^d$ for a straight-forward analysis. Here, we have the basis

$$\eta_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^{d} \exp(2\pi\mathrm{i}k_j x_j) = \exp(2\pi\mathrm{i}\langle \boldsymbol{k}, \boldsymbol{x}\rangle)$$

for $\boldsymbol{k} \in \mathbb{Z}^d$ as the concepts are well-studied and approachable in this case. Since there is no natural ordering of the multi indices $\boldsymbol{k} \in \mathbb{Z}^d$, a first goal is a characterization in terms of the decay of the Fourier coefficients $\langle f, \eta_{\boldsymbol{k}}\rangle_{L_2}$ similar to Theorem 3.24. To achieve that we need the following lemma on the equivalence of finite ℓ_p-norms, which extends to $0 < p < 1$ even though it is no norm in this case.

Lemma 3.38. *For* $0 < p \le q \le \infty$ *and* $\boldsymbol{k} \in \mathbb{C}^d$ *we have*

$$\|\boldsymbol{k}\|_q \le \|\boldsymbol{k}\|_p \le d^{\frac{1}{p}-\frac{1}{q}} \|\boldsymbol{k}\|_q\,.$$

Proof. **Step 1.** We first prove the left-hand inequality. Since $p \mapsto x^p$ is increasing for $x \in [0,1]$, we have

$$\|\boldsymbol{k}\|_p^{-q} \|\boldsymbol{k}\|_q^q = \left\| \frac{\boldsymbol{k}}{\|\boldsymbol{k}\|_p} \right\|_q^q = \sum_{j=1}^{d} \left| \frac{k_j}{\|\boldsymbol{k}\|_p} \right|^q \le \sum_{j=1}^{d} \left| \frac{k_j}{\|\boldsymbol{k}\|_p} \right|^p = 1\,.$$

Step 2. Now we use Jensen's inequality to obtain the second inequality:

$$\|\boldsymbol{k}\|_p = \left(\left(\frac{\sum_{j=1}^{d} |k_j|^p}{d} \right)^{\frac{q}{p}} d^{\frac{q}{p}} \right)^{\frac{1}{q}} \le \left(\sum_{j=1}^{d} |k_j|^q d^{\frac{q}{p}-1} \right)^{\frac{1}{q}} = d^{\frac{1}{p}-\frac{1}{q}} \|\boldsymbol{k}\|_q\,. \quad \blacksquare$$

Using this lemma, we show the decay of the Fourier coefficients for functions from isotropic Sobolev spaces on $\mathbb{T}^d$.

Lemma 3.39. *Let $s \in \mathbb{N}_0$. Then $f \in H^s(\mathbb{T}^d)$ if and only if for every $0 < p \leq \infty$ we have*

$$\sum_{\boldsymbol{k} \in \mathbb{Z}^d} \|\boldsymbol{k}\|_p^{2s} |\langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}|^2 < \infty \,.$$

Proof. **Step 1.** Establishing an explicit formula for derivatives of f. Since $\eta_{\boldsymbol{k}}$ are the trigonometric polynomials, we have the relation $D^{\boldsymbol{\alpha}} \eta_{\boldsymbol{k}} = (2\pi\mathrm{i}\boldsymbol{k})^{\boldsymbol{\alpha}} \eta_{\boldsymbol{k}}$ where $(2\pi\mathrm{i}\boldsymbol{k})^{\boldsymbol{\alpha}} := \prod_{j=1}^{d} (2\pi\mathrm{i}k_j)^{\alpha_j}$. Thus derivatives of $f \in L_2(\mathbb{T}^d)$ are given by

$$D^{\boldsymbol{\alpha}} f = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} (2\pi\mathrm{i}\boldsymbol{k})^{\boldsymbol{\alpha}} \langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}$$

and the condition $f \in H^s(\mathbb{T}^d)$ is equivalent to

$$\sum_{\boldsymbol{k} \in \mathbb{Z}^d} |\boldsymbol{k}^{\boldsymbol{\alpha}} \langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}|^2 < \infty \quad \text{for all} \quad \|\boldsymbol{\alpha}\|_1 \leq s \,. \tag{3.16}$$

Step 2. Showing $f \in H^s(\mathbb{T}^d)$ implies Fourier coefficient decay. By Theorem 3.38 all ℓ_p norms are the same up to a d-dependent constant. Thus, showing the result for $p = 1$ is enough. Jensen's inequality gives $\|\boldsymbol{k}\|_1^{2s} \leq d^{2s-1} \sum_{j=1}^{d} k_j^{2s}$ and thus

$$\begin{aligned} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \|\boldsymbol{k}\|_1^{2s} |\langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}|^2 &\leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \Big(d^{2s-1} \sum_{j=1}^{d} k_j^{2s} \Big) |\langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}|^2 \\ &\leq d^{2s-1} \sum_{j=1}^{d} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} k_j^{2s} |\langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}|^2 \end{aligned}$$

where the latter sums are all finite by (3.16) with the choices $\boldsymbol{\alpha} = s\boldsymbol{e}_j$ for $j = 1, \ldots, d$ and $\boldsymbol{e}_j$ the j-th unit vector.

Step 3. Showing Fourier coefficient decay implies $f \in H^s(\mathbb{T}^d)$. We have

$$\sum_{\boldsymbol{k} \in \mathbb{Z}^d} \boldsymbol{k}^{2\boldsymbol{\alpha}} |\langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}|^2 \leq \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \|\boldsymbol{k}\|_\infty^{2\|\boldsymbol{\alpha}\|_1} |\langle f, \eta_{\boldsymbol{k}} \rangle_{L_2}|^2 \,,$$

which is finite for all $\|\boldsymbol{\alpha}\|_1 \leq s$ by assumption. ■

From Theorem 3.39 we have that the Fourier coefficients decay according to an ℓ_p-norm of their frequency. If we want to approximate a function by

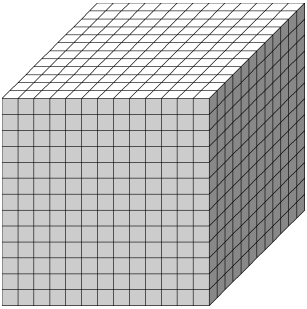


Figure 3.3: Cube of frequencies $\{\boldsymbol{k} \in \mathbb{Z}^3 : \|\boldsymbol{k}\|_\infty \leq 6\}$.

truncating its Fourier coefficients have to truncate accordingly. The next lemma shows that the d-dependence is independent of the chosen p and enters exponentially.

Lemma 3.40. *Let* $0 < p \leq \infty$ *and* $I_N^{d,p} := \{\boldsymbol{k} \in \mathbb{Z}^d : \|\boldsymbol{k}\|_p \leq N\}$. *Then* $|I_N^{d,p}| \sim N^d$.

Proof. **Step 1.** Showing the assertion for $1 \leq p$. In the boundary cases we have $|I_N^{d,1}| \sim |I_N^{d,\infty}| \sim N^d$. For $1 < p \leq \infty$ we use $|I_N^{d,1}| \leq |I_N^{d,p}| \leq |I_N^{d,\infty}|$, which follows from Theorem 3.38.

Step 2. It remains to show the assertion for $0 < p < 1$. By Theorem 3.38 we have $I_{d^{1-1/p}N}^{d,1} \subseteq I_N^{d,p}$. Thus, $|I_N^{d,p}| \leq |I_{d^{1-1/p}N}^{d,1}| \sim N^d$. ■

If we are solely concerned about the d-dependence, Theorem 3.40 states that we might as well chose the largest ℓ_p ball with $p = \infty$, which is depicted in Figure 3.3. For this frequency set multidimensional fast Fourier algorithms [CT65, KKP09] are applicable which makes this approach computationally interesting. On the downside, as the ℓ_p-balls grow exponentially in the dimension, the curse of dimensionality enters. In fact, in [KSU14] the behavior of the approximation numbers was determined $a_n(I_K \colon H^s(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \sim n^{-s/d}$, which shows that the number of information has to grow exponentially in the dimension to achieve the same accuracy. Nevertheless, in lower dimensions $d \in \{1, 2, 3\}$ this approach is feasible and in [KMU16] the preasymptotic behavior was studied where the involved constants decay polynomial in d.

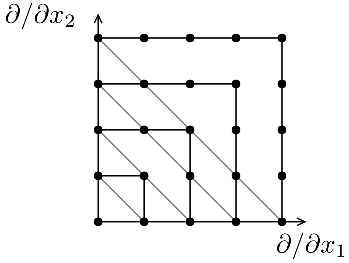


Figure 3.4: Visualization of the isotropic Sobolev spaces (gray) and Sobolev spaces with dominating mixed smoothness (black).

3.6.2 Sobolev spaces with dominating mixed smoothness

As we have seen in Section 3.6.1 the canonical approach to extend smoothness concepts to higher dimensions yields isotropic Sobolev spaces which are troublesome in higher dimensions. Another approach is do demand the existence of all derivatives of dominating mixed smoothness. This leads to the **Sobolev spaces with dominating mixed smoothness** $s \in \mathbb{N}_0$

$$H^s_{\text{mix}}(D) \coloneqq \Big\{ f \in L_2(D) : \|f\|^2_{H^s_{\text{mix}}} = \sum_{\boldsymbol{\alpha} \in \{0,s\}^d} \|D^{\boldsymbol{\alpha}} f\|^2_{L_2} < \infty \Big\}.$$

When comparing them to the isotropic Sobolev spaces the existence of partial derivatives $D^{\boldsymbol{\alpha}} f$ is required for the ℓ_∞ ball instead of ℓ_1 as depicted in Figure 3.4 for dimension $d = 2$. From this we obtain the immediate embedding $H^{ds}(D) \hookrightarrow H^s_{\text{mix}}(D) \hookrightarrow H^s(D)$. The original idea to consider these spaces goes back to K. I. Bebenko and S. M. Nikol'skii, cf. [Bab60, Nik63]. They occur naturally as solution for hyperbolic partial differential equations [Mam15], in quantum mechanics [Yse04], in discrepancy theory [BLV08], for computing entropy numbers [KL93], or approximation theory [Tem93a, DuTU18].

It was shown in [SU09] that the Sobolev spaces with dominating mixed smoothness are a tensor product of the one-dimensional Sobolev spaces

$$H^s_{\text{mix}}(D) = H^s(D_1) \otimes \cdots \otimes H^s(D_1)$$

for $D = D_1 \otimes \cdots \otimes D_1$. Because of this structure the basis and singular values

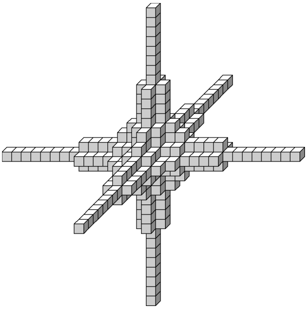


Figure 3.5: Hyperbolic cross for dimension $d = 3$ and radius $R = 15$.

of the embedding $H^s_{\text{mix}}(D) \hookrightarrow L_2(D)$ are given by

$$\eta_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^{d} \eta_{k_j}(x_j) \quad \text{and} \quad \sigma^2_{\boldsymbol{k}} = \prod_{j=1}^{d} \sigma^2_{k_j} \, .$$

For approximation the singular functions corresponding to the largest singular values are most important. Collecting all these multi-indices $\boldsymbol{k} \in \mathbb{Z}^d$ we end up with the so-called **hyperbolic cross** of radius R

$$I_{\text{hc}} := \left\{ \boldsymbol{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d : \prod_{j=1}^{d} \max\{1, k_j\} \leq R \right\},$$

where one is depicted in Figure 3.5. Note, that we included negative multi-indices as this originates from the frequencies used in trigonometric polynomials. When approximating with polynomials or the $H^1([0,1])$ or $H^2([0,1])$ basis from Sections 3.5.2 and 3.5.3, we only need the first quadrant. A huge advantage is its cardinality of merely $|I_{\text{hc}}| \asymp R(\log R)^{d-1}$ frequencies, cf. Theorem 7.11 or [BKUV17]. Thus, the dimension d only enters in the logarithm, which makes these spaces applicable in higher dimensions compared to the isotropic Sobolev spaces from Section 3.6.1.

3.6.3 Truncated ANOVA decomposition

Another way of approaching high-dimensional problems, gaining recent popularity, is via the truncated ANOVA decomposition, which we address in this section based on [BPS22]. In-depth works on this topic are [CMO97, RFA99, LO06, NW08, KSWW09, Gu13, Sch22] where we only give a brief introduction and restrict ourselves to $D = \mathbb{T}^d$ for simplicity. The core idea is that certain functions are representable as a sum of lower-dimensional functions, e.g.

$$f(x_1, \dots, x_9) = \exp(x_1) + \sin(x_1)\cos(x_2) + x_5 x_6^3 x_7^5 .$$

The above function f is nine-dimensional but may be decomposed into a sum of one one-dimensional function, one two-dimensional, and one three-dimensional one. This assumption occurs, e.g. naturally in calculations of the electronic structure problem for molecules in [GHH11] where componentwise interactions are intrinsic. Even when this assumption is not given, the truncation to lower-dimensional terms has been proven to beat past methods in practice on benchmark problems, cf. [Sch22, Chapter 6].

A central tool are integral projections

$$P_{\boldsymbol{u}} f(\boldsymbol{x}) = \int_{\mathbb{T}^{d-|\boldsymbol{u}|}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}_{\boldsymbol{u}^\complement}$$

for a subset of coordinate indices $\boldsymbol{u} \subseteq \{1, \dots, d\}$ and its complement $\boldsymbol{u}^\complement = \{1, \dots, d\} \setminus \boldsymbol{u}$. Additionally, for vectors $\boldsymbol{x} \in \mathbb{T}^d$ indexed with a subset $\boldsymbol{u} \subseteq \{1, \dots, d\}$ we define $\boldsymbol{x}_{\boldsymbol{u}} := (x_j)_{j \in \boldsymbol{u}}$. For $\boldsymbol{u} \subseteq \{1, \dots, d\}$, the **ANOVA terms** and **ANOVA decomposition** are given by

$$f_{\boldsymbol{u}} = P_{\boldsymbol{u}} f - \sum_{\boldsymbol{v} \subsetneq \boldsymbol{u}} f_{\boldsymbol{v}} \quad \text{and} \quad f = \sum_{\boldsymbol{u} \subseteq \{1, \dots, d\}} f_{\boldsymbol{u}} .$$

For $D = \mathbb{T}^d$ the ANOVA terms are orthogonal and expressible in terms of its Fourier coefficients

$$f_{\boldsymbol{u}} = \sum_{\substack{\boldsymbol{k} \in \mathbb{Z}^d \\ \operatorname{supp} \boldsymbol{k} = \boldsymbol{u}}} \langle f, \eta_{\boldsymbol{k}} \rangle_{L_2} \eta_{\boldsymbol{k}}$$

with $\operatorname{supp} \boldsymbol{k} = \{j \in \{1, \dots, d\} : k_j \neq 0\}$. A depiction of the frequencies corresponding to different dimensions is in Figure 3.6. For other domains with orthonormal systems this works similar, cf. [Sch22], or a bit more involved with wavelets, cf. [LPU23].

The number of ANOVA terms is 2^d and therefore grows exponentially in the dimension which reflects the well known curse of dimensionality. The idea to circumvent this is to truncate the decomposition and only take a certain number of terms into account. It is common to truncate to lower-dimensional terms $f_{\boldsymbol{u}}$ with $|\boldsymbol{u}| \leq d_s$. Then, the number of terms with respect to the spatial dimension d is $\sum_{j=1}^{d_s} \binom{d_s}{d} \in \mathcal{O}(d^{d_s})$, which grows polynomially instead of

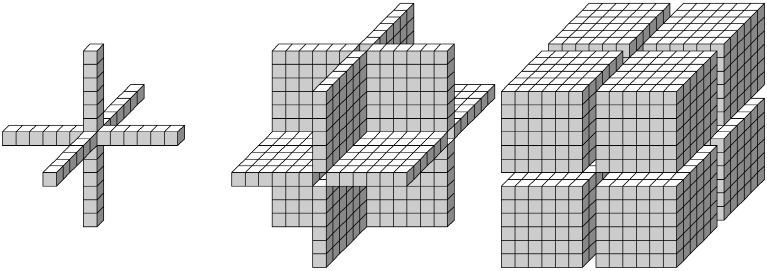


Figure 3.6: Decomposition of the frequency into the different dimensions of the ANOVA decomposition.

exponentially. Furthermore, amongst the terms $f_{\boldsymbol{u}}$ it is possible to find the ones contributing most to the overall function via sensitivity analysis decreasing the number of terms even more. Next to that, this approach is combinable with fast Fourier methods, cf. [BPS22, Sch22].

Chapter 4

Concentration inequalities

Randomness and random constructions are used in e.g. compressive sensing, Monte Carlo methods, or function approximation. To show error bounds of these constructions, concentration inequalities are used. In this section we repeat on a selection of them which we will use later on. A good collection and introduction can be found in [Ver18].

In particular, Section 4.1 is about concentration inequalities in random vectors starting with the well-known Bernstein inequality followed by the more general Hanson-Wright inequality, which we prove for complex random vectors under Bernstein conditions. In Section 4.2 we consider matrix-valued concentration inequalities for finite- and infinite-dimensional matrices.

4.1 Concentration inequalities for random vectors

The bounds in this section are on random variables of the form $f\colon \mathbb{C}^n \to \mathbb{R}$. We present Bernstein's inequality where f is the sum, the Hanson-Wright inequality for quadratic forms, and McDiarmid's inequality for general f satisfying a certain $\boldsymbol{c}$- boundedness condition.

We start with Bernstein's inequality, which is found in the standard literature, cf. [FR13, Corollary 7.31] or [SC08, Theorem 6.12].

Theorem 4.1 (Bernstein). *Let $\xi_1, \dots, \xi_n$ be independent real-valued mean-zero random variables satisfying $\mathbb{E}(\xi_i^2) \le \sigma^2$ and $\|\xi_i\|_\infty \le K$ for $i = 1, \dots, n$ and real numbers σ^2 and K. Then*

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i \le \frac{2Kt}{3n} + \sqrt{\frac{2\sigma^2 t}{n}}$$

with probability exceeding $1 - \exp(-t)$.

Bernstein's inequality gives a concentration bound for the sum of independent random variables. We need similar bounds for quadratic forms in random vectors, which are known as Hanson-Wright inequalities. To formulate them, we need to introduce the spectral norm and the Frobenius norm of a matrix

$\boldsymbol{A} \in \mathbb{C}^{m\times n}$

$$\|\boldsymbol{A}\|_{2\to 2} = \sqrt{\lambda_{\max}(\boldsymbol{A}^*\boldsymbol{A})} = \sigma_{\max}(\boldsymbol{A}) \quad \text{and} \quad \|\boldsymbol{A}\|_F = \sqrt{\sum_{k=1}^{m}\sum_{i=1}^{n}|a_{k,i}|^2}\,, \tag{4.1}$$

where $\lambda_{\max}$ and $\sigma_{\max}$ denote the largest eigenvalue and singular value, respectively. The following result is such an inequality with a Bernstein condition on the random variables.

Theorem 4.2 (Hanson-Wright)**.** *Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^{\mathsf{T}}$ be a vector of independent complex-valued mean-zero random variables such that*

$$\mathbb{E}(|\xi_i|^{2p}) \leq p!K^{2p-2}\sigma_i^2/2 \tag{4.2}$$

for $0 \leq \sigma_i^2 \leq K$, $i = 1, \ldots, n$, and $p \in \mathbb{N}$. Let further $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ Hermitian, $m = \mathbb{E}(\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi})$, and $\boldsymbol{D}_\sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$. Then

$$\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi} - m \leq \max\left\{256K^2\|\boldsymbol{A}\|_{2\to 2}t, 8\sqrt{3}K\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F\sqrt{t}\right\}$$

with probability exceeding $1 - \exp(-t)$.

Remark 4.3. *In [BLM13, Example 2.12] is a real-valued version of the Hanson-Wright inequality for Gauss distributed random variables, which is a special case of the above with better constants (the rotation invariance of a Gauss distributed vector is used). There is also a more general Hanson-Wright type inequality for sub-Gaussian random variables in [RV13] without specified constants.*

Theorem 4.2 was shown in [Bel19, Theorem 3] for real valued random variables and real matrices. With some adjustments the proof works for the complex case as well. For that we need a preliminary lemma.

Lemma 4.4. *Let $\boldsymbol{\xi}$ be as in Theorem 4.2, $a \in \mathbb{C}$, $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathsf{T}} \in \mathbb{C}^n$, and $\boldsymbol{A}_0 \in \mathbb{C}^{n\times n}$ Hermitian with zeros on the diagonal. Then*

$$\mathbb{E}(\exp(|a\xi|)) \leq \exp(|aK|^2)\,,$$

$$\mathbb{E}(|\exp(\langle \boldsymbol{a}, \boldsymbol{\xi}\rangle)|) \leq \exp(K^2\|\boldsymbol{a}\|_2^2)\,,$$

and

$$\mathbb{E}(\exp(\boldsymbol{\xi}^*\boldsymbol{A}_0\boldsymbol{\xi})) \leq \mathbb{E}(\exp(16K^2\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2)\,.$$

Furthermore, for $|a|K^2 \leq 1/2$, we have

$$\mathbb{E}(|a\xi_i^2 - a\sigma_i^2|) \leq \exp(|a|^2\sigma_i^2K^2)$$

and

$$\mathbb{E}(|a\xi_i^2|) \leq \exp\left(\frac{3}{2}|a|\sigma_i^2\right)\,.$$

Proof. The first, fourth, and fifth inequality are stated in [Bel19, Proposition 4] for real numbers. The extension to complex numbers is straight-forward by inserting absolute values in appropriate places.

Second inequality. Using the first inequality we have

$$\begin{aligned}\mathbb{E}(|\exp(\langle \boldsymbol{a}, \boldsymbol{\xi}\rangle)|) &= \mathbb{E}\Big(\Big|\prod_{i=1}^{n}\exp(a_i\overline{\xi_i})\Big|\Big) \leq \mathbb{E}\Big(\prod_{i=1}^{n}\exp(|a_iK|^2)\Big)\\ &= \mathbb{E}\Big(\exp(K^2\|\boldsymbol{a}\|_2^2)\Big)\,.\end{aligned}$$

Third inequality. Since $|\exp(z)|$ is convex on the complex plane, we can use the decoupling theorem from [Ver11] or [FR13, Theorem 8.11]. For $\boldsymbol{\xi}'$ an independent copy of $\boldsymbol{\xi}$, we obtain

$$\mathbb{E}_{\boldsymbol{\xi}}(|\exp(\boldsymbol{\xi}^*\boldsymbol{A}_0\boldsymbol{\xi})|) \leq \mathbb{E}_{\boldsymbol{\xi},\boldsymbol{\xi}'}(|\exp(4\boldsymbol{\xi}^*\boldsymbol{A}_0\boldsymbol{\xi}')|) \leq \mathbb{E}_{\boldsymbol{\xi}}(\exp(16K^2\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2))\,,$$

where we used the second inequality in the last line. ■

Proof of Theorem 4.2. We decompose $\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi}$ into the diagonal and the off-diagonal part

$$\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi} - m = \sum_{i=1}^{n} a_{ii}(|\xi_i|^2 - \sigma_i^2) + \boldsymbol{\xi}^*\boldsymbol{A}_0\boldsymbol{\xi} =: S_1 + S_2\,.$$

In the following we bound the moment generating functions of S_1 and S_2 in order to apply Chernoff bound. Let $\lambda > 0$ satisfy

$$128\|\boldsymbol{A}\|_{2\to2}K^2\lambda \leq 1\,. \tag{4.3}$$

Step 1. Bounding of the moment generating function of S_1. We apply the fourth inequality of Theorem 4.4 (applicable because of (4.3))

$$\begin{aligned}\mathbb{E}(\exp(\lambda S_1)) &= \mathbb{E}\Big(\exp\Big(\lambda\sum_{i=1}^{n} a_{ii}(|\xi_i|^2 - \sigma_i^2)\Big)\Big)\\ &\leq \exp\Big(\lambda^2K^2\sum_{i=1}^{n}|a_{ii}|^2\sigma_i^2\Big)\,.\end{aligned}$$

Step 2. Preparing auxiliary linear algebra. We define the matrix $\boldsymbol{A}_0 = \boldsymbol{A} - \operatorname{diag}(a_{11}, \ldots, a_{nn})$ being $\boldsymbol{A}$ with zeros on the diagonal. Since $|a_{ii}| \leq \|\boldsymbol{A}\|_{2\to2}$, we have by triangle inequality

$$\|\boldsymbol{A}_0\|_{2\to2} \leq 2\|\boldsymbol{A}\|_{2\to2}\,.$$

Let $\boldsymbol{B} = \boldsymbol{A}_0^* \boldsymbol{A}_0 = (b_{ij})_{i,j=1,\ldots,n}$ and $\boldsymbol{B}_0 = \boldsymbol{B} - \operatorname{diag}(b_{11}, \ldots, b_{nn})$. Then,

$$0 \le b_{ii} = \sum_{j \ne i} |a_{ji}|^2 \le \|\boldsymbol{A}\|_{2\to2}^2$$

and
$$\begin{aligned} \|\boldsymbol{B}_0 \boldsymbol{\xi}\|_2^2 &\le 2\|\boldsymbol{B}\boldsymbol{\xi}\|_2^2 + 2\sum_{i=1}^n |b_{ii}\xi_i|^2 \\ &\le 2\|\boldsymbol{A}_0\|_{2\to2}^2 \|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2 + 2\|\boldsymbol{A}\|_{2\to2}^2 \sum_{i=1}^n |b_{ii}||\xi_i|^2 \\ &\le 2\|\boldsymbol{A}\|_{2\to2}^2 \Big(2\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2 + \sum_{i=1}^n |b_{ii}||\xi_i|^2 \Big) . \end{aligned}$$

Step 3. Bounding of the moment generating function of S_2. Let $\eta = 32K^2\lambda^2$. We use the third inequality of Theorem 4.4

$$\mathbb{E}(\exp(\lambda S_2)) \le \mathbb{E}(\exp(16\lambda^2 K^2 \|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2)) = \mathbb{E}\Big(\exp\Big(\frac{\eta}{2}\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2\Big)\Big) . \tag{4.4}$$

Using $\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2 = \sum_{i=1}^n b_{ii}|\xi_i|^2 + \boldsymbol{\xi}^* \boldsymbol{B}_0 \boldsymbol{\xi}$, the Cauchy-Schwarz inequality, and the third inequality of Theorem 4.4 again, we obtain

$$\begin{aligned} \mathbb{E}\Big(\exp\Big(\frac{\eta}{2}\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2\Big)\Big)^2 &\le \mathbb{E}\Big(\exp\Big(\eta \sum_{i=1}^n b_{ii}|\xi_i|^2\Big)\Big) \mathbb{E}(\exp(\eta \boldsymbol{\xi}^* \boldsymbol{B}_0 \boldsymbol{\xi})) \\ &\le \mathbb{E}\Big(\exp\Big(\eta \sum_{i=1}^n b_{ii}|\xi_i|^2\Big)\Big) \mathbb{E}(\exp(16\eta^2 K^2 \|\boldsymbol{B}_0\boldsymbol{\xi}\|_2^2)) . \end{aligned}$$

Applying the result from Step 2 and $32K^2\|\boldsymbol{A}\|_{2\to2}^2 \eta \le 1/16$ (implied by (4.3)), we obtain

$$\begin{aligned} \mathbb{E}\Big(\exp\Big(\frac{\eta}{2}\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2\Big)\Big)^2 &\le \mathbb{E}\Big(\exp\Big(\eta \sum_{i=1}^n b_{ii}|\xi_i|^2\Big)\Big) \cdot \\ &\quad \mathbb{E}\Big(\exp\Big(32\eta^2 K^2 \|\boldsymbol{A}\|_{2\to2}^2 \Big(2\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2 + \sum_{i=1}^n |b_{ii}||\xi_i|^2 \Big)\Big)\Big) \\ &\le \mathbb{E}\Big(\exp\Big(\eta \sum_{i=1}^n b_{ii}|\xi_i|^2\Big)\Big) \mathbb{E}\Big(\exp\Big(\frac{\eta}{16}\Big(2\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2 + \sum_{i=1}^n |b_{ii}||\xi_i|^2 \Big)\Big)\Big) . \end{aligned}$$

By the Cauchy-Schwarz inequality, we further have

$$\mathbb{E}\Big(\exp\Big(\frac{\eta}{2}\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2\Big)\Big)^2 \le \mathbb{E}\Big(\exp\Big(\eta\sum_{i=1}^n b_{ii}|\xi_i|^2\Big)\Big)\cdot$$

$$\sqrt{\mathbb{E}\Big(\exp\Big(\frac{\eta}{4}\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2\Big)\Big)}\sqrt{\mathbb{E}\Big(\exp\Big(\frac{\eta}{8}\sum_{i=1}^n b_{ii}|\xi_i|^2\Big)\Big)}\,.$$

Thus,

$$\mathbb{E}\Big(\exp\Big(\frac{\eta}{2}\|\boldsymbol{A}_0\boldsymbol{\xi}\|_2^2\Big)\Big)^{3/2} \le \mathbb{E}\Big(\exp\Big(\eta\sum_{i=1}^n b_{ii}|\xi_i|^2\Big)\Big)^{3/2},$$

where we can drop the power because of the positivity of b_{ii}. Plugging this into (4.4) and using the fifth equation of Theorem 4.4 (applicable because of (4.3)), we obtain

$$\mathbb{E}(\exp(\lambda S_2)) \le \mathbb{E}\Big(\exp\Big(\frac{3}{2}\eta\sum_{i=1}^n b_{ii}\sigma_i^2\Big)\Big)\,.$$

Resubstituting λ for η and using $\sum_{i=1}^n b_{ii}\sigma_i^2 = \|\boldsymbol{A}_0\boldsymbol{D}_\sigma\|_F^2$, we have an estimate for the moment generating function of S_2

$$\mathbb{E}(\exp(\lambda S_2)) \le \exp\Big(48K^2\lambda^2\|\boldsymbol{A}_0\boldsymbol{D}_\sigma\|_F^2\Big)\,.$$

Step 4. Bringing everything together. By the Chernoff bound and the Chauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbb{P}(\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi} - m > \varepsilon) &\le \exp(-\lambda\varepsilon)\mathbb{E}(\exp(\lambda S_1 + \lambda S_2))\\ &\le \exp(-\lambda\varepsilon)\sqrt{\mathbb{E}(\exp(2\lambda S_1))\mathbb{E}(\exp(2\lambda S_2))}\,.\end{aligned}$$

Using Steps 1 and 3, we have

$$\begin{aligned}&\mathbb{P}(\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi} - m > \varepsilon)\\ &\quad\le \exp(-\lambda\varepsilon)\exp\Big(48\lambda^2K^2\Big(\sum_{i=1}^n |a_{ii}|^2\sigma_i^2 + \|\boldsymbol{A}_0\boldsymbol{D}_\sigma\|_F^2\Big)\Big)\\ &\quad= \exp\Big(48K^2\lambda^2\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F^2 - \lambda\varepsilon\Big)\,.\end{aligned}$$

It remains to choose λ. The minimum for the unconstrained problem is attained at $\lambda^\star = \varepsilon/(96K^2\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F^2)$. If $\lambda^\star$ satisfies (4.3), then

$$\mathbb{P}(\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi} - m > \varepsilon) \leq \exp\Big(\frac{-\varepsilon^2}{192K^2\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F^2}\Big).$$

Otherwise we use the boundary value $\lambda^b = 1/(128\|\boldsymbol{A}\|_{2\to 2}K^2) \leq \lambda^\star$ and

$$\begin{aligned}-\varepsilon\lambda^b + (\lambda^b)^2 48K^2\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F^2 &\leq -\varepsilon\lambda^b + \frac{\lambda^b 48K^2\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F^2}{96K^2\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F^2}\\ &= \frac{-\varepsilon}{256K^2\|\boldsymbol{A}\|_{2\to 2}}.\end{aligned}$$

Overall, we obtain

$$\mathbb{P}(\boldsymbol{\xi}^*\boldsymbol{A}\boldsymbol{\xi} - m > \varepsilon) \leq \exp\Big(\max\Big\{\frac{-\varepsilon^2}{192K^2\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F^2}, \frac{-\varepsilon}{256K^2\|\boldsymbol{A}\|_{2\to 2}}\Big\}\Big).$$

The assertion follows by choosing

$$\varepsilon = \max\Big\{8\sqrt{3t}K\|\boldsymbol{A}\boldsymbol{D}_\sigma\|_F,\, 256K^2\|\boldsymbol{A}\|_{2\to 2}t\Big\}. \qquad \blacksquare$$

We are interested in the special case where the quadratic form is expressed as the Euclidean norm of a matrix-vector product and will formulate this in the next corollary.

Corollary 4.5. *Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^\mathsf{T}$ be a vector of independent complex-valued mean-zero random variables satisfying $\mathbb{E}(|\xi_i|^2) \leq \sigma^2$ and $|\xi_i| \leq K$ for $i = 1, \ldots, n$. Then for all $\boldsymbol{L} \in \mathbb{C}^{m\times n}$*

$$\|\boldsymbol{L}\boldsymbol{\xi}\|_2^2 \leq (2m\sigma^2 + 256K^2 t)\|\boldsymbol{L}\|_{2\to 2}^2$$

with probability exceeding $1 - \exp(-t)$.

Proof. **Step 1.** Applying the Hanson-Wright inequality. Since $\|\boldsymbol{L}\boldsymbol{\xi}\|_2^2 = \boldsymbol{\xi}^*\boldsymbol{L}^*\boldsymbol{L}\boldsymbol{\xi}$ is a quadratic form we will apply Theorem 4.2 on $\boldsymbol{L}^*\boldsymbol{L}$. For that we check the moment condition (4.2) on $\xi_1^2, \ldots, \xi_n^2$. For $p = 1$ it is fulfilled for constants $K/\sqrt{2}$ and $(\sqrt{2}\sigma_i)^2$. For $p \geq 2$, we have $p! \geq 2^{p-1}$ and obtain

$$\begin{aligned}\mathbb{E}(|a_i\xi_i|^{2p}) &\leq \|\xi_i\|_\infty^{2p-2}\mathbb{E}(|\xi_i|^2)\\ &\leq (K)^{2p-2}\sigma^2\\ &\leq p!\Big(\frac{K}{\sqrt{2}}\Big)^{2p-2}\frac{(\sqrt{2}\sigma)^2}{2}.\end{aligned}$$

Therefore, Theorem 4.2 is applicable.

Step 2. Calculation of the expected value. Since $\xi_1, \dots, \xi_n$ are independent and have bounded variance, we obtain

$$\begin{aligned}\mathbb{E}(\|\boldsymbol{L}\boldsymbol{\xi}\|_2^2) &= \sum_{k=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ik}\overline{a_{j,k}}\,\mathbb{E}(\xi_i\overline{\xi_j}) \\ &= \sum_{k=1}^{m}\Big(\sum_{i=1}^{n}\sum_{j\neq i} a_{ik}\overline{a_{j,k}}\,\mathbb{E}(\xi_i\overline{\xi_j})\Big) + \sum_{i=1}^{n}|a_{ik}|^2\mathbb{E}(|\xi_i|^2) \\ &\le \sigma^2\|\boldsymbol{L}\|_F^2 \,.\end{aligned}$$

Step 3. Bringing everything together. By Step 1 and 2 we obtain

$$\|\boldsymbol{L}\boldsymbol{\xi}\|_2^2 \le \sigma^2\|\boldsymbol{L}\|_F^2 + \max\Big\{128K^2\|\boldsymbol{L}^*\boldsymbol{L}\|_{2\to 2}t,\, 8\sqrt{3\sigma^2 t}K\|\boldsymbol{L}^*\boldsymbol{L}\|_F\Big\}$$

with probability exceeding $1 - \exp(-t)$. With

$$\begin{aligned}&\|\boldsymbol{L}\|_F^2 = \operatorname{trace}(\boldsymbol{L}^*\boldsymbol{L}) \le m\|\boldsymbol{L}^*\boldsymbol{L}\|_{2\to 2} \le m\|\boldsymbol{L}\|_{2\to 2}^2 \\ \text{and}\qquad &\|\boldsymbol{L}^*\boldsymbol{L}\|_F = \sqrt{\operatorname{trace}(\boldsymbol{L}^*\boldsymbol{L}\boldsymbol{L}^*\boldsymbol{L})} \le \sqrt{m}\|\boldsymbol{L}\|_{2\to 2}^2 \,,\end{aligned}$$

we obtain

$$\begin{aligned}\|\boldsymbol{L}\boldsymbol{\xi}\|_2^2 &\le \Big(\sigma^2 m + \max\Big\{128K^2 t,\, 8\sqrt{3m\sigma^2 t}K\Big\}\Big)\|\boldsymbol{L}\|_{2\to 2}^2 \\ &\le \Big((\sqrt{m\sigma^2})^2 + 2\cdot\sqrt{m\sigma^2}\cdot 4\sqrt{3t}K + (\sqrt{128t}K)^2\Big)\|\boldsymbol{L}\|_{2\to 2}^2 \\ &\le \Big(\sqrt{m\sigma^2} + \sqrt{128t}K\Big)^2\|\boldsymbol{L}\|_{2\to 2}^2 \\ &\le (2m\sigma^2 + 256tK^2)\|\boldsymbol{L}\|_{2\to 2}^2 \,.\end{aligned}$$

■

Note that the m on the right-hand side is important. If we would use $\|\boldsymbol{L}\boldsymbol{\xi}\|_2^2 \le \|\boldsymbol{L}\|_{2\to 2}^2\|\boldsymbol{\xi}\|_2^2$ and estimate the latter we would get an n instead. Thus, we gain whenever $m < n$.

Next, we state a concentration result for more general $f\colon \mathbb{C}^n \to \mathbb{R}$. It was originally stated by McDiarmid in [McD89] and requires some notation.

Definition 4.6. *A function $f\colon \Omega^n \to \mathbb{R}$ is said to be **c-bounded** on $\Xi \subseteq \Omega^n$ for $\boldsymbol{c} = (c_1, \dots, c_n) \in [0,\infty)^n$ if and only if*

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}')| \le d_{\boldsymbol{c}}(\boldsymbol{x}, \boldsymbol{x}')$$

for all $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{x}' = (x_1', \ldots, x_n') \in \Xi$ where the distance $d_{\boldsymbol{c}}$ is defined by

$$d_{\boldsymbol{c}}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i : x_i \neq x_i'} c_i.$$

Note, that a function is $\boldsymbol{c}$-bounded if changing a single variable x_i, $i = 1, \ldots, n$ changes $f(\boldsymbol{x})$ only by c_i, i.e.,

$$\Big| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \Big| \leq c_i$$

for all $(x_1, \ldots, x_n)$, $(x_1', \ldots, x_n') \in \Xi$.

With that, we can formulate an extension of McDiarmid [McD89] due to R. Combes:

Theorem 4.7 (McDiarmid). *Let $\xi_1, \ldots, \xi_n$ be independent random variables with values in D. Furthermore, let $f \colon D^n \to \mathbb{R}$ be **c**-bounded on $\Xi \subseteq D^n$, $m = \mathbb{E}(f(\xi_1, \ldots, \xi_n) | (\xi_1, \ldots, \xi_n) \in \Xi)$, and $\gamma = \mathbb{P}((\xi_1, \ldots, \xi_n) \notin \Xi)$. Then*

$$|f(\xi_1, \ldots, \xi_n) - m| \leq \sqrt{\frac{t}{2}} \|\boldsymbol{c}\|_2 + \gamma \|\boldsymbol{c}\|_1$$

with probability exceeding $1 - 2\gamma - 2\exp(-t)$.

Proof. [Com15, Theorem 2.1] with $\varepsilon = \gamma \|\boldsymbol{c}\|_1 + \sqrt{t/2} \|\boldsymbol{c}\|_2$. ■

4.2 Concentration inequalities for the spectral norm of (infinite) matrices

The bounds in this section concern sums of random matrices and bound the maximal singular values of them, which we will apply for the least squares matrix (2.1) later on. The following works for finite matrices and was shown in [Tro12, Theorem 1.1].

Lemma 4.8 (Matrix Chernoff). *Let $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n \in \mathbb{C}^{m \times m}$ be a finite sequence of independent, Hermitian, positive semi-definite random matrices satisfying $\lambda_{\max}(\boldsymbol{A}_i) \leq K$ almost surely. Furthermore, we set $\mu_{\min} :=$*

$\lambda_{\min}(\sum_{i=1}^n \mathbb{E}(\boldsymbol{A}_i))$ *and* $\mu_{\max} := \lambda_{\max}(\sum_{i=1}^n \mathbb{E}(\boldsymbol{A}_i))$. *Then*

$$\begin{aligned}
&\mathbb{P}\Big(\lambda_{\min}\Big(\sum_{i=1}^n \boldsymbol{A}_i\Big) \le (1-\varepsilon)\mu_{\min}\Big)\\
&\quad \le m \exp\Big(-\frac{\mu_{\min}}{K}(\varepsilon + (1-\varepsilon)\log(1-\varepsilon))\Big) \le m\exp\Big(-\frac{\mu_{\min}\varepsilon^2}{2K}\Big),\\
&\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{i=1}^n \boldsymbol{A}_i\Big) \ge (1+\varepsilon)\mu_{\max}\Big)\\
&\quad \le m \exp\Big(-\frac{\mu_{\max}}{K}(-\varepsilon + (1+\varepsilon)\log(1+\varepsilon))\Big) \le m\exp\Big(-\frac{\mu_{\max}\varepsilon^2}{3K}\Big)
\end{aligned}$$

for $0 \le \varepsilon \le 1$.

Proof. The first estimates are provided by [Tro12, Theorem 1.1]. Based on the Taylor expansion

$$(1+t)\log(1+t) = t + \sum_{k=2}^{\infty} \frac{(-1)^k}{k(k-1)} t^k\,,$$

which holds true for $t \in [-1,1]$, we further derive the inequalities

$$t + (1-t)\log(1-t) = \sum_{k=2}^{\infty} \frac{1}{k(k-1)} t^k \ge \frac{t^2}{2}$$

and
$$-t + (1+t)\log(1+t) = \sum_{k=2}^{\infty} \frac{(-1)^k}{k(k-1)} t^k \ge \frac{t^2}{2} - \frac{t^3}{6} \ge \frac{t^2}{3}$$

for the range $0 \le t \le 1$. ■

The next bound bounds the maximal singular value as well and works for infinite matrices.

Theorem 4.9 ([MU21, Proposition 3.8]). *Let* $\boldsymbol{u}^1, \dots, \boldsymbol{u}^n$ *be i.i.d. random sequences from* ℓ_2. *Let further* $n \ge 3$, $t > 0$, $M > 0$ *such that* $\|\boldsymbol{u}^i\|_2 \le M$ *almost surely and* $\mathbb{E}(\boldsymbol{u}^i \otimes \boldsymbol{u}^i) = \boldsymbol{\Lambda}$ *for all* $i = 1, \dots, n$. *Then*

$$\Big\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{u}^i \otimes \boldsymbol{u}^i\Big\|_{2\to 2} \le \frac{21(\log(n)+t)}{n} M^2 + 2\|\boldsymbol{\Lambda}\|_{2\to 2}$$

with probability exceeding $1 - 2^{3/4}\exp(-t)$.

Chapter 5

Subsampling of finite frames

In this chapter we deal with the basic linear algebra concept of frames. Their notion goes back to 1952 introduced by R. J. Duffin and A. C. Schaeffer [DS52]. Let H be a complex Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. A countable subset $(\boldsymbol{y}^i)_i$ in H is said to be a **frame** if there are constants $0 < A \le B < \infty$ such that

$$A\|\boldsymbol{x}\|^2 \le \sum_i |\langle \boldsymbol{x}, \boldsymbol{y}^i \rangle|^2 \le B\|\boldsymbol{x}\|^2 \quad \text{for all} \quad \boldsymbol{x} \in H\,. \tag{5.1}$$

We are mostly interested in frames consisting of finitely many elements of a finite dimensional Hilbert space H, see e.g. O. Christensen [Chr08] or P. Casazza and G. Kutyniok [CK13] for an introduction to frame theory. Systems of this kind may be represented by vectors $(\boldsymbol{y}^i)_{i=1}^M \subseteq \mathbb{C}^m$. For ensuring $0 < A$, we need the condition $M \ge m$ for the number of frame elements. The question of finding good subframes in such a system is rather fundamental and important for many applications ranging from graph sparsifiers [BSS09, SS11], the Kadison-Singer problem [MSS15, Wea04], to optimal discretization and sampling recovery of multivariate functions [DKU23, KU21a, KU21b, NSU21, LT22, Tem21, PU22]. In this context, let us also mention the possibility of generating "approximations" of Hadamard matrices, a problem which has been considered in [DR22] for example. Subsampling of a tight Hadamard frame $(\boldsymbol{y}^i)_{i=1}^M$, where all entries of the $\boldsymbol{y}^i$ are ± 1 (or in a complex setting of modulus 1), may lead to an almost square Hadamard-type matrix with good condition. For our goal of function approximation we want to achieve a small error in some norm defined on the whole domain, i.e., a quantity where every single point matters. The discrete set of points, in which we have given function evaluations for the sampling problem, is a subset thereof. Therefore, the task can be understood as a subsampling procedure while trying to keep the inherent information, which corresponds to the subsampling of frames.

Some results of this section are already published in [BSU23], of which we present the following contributions: We introduce a random subsampling technique in Theorem 5.1 yielding a logarithmic gap in terms of optimality in the number of frame elements, i.e., we have n of order $m \log m$ where n is the number of frame elements and m their dimension. This logarithmic gap is closed in Theorem 5.3 by the deterministic BSS algorithm to merely linear oversampling, i.e., n of order m. In the function approximation application later on it is of interest if the subsampling can be done without weights. For that reason we present approaches in Theorems 5.7 and 5.10 eliminating the weights saving the more important lower frame bound as this bound is crucial for the reconstruction of a vector $\boldsymbol{a}$ from its frame coefficients $\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle$. We conclude this section with some numerical examples in Section 5.4 to indicate the applicability of the theory.

5.1 Random weighted subsampling of finite frames

We begin with a random subsampling strategy that allows to extract "good" subframes of $\mathcal{O}(m \log m)$ elements out of any given frame in $\mathbb{C}^m$. This goes back to M. Rudelson and R. Vershynin [RV07], see also D. A. Spielman and N. Srivastava [SS11], where the goal was to efficiently find a low rank approximation of a given matrix such that the error with respect to the spectral norm remains small. The method is rather simple since it relies on a random sub-selection where the discrete probability mass ϱ_i for selecting one particular frame element $\boldsymbol{y}^i$ is directly linked to its contribution to the sum of the norms, i.e., the Frobenius norm $\|\boldsymbol{Y}\|_F^2$, see (4.1), of the matrix

$$\boldsymbol{Y} := \left[\begin{array}{c} (\boldsymbol{y}^1)^* \\ \hline \vdots \\ \hline (\boldsymbol{y}^M)^* \end{array}\right] \in \mathbb{C}^{M \times m} . \tag{5.2}$$

Note that for a given frame $(\boldsymbol{y}^i)_{i=1}^M \subseteq \mathbb{C}^m$, with $M \geq m$ and $m \in \mathbb{N}$, this matrix represents the analysis operator of the frame and that

$$mA \leq \operatorname{trace}(\boldsymbol{Y}^*\boldsymbol{Y}) = \|\boldsymbol{Y}\|_F^2 \leq m\|\boldsymbol{Y}^*\boldsymbol{Y}\|_{2\to 2} = m\lambda_{\max}(\boldsymbol{Y}^*\boldsymbol{Y}) \leq mB . \tag{5.3}$$

The main result of this section relies on a matrix Chernoff bound in Theorem 4.8 proven by Tropp [Tro12, Thm. 1.1]. It shows how one can randomly subsample a finite frame of arbitrary size in $\mathbb{C}^m$ to a weighted subframe with $\mathcal{O}(m \log m)$ elements while essentially keeping its stability properties.

Theorem 5.1. *Let $(\boldsymbol{y}^i)_{i=1}^M \subseteq \mathbb{C}^m$ be a frame with constants $0 < A \leq B < \infty$. Let further $p, t \in (0,1)$ and $n \in \mathbb{N}$ be such that*

$$n \geq \frac{3B}{At^2} m \log\Big(\frac{2m}{p}\Big) .$$

Drawing n indices $J \subseteq \{1, \ldots, M\}$ (with duplicates) i.i.d. according to the discrete probability density $\varrho_i = \|\boldsymbol{y}^i\|_2^2 / \|\boldsymbol{Y}\|_F^2$, $i \in \{1, \ldots, M\}$, then gives a rescaled random subframe $(\varrho_i^{-1/2} \boldsymbol{y}^i)_{i \in J}$ such that

$$(1-t)A\|\boldsymbol{a}\|_2^2 \leq \frac{1}{n} \sum_{i \in J} \Big| \Big\langle \boldsymbol{a}, \varrho_i^{-1/2} \boldsymbol{y}^i \Big\rangle \Big|^2 \leq (1+t)B\|\boldsymbol{a}\|_2^2 \quad \textit{for all } \boldsymbol{a} \in \mathbb{C}^m$$

with probability exceeding $1 - p$.

Proof. The result is a direct consequence of Tropp's concentration inequality in Lemma 4.8. For a randomly chosen index $i \in \{1, \ldots, M\}$ we define the rank-one random matrix $\boldsymbol{A}_i := \frac{1}{n} \varrho_i^{-1} (\boldsymbol{y}^i \otimes \boldsymbol{y}^i)$. Clearly, it holds

$$\lambda_{\max}(\boldsymbol{A}_i) = \lambda_{\max}\Big(\frac{1}{n} \varrho_i^{-1} (\boldsymbol{y}^i \otimes \boldsymbol{y}^i) \Big) = \frac{1}{n} \varrho_i^{-1} \|\boldsymbol{y}^i\|_2^2 = \frac{1}{n} \|\boldsymbol{Y}\|_F^2 \,.$$

Furthermore, having n independent copies $(\boldsymbol{A}_i)_{i \in J}$, we obtain

$$\sum_{i \in J} \mathbb{E} \boldsymbol{A}_i = \sum_{i \in J} \mathbb{E} \Big(\frac{1}{n} \varrho_i^{-1} (\boldsymbol{y}^i \otimes \boldsymbol{y}^i) \Big) = \sum_{i \in J} \frac{1}{n} \boldsymbol{Y}^* \boldsymbol{Y} = \boldsymbol{Y}^* \boldsymbol{Y} \,.$$

This gives for $\mu_{\min} := \lambda_{\min}(\sum_{i \in J} \mathbb{E} \boldsymbol{A}_i)$ and $\mu_{\max} := \lambda_{\max}(\sum_{i \in J} \mathbb{E} \boldsymbol{A}_i)$ that

$$\mu_{\min} = \lambda_{\min}(\boldsymbol{Y}^* \boldsymbol{Y}) \geq A \quad \text{and} \quad \mu_{\max} = \lambda_{\max}(\boldsymbol{Y}^* \boldsymbol{Y}) = \|\boldsymbol{Y}^* \boldsymbol{Y}\|_{2 \to 2} \,.$$

Since $\|\boldsymbol{Y}\|_F^2 = \operatorname{trace}(\boldsymbol{Y}^* \boldsymbol{Y})$, Lemma 4.8 and (5.3) gives

$$\begin{aligned} &\mathbb{P}\Big(\lambda_{\max}\Big(\frac{1}{n} \sum_{i \in J} \varrho_i^{-1} \boldsymbol{y}^i \otimes \boldsymbol{y}^i \Big) \geq (1+t)B \Big) \\ &\leq m \exp\Big(- \frac{n \|\boldsymbol{Y}^* \boldsymbol{Y}\|_{2 \to 2}}{\operatorname{trace}(\boldsymbol{Y}^* \boldsymbol{Y})} \frac{t^2}{3} \Big) \\ &\leq m \exp\Big(- \frac{n}{m} \frac{t^2}{3} \Big) \,. \end{aligned}$$

For the smallest eigenvalue things are a bit different. Here we obtain

$$\begin{aligned} &\mathbb{P}\Big(\lambda_{\min}\Big(\frac{1}{n} \sum_{i \in J} \varrho_i^{-1} \boldsymbol{y}^i \otimes \boldsymbol{y}^i \Big) \leq (1-t)A \Big) \\ &\leq m \exp\Big(- \frac{nA}{\operatorname{trace}(\boldsymbol{Y}^* \boldsymbol{Y})} \frac{t^2}{2} \Big) \\ &\leq m \exp\Big(- \frac{An}{Bm} \frac{t^2}{2} \Big) \,. \end{aligned}$$

For the probability of our assertion we need the complement of the two events above:

$$
\begin{aligned}
&1 - m \exp\Big(-\frac{n}{m}\frac{t^2}{3}\Big) - m \exp\Big(-\frac{An}{Bm}\frac{t^2}{2}\Big) \\
&\geq 1 - 2m \exp\Big(-\frac{An}{Bm}\frac{t^2}{3}\Big) \geq 1 - p\,,
\end{aligned}
$$

which follows from the assumption on n. ■

Theorem 5.1 shows that drawing frame elements $\boldsymbol{y}^i$, $i \in \{1, \dots, M\}$, according to the probabilities $\varrho_i := \|\boldsymbol{y}^i\|_2^2/(\sum_j \|\boldsymbol{y}^j\|_2^2)$ yields a reweighted subframe $(\varrho_i^{-1/2}\boldsymbol{y}^i)_{i\in J}$ with similar frame bounds, with high probability provided $|J| = \mathcal{O}(m \log m)$. In terms of computational complexity this strategy is very efficient. It is not optimal with respect to the number of frame elements, however.

Remark 5.2. *The rescaled random subframe $(\varrho_i^{-1/2}\boldsymbol{y}^i)_{i\in J}$ in Theorem 5.1 is an equal-norm frame. Thus, starting with a tight frame, we are able to construct an "almost tight" frame with unit-norm (UNTF). These are important in robust data transmission and have proven notoriously difficult to construct, cf. [CFM12, CK03].*

5.2 Deterministic weighted subsampling of finite frames

The shortcoming of the random subsampling being non-optimal with respect to the number of frame elements is dealt with in this section. We next present a deterministic subsampling algorithm for finite frames in $\mathbb{C}^m$ which we subsequently call (generalized) `BSS` algorithm. A version for real-valued tight frames in $\mathbb{R}^m$ was originally introduced by J. D. Batson, D. A. Spielman, and N. Srivastava in the context of graph sparsification [BSS09]. It allows to extract from any given finite frame in $\mathbb{C}^m$ a comparably well-conditioned reweighted subframe of cardinality $\mathcal{O}(m)$. This is the statement of Theorem 5.3 below which generalizes [BSS09, Thm. 3.1].

In contrast to related non-weighted subsampling results like in [NSU21, Thm. 2.3], which are all based on Weaver's theorem, a deep result equivalent to the famous Kadison-Singer theorem [MSS15], the proof of Theorem 5.3 is elementary and constructive. The underlying `BSS` algorithm lends itself to practical polynomial time implementation.

The careful analysis of the algorithm is published in [BSU23, Section 3] and is accredited to M. Schäfer. Because of its technical nature we omit it here and refer to the mentioned reference for the eager reader.

Theorem 5.3. *Let $(\boldsymbol{y}^i)_{i=1}^M \subseteq \mathbb{C}^m$ be a frame* (5.1) *with frame constants $0 < A \leq B < \infty$ and let $b > \kappa^2 \geq 1$ with*

$$\kappa := \Big(\frac{B}{2A} + \frac{1}{2}\Big) + \sqrt{\Big(\frac{B}{2A} + \frac{1}{2}\Big)^2 - 1}\,. \tag{5.4}$$

Then the `BSS` *algorithm, cf. Algorithm 1, computes a subset $J \subseteq \{1, \dots, M\}$ with $|J| \leq \lceil bm \rceil$ and nonnegative weights s_i, $i \in J$, such that*

$$A\|\boldsymbol{a}\|_2^2 \leq \sum_{i \in J} s_i |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \leq \gamma \cdot B\|\boldsymbol{a}\|_2^2 \quad \textit{for all} \quad \boldsymbol{a} \in \mathbb{C}^m \tag{5.5}$$

with

$$\gamma := \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)(\sqrt{b}-\kappa)}\,. \tag{5.6}$$

Remark 5.4. *(i) The* `BSS` *algorithm computes the index subset J and the corresponding weights s_i in $\mathcal{O}(bMm^3)$. An implementation and runtime analysis is given on page 87, see also [BSS09, Sec.3]. Better guarantees on the bound can be obtained by a "preconditioning" of the frame, given by Lemma 5.6. The resulting algorithm is called* `BSS`$^\perp$. *In* `BSS`$^\perp$ *also the restriction $b > \kappa^2$ can be evaded. Some empirical results are presented in Section 5.4.*

(ii) The theorem neither gives control over the weights s_i nor provides an unweighted version of itself. The latter would actually be useful for applications. We refer to [NSU21] for an unweighted result which is called "Weaver subsampling" and relies on the Kadison-Singer theorem [MSS15]. In Section 5.3 below we will use a special construction from Lemma 5.9 to deduce an unweighted version that preserves the lower frame bound, cf. Corollary 5.11.

In the following we give an insight into the principal structure of the `BSS` algorithm. The frame property of the vectors $(\boldsymbol{y}^i)_{i=1}^M$ can be formulated as

$$A\boldsymbol{I} \preceq \sum_{i=1}^M \boldsymbol{y}^i (\boldsymbol{y}^i)^* \preceq B\boldsymbol{I}\,,$$

where $\boldsymbol{I}$ denotes the identity matrix in $\mathbb{C}^{m \times m}$ and $\boldsymbol{A} \preceq \boldsymbol{B}$ denotes $\boldsymbol{B} - \boldsymbol{A}$ being positive semi-definite.

Furthermore, condition (5.5) of the subsampled frame can be rewritten as

$$A\boldsymbol{I} \preceq \sum_{i\in J} s_i \boldsymbol{y}^i(\boldsymbol{y}^i)^* \preceq \gamma \cdot B\boldsymbol{I}\,. \tag{5.7}$$

The idea of the BSS algorithm is to build the sum $\sum_{i\in J} s_i\boldsymbol{y}^i(\boldsymbol{y}^i)^*$ iteratively in $n := \lceil bm \rceil$ steps. Starting with the zero-matrix $\boldsymbol{A}^{(0)} := \boldsymbol{0}$, a sequence of Hermitian matrices

$$\boldsymbol{A}^{(0)},\ \boldsymbol{A}^{(1)},\ \boldsymbol{A}^{(2)},\ \dots,\ \boldsymbol{A}^{(n)} \tag{5.8}$$

is computed via rank-1 updates of the form

$$\boldsymbol{A}^{(k)} = \boldsymbol{A}^{(k-1)} + t^{(k)}\boldsymbol{y}^{i^{(k)}}(\boldsymbol{y}^{i^{(k)}})^*\,, \quad k \in \{1,\dots,n\}\,,$$

with suitably selected indices $i^{(k)} \in \{1,\dots,M\}$ and weights $t^{(k)} > 0$. After n iterations we have thus constructed a matrix $\boldsymbol{A}^{(n)}$ of the form

$$\boldsymbol{A}^{(n)} = \sum_{k=1}^{n} t^{(k)}\boldsymbol{y}^{i^{(k)}}(\boldsymbol{y}^{i^{(k)}})^* = \sum_{i\in J} \tilde{s}_i\boldsymbol{y}^i(\boldsymbol{y}^i)^* \tag{5.9}$$

where

$$\tilde{s}_i := \sum_{k:i^{(k)}=i} t^{(k)} \quad \text{and} \quad J := \left\{ i^{(k)} : k = 1,\dots,n \right\}.$$

Clearly, $|J| \le n = \lceil bm \rceil$. During the whole process the spectra of the constructed matrices $\boldsymbol{A}^{(k)}$ are controlled by means of so-called spectral barriers, i.e., numbers $l^{(k)}, u^{(k)} \in \mathbb{R}$ such that

$$\sigma(\boldsymbol{A}^{(k)}) \subseteq (l^{(k)}, u^{(k)})\,, \quad k \in \{0,\dots,n\}\,. \tag{5.10}$$

Note, that the barriers may be negative even though the eigenvalues of Hermitian matrices are always non-newgative.

Whereas the precise location of the eigenvalues of $\boldsymbol{A}^{(k)}$ may not be known, in this way we have enclosed their location in open intervals $(l^{(k)}, u^{(k)})$, in particular it holds

$$l^{(k)}\boldsymbol{I} \preceq \boldsymbol{A}^{(k)} \preceq u^{(k)}\boldsymbol{I}\,.$$

The algorithm starts with initial barriers $l^{(0)} < 0$ and $0 < u^{(0)}$ for $\boldsymbol{A}^{(0)} = \boldsymbol{0}$. From each step to the next, the barriers are then shifted to the right, most simply by certain fixed lengths $\delta_L > 0$ and $\delta_U > 0$. In the k-th iteration we thus have $l^{(k)} = l^{(0)} + k\delta_L$ and $u^{(k)} = u^{(0)} + k\delta_U$ (see Figure 5.1). For each $\boldsymbol{A}^{(k)}$

the indices and weights $i^{(k+1)}$ and $t^{(k+1)}$ are further chosen such that (5.10) remains valid for the updated matrix $\boldsymbol{A}^{(k+1)}$. Under these conditions, the final matrix $\boldsymbol{A}^{(n)}$ then has property (5.10) for

$$l^{(n)} = l^{(0)} + n\delta_L \quad \text{and} \quad u^{(n)} = u^{(0)} + n\delta_U \,.$$

For the "right" choice of $l^{(0)}$, $u^{(0)}$, δ_L, and δ_U we end up with final barriers satisfying

$$l^{(n)} > 0 \quad \text{and} \quad \frac{u^{(n)}}{l^{(n)}} \leq \gamma \cdot \frac{B}{A} \,.$$

This finally allows to rescale the weights $\tilde{s}_i$ in (5.9) appropriately, giving the desired weights s_i such that (5.7) is fulfilled.

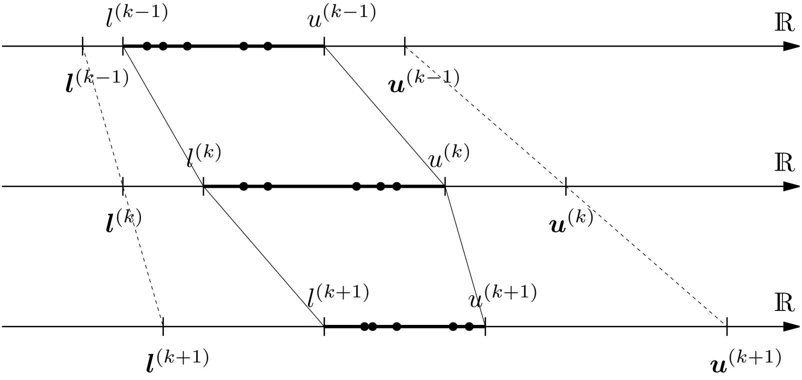


Figure 5.1: Spectral shifting via **constant** and *variable* barrier shifts.

The subsequent version, Algorithm 1, was implemented for the purpose of empirical analysis (see Section 5.4). Instead of fixed barrier shifts δ_L and δ_U it uses variable shifts $\delta_L^{(k)}$ and $\delta_U^{(k)}$ depending on the iteration step k.

Algorithm 1 BSS

Input: Frame $\boldsymbol{y}^1, \dots, \boldsymbol{y}^M \in \mathbb{C}^m$ with frame bounds $0 < A \leq B < \infty$;
Oversampling factor $b > \kappa^2$ with κ as in (5.4); Stability factor $\Delta \geq 0$.

Output: Nonnegative weights s_i such that $\sqrt{s_1}\boldsymbol{y}^1, \dots \sqrt{s_M}\boldsymbol{y}^M$ is a frame with $|\{i : s_i > 0\}| \leq \lceil bm \rceil$ and bounds $0 < A \leq B\gamma(1+\Delta) < \infty$. ($\gamma := \gamma(b,\kappa)$ is the value from (5.6).)

```
Put $n := \lceil bm \rceil$ and $\kappa := \kappa(A,B)$ as in (5.4). Further, set $\boldsymbol{A}^{(0)} := \boldsymbol{0}$,
    $$l^{(0)} := -m\frac{\sqrt{b}\kappa}{1+\Delta}, \quad u^{(0)} := m\frac{b+\sqrt{b}}{\sqrt{b}-1}\frac{B}{A},$$
    $$\delta_L^{(0)} := \frac{1}{1+\Delta}, \quad \delta_U^{(0)} := \frac{\sqrt{b}+1}{\sqrt{b}-1}\frac{B}{A}.$$
▷ $\boldsymbol{A}^{(0)} \in \mathbb{C}^{m\times m}$ is the zero matrix, $l^{(0)}$, $u^{(0)}$ associated lower and upper spectral barriers. The initial barrier shifts are given by $\delta_L^{(0)}$, $\delta_U^{(0)}$.
for $k = 1$ to $n$ do
    Compute the eigenvalues $\lambda_1^{(k-1)}, \dots, \lambda_m^{(k-1)}$ of $\boldsymbol{A}^{(k-1)}$.
    Compute the so-called lower and upper potentials
        $\epsilon_L^{(k-1)} := \Phi_{l^{(k-1)}}(\boldsymbol{A}^{(k-1)}) = \sum_{j=1}^m \left(\lambda_j^{(k-1)} - l^{(k-1)}\right)^{-1}$,
        $\epsilon_U^{(k-1)} := \Phi^{u^{(k-1)}}(\boldsymbol{A}^{(k-1)}) = \sum_{j=1}^m \left(u^{(k-1)} - \lambda_j^{(k-1)}\right)^{-1}$.
    Put $\delta_L^{(k-1)} := \left(\frac{1}{\delta_L^{(0)}} - \kappa\epsilon_L^{(0)} + \kappa\epsilon_L^{(k-1)}\right)^{-1}$ and
        $\delta_U^{(k-1)} := \left(\frac{1}{\delta_U^{(0)}} + \epsilon_U^{(0)} - \epsilon_U^{(k-1)}\right)^{-1}$.
    Increment $l^{(k-1)}$ and $u^{(k-1)}$:
        $l^{(k)} := l^{(k-1)} + \delta_L^{(k-1)}$, $u^{(k)} := u^{(k-1)} + \delta_U^{(k-1)}$.
    Compute the factors
        $f_L^{(k-1)} := \Phi_{l^{(k)}}(\boldsymbol{A}^{(k-1)}) = \sum_{j=1}^m \left(\lambda_j^{(k-1)} - l^{(k)}\right)^{-1}$ and
        $f_U^{(k-1)} := \Phi^{u^{(k)}}(\boldsymbol{A}^{(k-1)}) = \sum_{j=1}^m \left(u^{(k)} - \lambda_j^{(k-1)}\right)^{-1}$.
    for $j = 1$ to $M$ do
        Compute
```

```
            $L^{(k-1)}(\boldsymbol{y}^j) := \frac{(\boldsymbol{y}^j)^*(\boldsymbol{A}^{(k-1)} - l^{(k)}\boldsymbol{I})^{-2}\boldsymbol{y}^j}{f_L^{(k-1)} - \epsilon_L^{(k-1)}} - (\boldsymbol{y}^j)^*(\boldsymbol{A}^{(k-1)} - l^{(k)}\boldsymbol{I})^{-1}\boldsymbol{y}^j$,
            $U^{(k-1)}(\boldsymbol{y}^j) := \frac{(\boldsymbol{y}^j)^*(\boldsymbol{A}^{(k-1)} - u^{(k)}\boldsymbol{I})^{-2}\boldsymbol{y}^j}{\epsilon_U^{(k-1)} - f_U^{(k-1)}} - (\boldsymbol{y}^j)^*(\boldsymbol{A}^{(k-1)} - u^{(k)}\boldsymbol{I})^{-1}\boldsymbol{y}^j$.
10:         if $L^{(k-1)}(\boldsymbol{y}^j) - U^{(k-1)}(\boldsymbol{y}^j) \geq \frac{\Delta}{2M}\left(1 - \frac{1}{\sqrt{b}}\right)$ then
11:             denote this index by $i^{(k)}$.
12:             break
13:         end if
14:     end for
15:     Compute
            $t^{(k)} := 2\big(L^{(k-1)}(\boldsymbol{y}^{i^{(k)}}) + U^{(k-1)}(\boldsymbol{y}^{i^{(k)}})\big)^{-1}$,
            $\tilde{s}_{i^{(k)}} := \tilde{s}_{i^{(k)}} + t^{(k)}$,
            $\boldsymbol{A}^{(k)} := \boldsymbol{A}^{(k-1)} + t^{(k)}\boldsymbol{y}^{i^{(k)}}(\boldsymbol{y}^{i^{(k)}})^*$.
16: end for
17: return rescaled weights $s_i := \frac{1}{2}\Big(\frac{A}{l^{(n)}} + \frac{B\gamma(1+\Delta)}{u^{(n)}}\Big)\tilde{s}_i$
    for $i = 1, \ldots, M$.
```

A crucial step is the index selection in line 10. The condition guarantees that the chosen $i^{(k)}$ and the subsequently computed $t^{(k)}$ lead to a new updated matrix $\boldsymbol{A}^{(k)}$ (in line 15) which fulfills (5.10) as the matrices $\boldsymbol{A}^{(0)}, \ldots, \boldsymbol{A}^{(k-1)}$ did before. To avoid numerical issues in the selection, which might occur due to calculation inaccuracies, the stability parameter $\Delta \geq 0$ comes into play. It ensures that $L^{(k-1)}(\boldsymbol{y}^{i^{(k)}}) \geq U^{(k-1)}(\boldsymbol{y}^{i^{(k)}})$ can be verified, via the condition in line 10, in a numerically stable manner. In the numerical experiments we used $\Delta = 10^{-14}$ which worked for every tested experiment and does not seem to cause any issues.

Remark 5.5. *Algorithm 1 also works for fixed barrier shifts. We can skip the update in line 5 and always use $\delta_L^{(0)}$ and $\delta_U^{(0)}$ in the subsequent incrementation step in line 6. The advantage of variable shifts is a sharper containment of the spectrum (see illustration in Fig. 5.1).*

Runtime analysis of Algorithm 1. The computation of the singular value decomposition (SVD) of $\boldsymbol{A}^{k-1}$ in the k-th iteration step has a complexity of $\mathcal{O}(m^3)$. Having the SVD decomposition at hand, matrix-vector products with $(\boldsymbol{A}^{(k-1)} - l^{(k)}\boldsymbol{I})^{-1}$, $(\boldsymbol{A}^{(k-1)} - l^{(k)}\boldsymbol{I})^{-2}$, $(\boldsymbol{A}^{(k-1)} - u^{(k)}\boldsymbol{I})^{-1}$, and $(\boldsymbol{A}^{(k-1)} - u^{(k)}\boldsymbol{I})^{-2}$ are computable in $\mathcal{O}(m^2)$. To eventually decide, which index is selected in line 10, $L^{(k-1)}(\boldsymbol{y}^i)$ and $U^{(k-1)}(\boldsymbol{y}^i)$ in the worst case need to be computed for all $i \in \{1, \ldots, M\}$. This thus may require $\mathcal{O}(Mm^2)$ multiplication steps. All in all, taking into account $M \geq m$, each iteration can be performed in $\mathcal{O}(Mm^2)$ time. Since the number of iterations is $\lceil bm \rceil$, the total time of the algorithm is $\mathcal{O}(bMm^3)$. In our implementation we used a random procedure to traverse the indices $i \in \{1, \ldots, M\}$ and noticed that this speeds up the algorithm. In particular, often it suffices to check one or two elements to find one fulfilling the barrier condition, see Section 5.4, Experiment 3.

Instead of computing the singular value decomposition from scratch every iteration it is possible to update it continuously, cf. [BN79, MVDV92], which we have not implemented.

By including a preceding orthogonalization procedure, it is possible to allow arbitrarily small oversampling factors $b > 1$. Further the guarantees on the bounds improve. The modified algorithm is called $\texttt{BSS}^\perp$. It is based on the following simple observation.

Lemma 5.6. *For every matrix* $\boldsymbol{Y} \in \mathbb{C}^{M \times m}$ *with* $M \geq m$ *there is a matrix* $\tilde{\boldsymbol{Y}} \in \mathbb{C}^{M \times m}$ *such that*

$$\operatorname{range}(\tilde{\boldsymbol{Y}}) \supset \operatorname{range}(\boldsymbol{Y}), \quad \tilde{\boldsymbol{Y}}^*\tilde{\boldsymbol{Y}} = \boldsymbol{I}, \quad \textit{and} \quad \|\tilde{\boldsymbol{Y}}\|_F^2 = m\,,$$

where $\operatorname{range}(\tilde{\boldsymbol{Y}})$ *and* $\operatorname{range}(\boldsymbol{Y})$ *denote the range in* $\mathbb{C}^M$ *of the respective operators.*

Proof. The matrix $\tilde{\boldsymbol{Y}}$ is constructed by applying the Gram-Schmidt algorithm to the columns of $\boldsymbol{Y}$. If we end up with less than m vectors, which happens if $\operatorname{rank}(\boldsymbol{Y}) < m$, we orthogonally extend them, which is possible since $M \geq m$. ■

Algorithm 2 $\texttt{BSS}^\perp$

Input:	Frame $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^M \in \mathbb{C}^m$ with frame bounds $0 < A \leq B < \infty$; Oversampling factor $b > 1$; Stability factor $\Delta \geq 0$.
Output:	Nonnegative weights s_i such that $\sqrt{s_1}\boldsymbol{y}^1, \ldots \sqrt{s_M}\boldsymbol{y}^M$ is a frame with $\lvert\{i : s_i > 0\}\rvert \leq \lceil bm \rceil$ and bounds $0 < A \leq B\gamma(1 + \Delta) < \infty$. ($\gamma$ is the value from (5.6) for $\kappa = 1$.)

```
1: Let $\boldsymbol{Y} \in \mathbb{C}^{M\times m}$ be the matrix with rows $\boldsymbol{y}^1, \dots, \boldsymbol{y}^M$ and construct
   $\tilde{\boldsymbol{Y}} \in \mathbb{C}^{M\times m}$ as in Lemma 5.6 via Gram-Schmidt orthogonalization of the
   columns of $\boldsymbol{Y}$.
2: return weights $s_1, \dots, s_M$, calculated by applying BSS (Algorithm 1)
   to the rows of $\tilde{\boldsymbol{Y}}$.
```

Note that the rows $\tilde{\boldsymbol{y}}^1, \dots, \tilde{\boldsymbol{y}}^M$ of $\tilde{\boldsymbol{Y}}$, constructed in line 1 of Algorithm 2, form a tight frame with $A = B = 1$. Hence, in lines 2 Algorithm 1 can be applied for arbitrarily small $b > 1$. In fact, the frame property of the initial system $(\boldsymbol{y}^i)_{i=1}^M$ is not needed for this. It is possible to run $\texttt{BSS}^\perp$ for any input vector sequence $(\boldsymbol{y}^i)_{i=1}^M$ in $\mathbb{C}^m$, satisfying $M \geq m$. The returned weights s_i always fulfill $|\{i : s_i \neq 0\}| \leq \lceil bm \rceil$ and it always holds

$$\sum_{i=1}^{M} \left|\left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle\right|^2 \leq \sum_{i=1}^{M} s_i \left|\left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle\right|^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2}(1+\Delta) \sum_{i=1}^{M} \left|\left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle\right|^2 \tag{5.11}$$

for all $\boldsymbol{a} \in \mathbb{C}^m$. Assuming the input sequence was a frame, we then further deduce

$$A\|\boldsymbol{a}\|_2^2 \leq \sum_{i=1}^{M} s_i \left|\left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle\right|^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2}(1+\Delta) B \|\boldsymbol{a}\|_2^2 \quad \text{for all } \boldsymbol{a} \in \mathbb{C}^m\,.$$

To verify (5.11), let us first reformulate this inequality as

$$\|\boldsymbol{Y}\boldsymbol{a}\|_2^2 \leq \|\boldsymbol{S}^{\frac{1}{2}}(\boldsymbol{Y}\boldsymbol{a})|_J\|_2^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2}(1+\Delta)\|\boldsymbol{Y}\boldsymbol{a}\|_2^2\,, \tag{5.12}$$

where $J := \{i : s_i \neq 0\}$, $(\boldsymbol{Y}\boldsymbol{a})|_J$ is for the restricted vector $([\boldsymbol{Y}\boldsymbol{a}]_i)_{i\in J} \in \mathbb{C}^{|J|}$, and $\boldsymbol{S} := \operatorname{diag}(s_i)_{i\in J} \in \mathbb{C}^{|J|\times|J|}$. By Theorem 5.3, applying BSS (Algorithm 1) to $(\tilde{\boldsymbol{y}}^i)_{i=1}^M$ yields s_i such that $|J| = |\{i : s_i \neq 0\}| \leq \lceil bm \rceil$ and

$$\|\boldsymbol{a}\|_2^2 \leq \sum_{i=1}^{M} s_i |\langle \boldsymbol{a}, \tilde{\boldsymbol{y}}^i \rangle|^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2}(1+\Delta)\|\boldsymbol{a}\|_2^2 \quad \text{for all} \quad \boldsymbol{a} \in \mathbb{C}^m\,.$$

Therefore, by the orthogonality of $\tilde{\boldsymbol{Y}}$, for all $\boldsymbol{a} \in \mathbb{C}^m$

$$\|\tilde{\boldsymbol{Y}}\boldsymbol{a}\|_2^2 \leq \|\boldsymbol{S}^{\frac{1}{2}}(\tilde{\boldsymbol{Y}}\boldsymbol{a})|_J\|_2^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2}(1+\Delta)\|\tilde{\boldsymbol{Y}}\boldsymbol{a}\|_2^2\,.$$

Using $\operatorname{range}(\tilde{\boldsymbol{Y}}) \supset \operatorname{range}(\boldsymbol{Y})$, as guaranteed by Lemma 5.6, we may finally replace $\tilde{\boldsymbol{Y}}$ in this last inequality with the original $\boldsymbol{Y}$, leading to (5.12).

5.3 Non-weighted subsampling of finite frames

We now turn to non-weighted versions of the subsampling strategies from Sections 5.1 and 5.2. Our approach is to give estimates on the occurring weights. In this way, we sacrifice the upper frame bound in order to keep a lower frame bound not much smaller than the initial one. For many applications the lower frame bound is the important one as it ensures the reconstruction of any vector $\boldsymbol{a} \in \mathbb{C}^m$ from its frame coefficients $\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle$. Results in this section will be of the following form:

Given vectors $\boldsymbol{y}^1, \dots, \boldsymbol{y}^M$, we seek inequalities of the type

$$\frac{1}{M} \sum_{i=1}^{M} |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \leq \frac{C}{|J|} \sum_{i \in J} |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \quad \text{for all} \quad \boldsymbol{a} \in \mathbb{C}^m \tag{5.13}$$

for $J \subseteq \{1, \dots, M\}$ and some fixed constant $C > 0$. If the initial $\boldsymbol{y}^1, \dots, \boldsymbol{y}^M$ satisfy a lower frame bound, (5.13) gives that the vectors $\boldsymbol{y}^i$, $i \in J$, satisfy a lower frame bound as well.

For the non-weighted version of the random subsampling in Theorem 5.1 the construction of $\tilde{\boldsymbol{Y}}$ is covered by Lemma 5.6. We obtain the following result with $|J| = \mathcal{O}(m \log m)$.

Theorem 5.7. *Let $(\boldsymbol{y}^i)_{i=1}^M \subseteq \mathbb{C}^m$ be a frame and $c, p, t \in (0,1)$ and $n \in \mathbb{N}$ be such that*

$$n \geq \frac{3}{ct^2} m \log\left(\frac{m}{p}\right).$$

Drawing n indices $J \subseteq \{1, \dots, M\}$ (with duplicates) i.i.d. according to the discrete probability density $\varrho_i = (1-c)/M + c \cdot \|\tilde{\boldsymbol{y}}^{\boldsymbol{i}}\|_2^2/m$ gives

$$\frac{1}{M} \sum_{i=1}^{M} |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \leq \frac{1}{(1-c)(1-t)} \frac{1}{|J|} \sum_{i \in J} |\langle \boldsymbol{a}, \boldsymbol{y}^i \rangle|^2 \quad \textit{for all} \quad \boldsymbol{a} \in \mathbb{C}^m$$

with probability exceeding $1 - p$.

Proof. Similar to Theorem 5.1, the result follows from Tropp's concentration inequality in Lemma 4.8. Here it is applied to the random rank-1 matrices $\boldsymbol{A}_i := \frac{1}{n} \varrho_i^{-1} \tilde{\boldsymbol{y}}^i \otimes \tilde{\boldsymbol{y}}^i$, where $\tilde{\boldsymbol{y}}^1, \dots, \tilde{\boldsymbol{y}}^M \in \mathbb{C}^m$ are the rows of the matrix $\tilde{\boldsymbol{Y}}$ obtained according to Lemma 5.6 from $\boldsymbol{Y}$, the analysis operator (5.2) of $(\boldsymbol{y}^i)_{i=1}^M$. The matrices $\boldsymbol{A}_i$ satisfy

$$\lambda_{\max}(\boldsymbol{A}_i) = \frac{1}{n} \varrho_i^{-1} \|\tilde{\boldsymbol{y}}^i\|_2^2 \leq \frac{m}{cn}.$$

For $n = |J|$ independent copies $(\boldsymbol{A}_i)_{i\in J}$, due to the orthogonality of $\tilde{\boldsymbol{Y}}$, we further have

$$\sum_{i\in J} \mathbb{E}\boldsymbol{A}_i = \sum_{i\in J} \mathbb{E}\Big(\frac{1}{n}\varrho_i^{-1}\tilde{\boldsymbol{y}}^i \otimes \tilde{\boldsymbol{y}}^i\Big) = \sum_{i\in J}\frac{1}{n}\tilde{\boldsymbol{Y}}^*\tilde{\boldsymbol{Y}} = \boldsymbol{I}\,,$$

where $\boldsymbol{I}$ is the $m \times m$-dimensional identity matrix. Consequently, we have $\mu_{\min} = \lambda_{\min}(\sum_{i\in J}\mathbb{E}\boldsymbol{A}_i) = 1$. Using this in combination with Lemma 4.8 gives

$$\mathbb{P}\Big(\lambda_{\min}\Big(\frac{1}{n}\sum_{i\in J}\varrho_i^{-1}\tilde{\boldsymbol{y}}^i \otimes \tilde{\boldsymbol{y}}^i\Big) \le 1-t\Big) \le m\exp\Big(-\frac{cn}{m}\frac{t^2}{3}\Big)\,,$$

which is smaller than p by the assumption on n. Using $\varrho_i \ge (1-c)/M$, we obtain

$$\begin{aligned}\|\tilde{\boldsymbol{Y}}\boldsymbol{a}\|_2^2 = \|\boldsymbol{a}\|_2^2 &\le \frac{1}{1-t}\frac{1}{n}\sum_{i\in J}\varrho_i^{-1}|\langle \boldsymbol{a},\tilde{\boldsymbol{y}}^i\rangle|^2 \\ &\le \frac{M}{(1-c)(1-t)}\frac{1}{n}\|(\tilde{\boldsymbol{Y}}\boldsymbol{a})|_J\|_2^2\end{aligned}$$

for all $\boldsymbol{a} \in \mathbb{C}^m$ with probability exceeding $1-p$. By the arguments in (5.12) and after we may replace $\tilde{\boldsymbol{Y}}$ with the original $\boldsymbol{Y}$ to obtain the assertion. ■

Next, we assume that we have a Bessel sequence in $\mathbb{C}^m$ with elements that are norm-bounded from below. Applying Algorithm 2 (BSS$^\perp$) then yields a non-weighted inequality of type (5.13) with $|J| = \mathcal{O}(m)$. In Section 5.4 this algorithm is used in the experiments 1-3.

Lemma 5.8. *Let $(\boldsymbol{y}^i)_{i=1}^M$ be a Bessel sequence in $\mathbb{C}^m$, i.e., a set of vectors satisfying the upper bound in* (5.1) *for some $B > 0$. Further assume $M \ge m$ and $\|\boldsymbol{y}^i\|_2^2 \ge \beta m/M$ for some $\beta > 0$ and all $i \in \{1,\dots,M\}$. Then, for any $b > 1$, there exists a subset $J \subseteq \{1,\dots,M\}$ with $|J| \le \lceil bm\rceil$ (without duplicates) such that*

$$\frac{1}{M}\sum_{i=1}^M |\langle \boldsymbol{a},\boldsymbol{y}^i\rangle|^2 \le \frac{B}{\beta}\frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2}\frac{1}{m}\sum_{i\in J}|\langle \boldsymbol{a},\boldsymbol{y}^i\rangle|^2 \quad \textit{for all} \quad \boldsymbol{a} \in \mathbb{C}^m\,.$$

Proof. Applying Algorithm 2 (BSS$^\perp$) to the sequence $(\boldsymbol{y}^i)_{i=1}^M$ yields weights $s_i \ge 0$, where $|\{i : s_i \ne 0\}| \le \lceil bm\rceil$. Recall that, by the discussion of Algorithm 2, its application to any input sequence $(\boldsymbol{y}^i)_{i=1}^M$ is possible provided

$M \geq m$. We obtain (5.11). Taking into account the Bessel property of $(\boldsymbol{y}^i)_{i=1}^M$ and choosing $\Delta = 0$ in Algorithm 2 yields

$$\sum_{i=1}^{M} \left|\left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle\right|^2 \leq \sum_{i \in J} s_i \left|\left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle\right|^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2} B \|\boldsymbol{a}\|_2^2 \tag{5.14}$$

for $J := \{i : s_i \neq 0\}$ and all $\boldsymbol{a} \in \mathbb{C}^m$. Setting $\boldsymbol{a} = \boldsymbol{y}^j$ for $j \in J$, we obtain by the assumption $\|\boldsymbol{y}^j\|_2^2 \geq \beta m/M$ and the upper estimate in (5.14)

$$s_j \leq \frac{(\sqrt{b}+1)^2 B}{(\sqrt{b}-1)^2 \|\boldsymbol{y}^j\|_2^2} \leq \frac{B}{\beta} \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2} \frac{M}{m} \,.$$

Thus, by the lower estimate in (5.14), we obtain the assertion. ■

The condition on the norms $\|\boldsymbol{y}^i\|_2$ in Lemma 5.8 can be dropped with a more elaborate subsampling strategy, `PlainBSS` (see below) instead of `BSS`$^\perp$. The "preconditioning" in `PlainBSS` is based on Lemma 5.9 rather than Lemma 5.6. The final result is stated in Corollary 5.11. The price we pay for this is the dependence of the constant in terms of the oversampling factor b. It deteriorates to $(b-1)^{-3}$ while in the previous result it is $(b-1)^{-2}$.

Lemma 5.9. *Let $\boldsymbol{Y} \in \mathbb{C}^{M \times m}$ be a matrix and $K \in \{0, \ldots, M\}$. Then there is a matrix $\tilde{\boldsymbol{Y}} \in \mathbb{C}^{M \times m'}$ with $m' \in \{K, \ldots, K+m\}$ and rows $\tilde{\boldsymbol{y}}^1, \ldots, \tilde{\boldsymbol{y}}^M \in \mathbb{C}^{m'}$ such that*

$$\operatorname{range}(\tilde{\boldsymbol{Y}}) \supset \operatorname{range}(\boldsymbol{Y}) \,, \quad \tilde{\boldsymbol{Y}}^* \tilde{\boldsymbol{Y}} = \boldsymbol{I} \,, \quad \textit{and} \quad \|\tilde{\boldsymbol{y}}^i\|_2^2 \geq \frac{K}{M} \,,$$

where $\boldsymbol{I}$ is the $m' \times m'$-dimensional identity matrix.

Proof. Let us denote the columns of $\boldsymbol{Y}$ with $\boldsymbol{c}^1, \ldots, \boldsymbol{c}^m$. Further define columns in $\mathbb{C}^M$ by

$$\boldsymbol{d}^k = \frac{1}{\sqrt{M}} \left[\exp\left(2\pi \mathrm{i} k \frac{j}{M}\right) \right]_{j=1}^{M}$$

for $k = 0, \ldots, K-1$, which are the first K columns of a Fourier matrix. By construction the system $(\boldsymbol{d}^k)_{k=1}^K$ is orthonormal. It can hence be extended by appropriate vectors $\tilde{\boldsymbol{c}}^1, \ldots, \tilde{\boldsymbol{c}}^l$ to an orthonormal basis of

$$\operatorname{span}\{\boldsymbol{d}^1, \ldots, \boldsymbol{d}^K, \boldsymbol{c}^1, \ldots, \boldsymbol{c}^m\} \,.$$

Those can be constructed e.g. via the Gram-Schmidt algorithm. Finally, we set up

$$\tilde{\boldsymbol{Y}} := \left[\, \boldsymbol{d}^1 \,\middle|\, \cdots \,\middle|\, \boldsymbol{d}^K \,\middle|\, \tilde{\boldsymbol{c}}^1 \,\middle|\, \cdots \,\middle|\, \tilde{\boldsymbol{c}}^l \,\right] = \begin{bmatrix} (\tilde{\boldsymbol{y}}^1)^* \\ \hline \vdots \\ \hline (\tilde{\boldsymbol{y}}^M)^* \end{bmatrix} \in \mathbb{C}^{M \times (K+l)} ,$$

which fulfills the stated conditions. ■

Theorem 5.10. *Let $(\boldsymbol{y}^i)_{i=1}^M$ be a sequence of vectors in $\mathbb{C}^m$ and let $K \in \{0, \ldots, M\}$. Then, for any $b > 1$, a set of indices $J \subseteq \{1, \ldots, M\}$ (without duplicates) can be constructed (in polynomial time) such that $|J| \leq \lceil b(K+m) \rceil$ and*

$$\frac{1}{M} \sum_{i=1}^{M} \left| \left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle \right|^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2} \frac{m}{K} \frac{1}{m} \sum_{i \in J} \left| \left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle \right|^2 \quad \textit{for all} \quad \boldsymbol{a} \in \mathbb{C}^m . \tag{5.15}$$

Proof. We construct the vectors $\tilde{\boldsymbol{y}}^1, \ldots, \tilde{\boldsymbol{y}}^M$ according to Lemma 5.9. They form a tight frame in $\mathbb{C}^{m'}$ with $m' \in \{K, \ldots, K+m\}$ and $\|\tilde{\boldsymbol{y}}^i\|_2^2 \geq \frac{K}{M}$ for all $i \in \{1, \ldots, M\}$. We can thus apply Lemma 5.8 (`BSS`$^\perp$, which in effect is here `BSS`) with $B = 1$. We obtain a subset $J \subseteq \{1, \ldots, M\}$ with $|J| \leq \lceil bm' \rceil \leq \lceil b(K+m) \rceil$ (without duplicates) such that

$$\frac{1}{M} \sum_{i=1}^{M} \left| \left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle \right|^2 \leq \frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2} \frac{1}{K} \sum_{i \in J} \left| \left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle \right|^2 \quad \text{for all} \quad \boldsymbol{a} \in \mathbb{C}^m$$

which finishes the proof. ■

A result in terms of the "real" oversampling factor b' in Theorem 5.10, determined by $\lceil b'm \rceil = \lceil b(K+m) \rceil$, is given in Corollary 5.11.

Corollary 5.11. *Let $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^M \in \mathbb{C}^m$ be vectors with $m \in \mathbb{N}$. Further, take $b' > 1 + \frac{1}{m}$ and assume $M \geq \lceil b'm \rceil$. We then obtain indices $J' \subseteq \{1, \ldots, M\}$ with $|J'| \leq \lceil b'm \rceil$ such that*

$$\frac{1}{M} \sum_{i=1}^{M} \left| \left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle \right|^2 \leq 89 \frac{(b'+1)^2}{(b'-1)^3} \frac{1}{m} \sum_{i \in J'} \left| \left\langle \boldsymbol{a}, \boldsymbol{y}^i \right\rangle \right|^2 \quad \textit{for all} \quad \boldsymbol{a} \in \mathbb{C}^m .$$

Proof. The idea is to apply Theorem 5.10 for specifically chosen $K \in \mathbb{N}$ and $b > 1$, such that $\lceil b(m+K) \rceil \leq \lceil b'm \rceil$ for the given b' and the prefactor in (5.15) becomes small. Theorem 5.10 yields the prefactor

$$\frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)^2}\frac{m}{K} = \frac{(\sqrt{b}+1)^4}{(b-1)^2}\frac{m}{K} \leq 4\frac{(b+1)^2}{(b-1)^2}\frac{m}{K} =: C(K)\,.$$

Choosing b and K such that $b' = b\frac{m+K}{m}$ gives

$$\begin{aligned}&b = b'/(1+K/m)\,, \quad b+1 = (b'+1+K/m)/(1+K/m)\,,\\ &b-1 = (b'-1-K/m)/(1+K/m)\,,\end{aligned}$$

and hence

$$C(K) = 4\Big(\frac{b'+1+\frac{K}{m}}{b'-1-\frac{K}{m}}\Big)^2\frac{m}{K}\,. \tag{5.16}$$

We now choose $K^\star = \lceil\frac{(b'-1)m}{8}\rceil \in \{1,\dots,M\}$. Assuming $b' \geq 1+4/m$, we can then bound

$$\frac{b'-1}{8} \leq \frac{K^\star}{m} \leq \frac{b'-1}{4}\,.$$

Further, since $b'-1-K^\star/m > 0$, we arrive at the estimate

$$\begin{aligned}C(K^\star) &\leq 4\Big(\frac{b'+1+(b'-1)/4}{b'-1-(b'-1)/4}\Big)^2\frac{8}{b'-1}\\ &= \frac{32}{b'-1}\Big(\frac{5b'/4+3/4)}{3(b'-1)/4}\Big)^2\\ &\leq 32\Big(\frac{5}{3}\Big)^2\frac{(b'+1)^2}{(b'-1)^3}\\ &\leq 89\frac{(b'+1)^2}{(b'-1)^3}\,.\end{aligned} \tag{5.17}$$

Next, we consider the cases $b' = 1+2/m$ and $b' = 1+3/m$ separately, where in both $K^\star = 1$. The associated b's are given by $b = 1+1/(m+1)$ and $b = 1+2/(m+1)$. Further $1/m = (b'-1)/2$ and $1/m = (b'-1)/3$. Inserting these values into (5.16), we obtain estimates for $C(K^\star)$ as in (5.17). The prefactors, being 72 and 48, are even smaller than 89. Finally, to extend the estimate (5.17) to the whole range $b' > 1+\frac{1}{m}$, note that the right-hand side of (5.17) is increasing for $b' \searrow 1$. Taking into account $\lceil b'm \rceil = m+k+1$ for each $k \in \mathbb{N}$ and $b' \in (1+\frac{k}{m}, 1+\frac{k+1}{m}]$, we are finished. ■

Building on the proof of Corollary 5.11, we now formulate Algorithm 3 (PlainBSS). Like Algorithm 1 (BSS) and Algorithm 2 ($\mathtt{BSS}^{\perp}$), it is polynomial in time.

Algorithm 3 PlainBSS

```
Input:  Vectors $\boldsymbol{y}^1,\dots,\boldsymbol{y}^M \in \mathbb{C}^m$ with $m \in \mathbb{N}$ and $M \geq m+2$;
        Oversampling factor $b'$ s.t. $m+2 \leq \lceil b'm \rceil \leq M$; Stability
        factor $\Delta \geq 0$.
Output: Indices $J \subseteq \{1,\dots,M\}$ such that $|J| \leq \lceil b'm \rceil$ and
        $\frac{1}{M}\sum_{i=1}^{M} |\langle \boldsymbol{a}, \boldsymbol{y}^i\rangle|^2 \leq 89\frac{(b'+1)^2}{(b'-1)^3}\frac{1+\Delta}{m}\sum_{i\in J} |\langle \boldsymbol{a}, \boldsymbol{y}^i\rangle|^2$ .
Compute $K^\star$ and $b$ from $b'$ as in the proof of Corollary 5.11.
Construct the vectors $\tilde{\boldsymbol{y}}^1,\dots,\tilde{\boldsymbol{y}}^M \in \mathbb{C}^{m'}$ with $K = K^\star$ according to
    Lemma 5.9, where the initial $\boldsymbol{Y} \in \mathbb{C}^{M\times m}$ is the matrix (5.2) with rows
    $(\boldsymbol{y}^1)^*,\dots,(\boldsymbol{y}^M)^*$ .
Apply Algorithm 1 (BSS) to $\tilde{\boldsymbol{y}}^1,\dots,\tilde{\boldsymbol{y}}^M$ with oversampling factor $b$ and
    stability factor $\Delta$ to obtain weights $s_1,\dots,s_M$.
return indices $J := \{i : s_i \neq 0\}$.
```

For a better runtime, it might sometimes be advantageous to combine BSS subsampling with a preceding random subsampling step. Theorem 5.1 could be used, for instance, to quickly reduce the number of vectors to $\mathcal{O}(m\log m)$ in case of very large M. In the following corollary such a two-step procedure is used to construct a unit-norm frame with few (close to m) elements and well-behaved frame bounds. Here it is crucial that the BSS algorithm returns no duplicates, which is used in the proof.

Corollary 5.12. *Assume that the vectors $\boldsymbol{y}^1,\dots,\boldsymbol{y}^M \in \mathbb{C}^m, m \in \mathbb{N}$, form a tight frame and let $b' > 1 + \frac{1}{m}$. Further choose $p, t \in (0,1)$ and draw*

$$n := \left\lceil \frac{3}{t^2} m \log\left(\frac{2m}{p}\right)\right\rceil$$

indices $J \subseteq \{1,\dots,M\}$ (with duplicates) i.i.d. according to the discrete probability density $\varrho_i = \|\boldsymbol{y}^i\|_2^2/\|\boldsymbol{Y}\|_F^2$. In case $n > \lceil b'm \rceil$, those can further be subsampled using BSS (with oversampling factor b') giving $J' \subseteq J$ with $|J'| \leq \lceil b'm \rceil$ and a unit-norm frame $(\boldsymbol{y}^i/\|\boldsymbol{y}^i\|_2)_{i\in J'}$ satisfying

$$\frac{(1-t)(b'-1)^3}{89(b'+1)^2}\|\boldsymbol{a}\|_2^2 \leq \sum_{i\in J'} \left|\left\langle \boldsymbol{a}, \frac{\boldsymbol{y}^i}{\|\boldsymbol{y}^i\|_2}\right\rangle\right|^2 \leq (1+t)\left\lceil \frac{3\log(2m/p)}{t^2}\right\rceil \|\boldsymbol{a}\|_2^2$$

for all $\boldsymbol{a} \in \mathbb{C}^m$ *with probability exceeding* $1-p$. *Otherwise, when* $n \leq \lceil b'm \rceil$, *the frame* $(\boldsymbol{y}^i/\|\boldsymbol{y}^i\|_2)_{i\in J'}$ *with* $J'=J$ *already satisfies* $|J'| \leq \lceil b'm \rceil$ *and* (5.18) *for all* $\boldsymbol{a} \in \mathbb{C}^m$ *with probability exceeding* $1-p$.

Proof. By (5.3) and $(\boldsymbol{y}^i)_{i=1}^M$ forming a tight frame, we have $\|\boldsymbol{Y}\|_F^2 = mA$. By Theorem 5.1 we first obtain a subframe with $n=|J|$ elements such that

$$\frac{1-t}{m}\|\boldsymbol{a}\|_2^2 \leq \frac{1}{n}\sum_{i\in J}\Big|\Big\langle \boldsymbol{a}, \frac{\boldsymbol{y}^i}{\|\boldsymbol{y}^i\|_2}\Big\rangle\Big|^2 \leq \frac{1+t}{m}\|\boldsymbol{a}\|_2^2 . \tag{5.18}$$

Next, if we apply Algorithm 3 (`PlainBSS`) to this subframe, we obtain $J' \subseteq J$ with $|J'| \leq \lceil b'm \rceil$ such that

$$\frac{1}{n}\sum_{i\in J}\Big|\Big\langle \boldsymbol{a}, \frac{\boldsymbol{y}^i}{\|\boldsymbol{y}^i\|_2}\Big\rangle\Big|^2 \leq 89\frac{(b'+1)^2}{(b'-1)^3}\frac{1}{m}\sum_{i\in J'}\Big|\Big\langle \boldsymbol{a}, \frac{\boldsymbol{y}^i}{\|\boldsymbol{y}^i\|_2}\Big\rangle\Big|^2 ,$$

which is used in the lower frame bound. For the upper frame bound we use that J' has no duplicates, wherefore

$$\frac{1}{m}\sum_{i\in J'}\Big|\Big\langle \boldsymbol{a}, \frac{\boldsymbol{y}^i}{\|\boldsymbol{y}^i\|_2}\Big\rangle\Big|^2 \leq \Big\lceil\frac{3\log(2m/p)}{t^2}\Big\rceil\frac{1}{n}\sum_{i\in J}\Big|\Big\langle \boldsymbol{a}, \frac{\boldsymbol{y}^i}{\|\boldsymbol{y}^i\|_2}\Big\rangle\Big|^2 .$$

Here, the relation of n and m was used. Last, we use the upper frame bound (5.18) and obtain the assertion. ■

5.4 Numerical experiments

In this section we test `BSS`, $\texttt{BSS}^{\perp}$, and `PlainBSS` (Algorithms 1, 2, and 3) in practice. Note that there are further recent attempts to reduce the sampling budget in least squares approximations in practice, see [HNP22]. A survey on different probabilistic sampling strategies for sparse recovery of multivariate functions can be found in [ACDM22] (here especially Sec. 1.4 provides many further references). In addition, let us mention [AB22], where B. Adcock and S. Brugiapaglia give theoretical and empirical evidence of the near-optimal performance of simple Monte Carlo sampling for the recovery of smooth functions in high dimensions.

For the first three experiments, we use the rows of a d-dimensional Fourier matrix as initial frame, i.e.,

$$\boldsymbol{y}^i = \Big[\frac{1}{\sqrt{M}}\exp(2\pi \mathrm{i}\langle \boldsymbol{k}, \boldsymbol{x}^i\rangle)\Big]_{\boldsymbol{k}\in I} \quad \text{for} \quad i \in \{1,\dots,M\} , \tag{5.19}$$

where $I \subseteq \mathbb{Z}^d$ are $|I| = m$ frequencies determining the dimension of the frame elements and the points $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^M\} \subseteq \mathbb{T}^d$ determine their number. In the experiments, we will have a look at different choices for these frequencies I and points $\boldsymbol{X}$. Note that construction (5.19) gives an equal-norm frame.

As for the stability factor Δ we used the fixed number 10^{-14} which worked for every example we tested and does not seem to alter the results.

Experiment 1 We choose dimension $d = 2$ and, in the frequency domain, we use a so-called dyadic hyperbolic cross

$$I = H_R^d = \bigcup_{\substack{\boldsymbol{l} \in \mathbb{N}_0^d \\ \|\boldsymbol{l}\|_1 = R}} \hat{G}_{\boldsymbol{l}} \quad \text{with} \quad \hat{G}_{\boldsymbol{l}} = \bigtimes_{j=1}^{d} \hat{G}_{l_j}$$

and $$\hat{G}_l = \mathbb{Z} \cap (-2^{l-1}, 2^{l-1}]\,,$$

which occurs naturally when approximating in Sobolev spaces with mixed smoothness, cf. Section 3.6.2 or [DuTU18]. Here, we use $R = 6$, which results in 256 frequencies. In spatial domain, the canonical candidate are sparse grids:

$$S_R^d = \bigcup_{\substack{j \in \mathbb{N}_0^d \\ \|\boldsymbol{l}\|_1 = R}} G_{\boldsymbol{l}} \quad \text{with} \quad G_{\boldsymbol{l}} = \bigtimes_{j=1}^{d} G_{l_j} \quad \text{and} \quad G_l = 2^{-l}(\mathbb{Z} \cap [0, 2^l))\,.$$

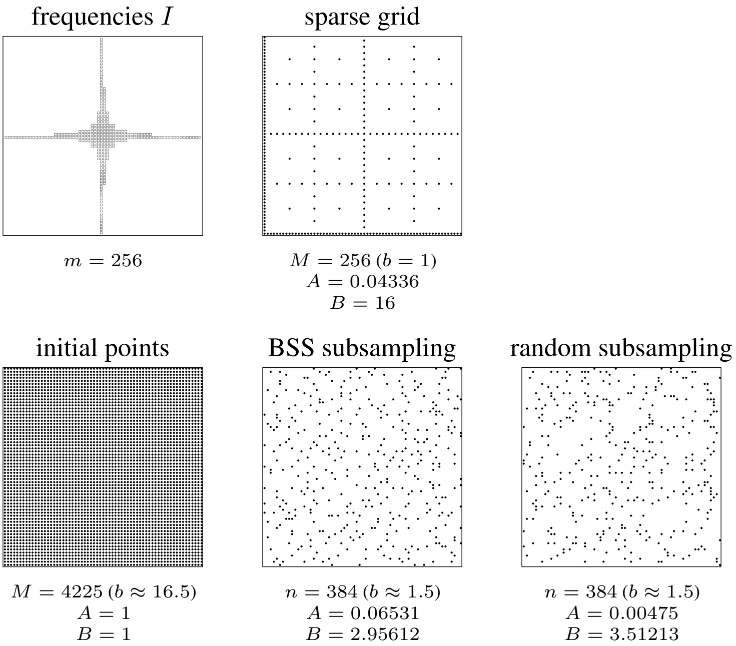
Sparse grids have the minimal amount of points $n = m$ and reconstruct every frequency $\boldsymbol{k} \in H_R^d$, i.e., $A > 0$. Precise estimates on the frame bounds of these matrices are found in [KK11, Thm. 3.1].

To test the `BSS` algorithm we use an initial 2-dimensional 65×65 equispaced grid

$$\boldsymbol{X} = \left\{ \frac{\boldsymbol{i}}{\sqrt[d]{M}} : \boldsymbol{i} \in \{0, \dots, \sqrt[d]{M} - 1\}^d \right\},$$

which has $M = 4\,225$ points and is exact ($A = B = 1$) for the M frequencies $\boldsymbol{k} \in \{-(\sqrt[d]{M} - 1)/2, \dots, (\sqrt[d]{M} - 1)/2\}^d$, cf. [PPST18, Sec. 4.4.3], in particular for the given dyadic hyperbolic cross. These initial frequencies and points can be seen in the first three graphs of Figure 5.2.

On the resulting frame constructed according to (5.19) we apply the unweighted `BSS` algorithm (discarding the weights s_i) with a target oversampling



frequencies I

$m = 256$

sparse grid

$M = 256\,(b = 1)$
$A = 0.04336$
$B = 16$

initial points

$M = 4225\,(b \approx 16.5)$
$A = 1$
$B = 1$

BSS subsampling

$n = 384\,(b \approx 1.5)$
$A = 0.06531$
$B = 2.95612$

random subsampling

$n = 384\,(b \approx 1.5)$
$A = 0.00475$
$B = 3.51213$

Figure 5.2: Two-dimensional experiment with sparse grid.

of $b = 1.5$ to obtain the subset J and compute the new frame bounds. For comparison, we draw a random subset (with replacement) of the same size and compute the frame bounds as well. Note, that we do not have theoretical bounds for these few random points. The results are depicted in the two rightmost graphs of Figure 5.2.

Since $\|\boldsymbol{y}^i\|_2^2 = m$, we obtain by Lemma 5.8 the theoretical lower frame bound $A = (\sqrt{b} - 1)^2/(\sqrt{b} + 1)^2 = 0.01021$ (cf. Lemma 5.8) where we observe $A = 0.06531$ in the experiment. This is better by a factor of 13 when compared to random subsampling, where we obtain a lower frame bound of $A = 0.00475$. Furthermore, the `BSS` algorithm gives a smaller upper frame constant than random subsampling, but this is not covered by our theory. The lower frame bound of the `BSS` subsampled points is bigger than the lower frame bound of the sparse grid. Even using the next biggest sparse grid with $n = 576$ points this still holds, as the frame bounds are $A = 0.04336$ and $B = 16$.

Following [KK11] the frame bounds worsen for the sparse grids in higher dimensions. We conducted the same experiment in five dimensions with dyadic hyperbolic cross with $m = 1002$ frequencies and included Frolov points for comparison as well, cf. [KOUU20]. The outcome is as follows:

		b	n	A	B
sparse grids	S_5^5	1.00	1002	0.00009	89.5249
	S_5^6	2.96	2972	0.00063	74.5446
	S_5^7	8.46	8472	0.00158	63.5213
Frolov points		1.02	1021	0.00008	3.13560
		2.05	2051	0.08128	2.14287
		4.08	4093	0.37502	1.79493
`BSS`		1.01	1013	0.00012	3.69835
		1.50	1503	0.04333	2.99637
		2.00	2004	0.10659	2.61729
		2.50	2505	0.16325	2.39153
		2.96	2966	0.20790	2.24841
		3.50	3507	0.25682	2.10744
		4.08	4089	0.30101	2.00187

We cannot set $b = 1$ with the `BSS` algorithm, but already for $b = 1.01$ we achieve a slightly better lower frame bound A than for the sparse grid. When b increases is where the `BSS` algorithm shows its advantage as the frame bounds become progressively better. In comparison to Frolov points the performance is similar but the Frolov points lack theoretical validation in this setting.

Experiment 2 As the components of the frame elements $\boldsymbol{y}^i$ are continuous in $\boldsymbol{x}^i$, we have similar frame elements for close points $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$. For the next experiment, we again are in dimension $d = 2$ and choose the full grid of frequencies $I = [-6, 6] \cap \mathbb{Z}^2$ with $m = 169$ frequencies for which the full grid of $13 \times 13 = 169$ points is barely exact. For the points we use two 13×13 point grids where one is slightly moved by $[0.01, 0.01]^\mathsf{T}$, which is depicted in the two leftmost plots of Figure 5.3. This setting is a union of two tight frames, itself a tight frame, where each element has a close duplicate which occur in pairs. A reasonable subsampling technique would pick at least one out of each pair. We set a target oversampling factor of $b = 1.1$ and apply the unweighted `BSS` algorithm and random subsampling for comparison. The results are depicted in the two rightmost graphs of Figure 5.3.

As in the first experiment, we have the theoretical lower frame bound $A = (\sqrt{b} - 1)^2/(\sqrt{b} + 1)^2 = 0.00057$ (cf. Lemma 5.8) where we observe $A = 0.02278$ in the experiment. For random subsampling we do not pick one

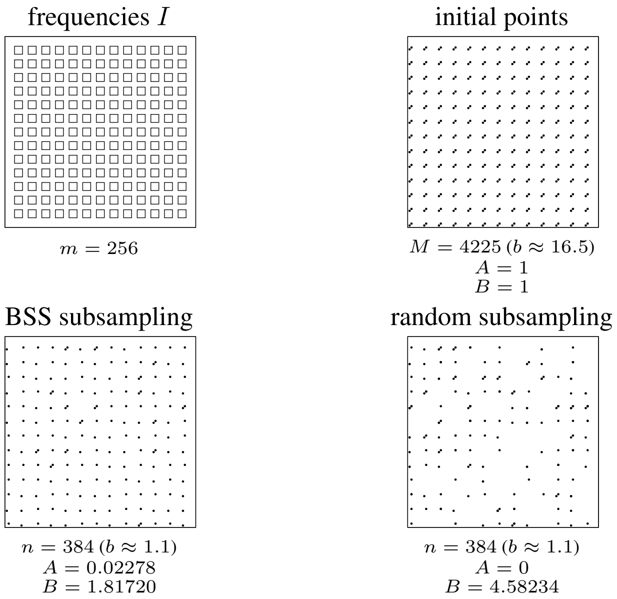


Figure 5.3: Two-dimensional experiment with frequencies on the grid.

frame element of each pair creating holes which spoil the lower frame bound. In fact, the subsampling is not even a frame anymore as $A = 0$.

Experiment 3 As our algorithms do not depend on the dimension, for the next experiment, we choose $d = 25$. In frequency domain we choose $m = 500$ random frequencies in $[-1000, 1000]^{25} \cap \mathbb{Z}^{25}$. In time domain we use two different choices:

- We use a full grid with $M = 2001^{25} > 10^{80}$ points, which is exact for all possible frequencies.
- We use $M = \lceil 6m \log(m) \rceil = 18\,644$ random points. In Theorem 6.4, we show that this gives frame bounds $A = 1/2$ and $B = 3/2$ with high probability.

For ten different choices of $b \in (1, 2]$ we use the unweighted `BSS` algorithm for the grid and unweighted $\texttt{BSS}^{\perp}$ for the random points. Note, that we do not compute all frame elements in advance but rather on the fly, which is possible

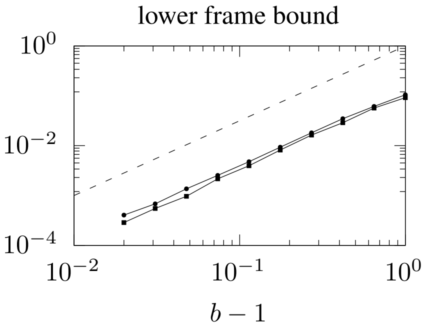


Figure 5.4: 25-dimensional experiment. Solid line with circles: lower frame bound A for the initial points being the full Grid. Solid line with squares: lower frame bound A for the initial points being drawn randomly. Dashed: $(b-1)^{3/2}$.

with the random subsampling and the BSS algorithm by addressing the frame elements by their index. For $\text{BSS}^{\perp}$ or PlainBSS this is not feasable since we construct an orthogonalized matrix of the original size $M \times m$.

We compute the new frame bounds and count the inner iterations (i.i.) of the BSS algorithm in line 6. Further, we compute the theoretical frame bounds $1/B \cdot (\sqrt{b}-1)^2/(\sqrt{b}+1)^2$ from Lemma 5.8. The results are shown in the table below and Figure 5.4.

		Grid points				**Random points**			
		$M = 2001^{25}$ ($b \approx 7 \cdot 10^{79}$)				$M = 18\,644$ ($b \approx 37$)			
		$A = B = 1$				$A = 0.70, B = 1.34$			
b	n	A	bound	B	i.i.	A	bound	B	i.i.
1.02	510	$3.72 \cdot 10^{-4}$	$2.45 \cdot 10^{-5}$	3.81	1.5	$2.70 \cdot 10^{-4}$	$1.83 \cdot 10^{-5}$	3.84	1.4
1.12	564	$5.59 \cdot 10^{-3}$	$8.02 \cdot 10^{-4}$	3.63	1.6	$4.68 \cdot 10^{-3}$	$5.99 \cdot 10^{-4}$	3.67	1.4
1.23	618	$1.46 \cdot 10^{-2}$	$2.67 \cdot 10^{-3}$	3.46	1.5	$1.26 \cdot 10^{-2}$	$2.00 \cdot 10^{-3}$	3.53	1.4
1.34	673	$2.63 \cdot 10^{-2}$	$5.33 \cdot 10^{-3}$	3.35	1.6	$2.28 \cdot 10^{-2}$	$3.98 \cdot 10^{-3}$	3.39	1.4
1.45	727	$3.92 \cdot 10^{-2}$	$8.58 \cdot 10^{-3}$	3.22	1.5	$3.56 \cdot 10^{-2}$	$6.40 \cdot 10^{-3}$	3.24	1.4
1.56	782	$5.14 \cdot 10^{-2}$	$1.23 \cdot 10^{-2}$	3.11	1.5	$4.89 \cdot 10^{-2}$	$9.15 \cdot 10^{-3}$	3.14	1.4
1.67	836	$6.63 \cdot 10^{-2}$	$1.63 \cdot 10^{-2}$	3.01	1.6	$5.77 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$	3.04	1.4
1.78	891	$7.90 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	2.94	1.5	$7.02 \cdot 10^{-2}$	$1.53 \cdot 10^{-2}$	3.01	1.4
1.89	940	$9.02 \cdot 10^{-2}$	$2.49 \cdot 10^{-2}$	2.90	1.6	$7.96 \cdot 10^{-2}$	$1.86 \cdot 10^{-2}$	2.91	1.4
2.00	1000	$1.02 \cdot 10^{-1}$	$2.94 \cdot 10^{-2}$	2.82	1.6	$9.55 \cdot 10^{-2}$	$2.20 \cdot 10^{-2}$	2.83	1.4

The message of this experiment is twofold:

- The rate of A for $b \to 1$ is cubic in the theoretical results, cf. Lemma 5.8 and Theorem 5.10. In this experiment we observe the rate of $3/2$ which is even smaller that the bound for the weighted `BSS` algorithm, cf. Theorem 5.3.
- As the number of points in the grid is larger than the estimated number of atoms in the observable universe, the `BSS` algorithm could have a slow runtime. The only difference in the computational effort could originate from the iterations in the inner loop of the `BSS` algorithm. From our theory we obtain M iterations in the worst case whereas we observe 1.5 iterations on average in both experiments.

Experiment 4 In this experiment we deal with two-dimensional hyperbolic Chui-Wang wavelets $\psi_{\boldsymbol{j},\boldsymbol{k}}$, which are compactly supported, piecewise linear, and $L_2([0,1]^d)$-normalized, see for instance [LPU23] for the precise construction. We define the index sets

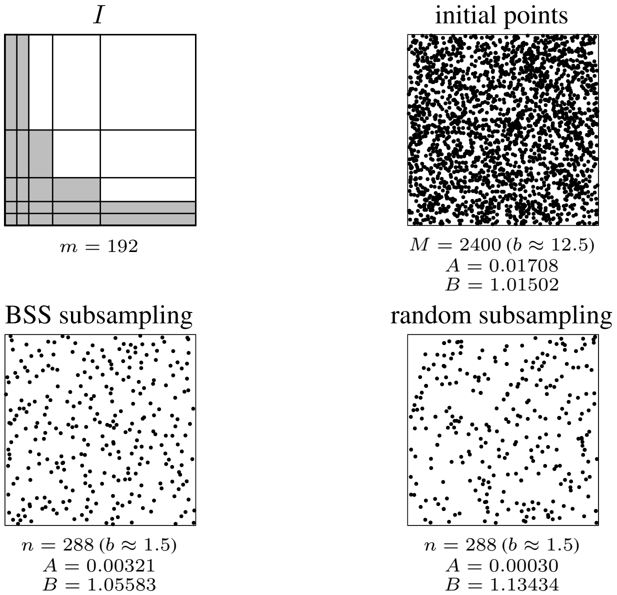
$$\mathcal{J}_n = \{(\boldsymbol{j},\boldsymbol{k}) \in \mathbb{N}_{-1}^d \times \mathbb{Z}^d : \boldsymbol{j} \geq -\mathbf{1}, |\boldsymbol{j}|_1 \leq N, \boldsymbol{k} \in I_{\boldsymbol{j}}\} \tag{5.20}$$

and

$$I_{\boldsymbol{j}} = \prod_{i=1}^{d} \begin{cases} \{0,1,\dots 2^{j_i}-1\} & \text{for } j_i \geq 0, \\ \{0\} & \text{for } j_i = -1. \end{cases}$$

The projection on the $\boldsymbol{j}$-component of this index set is displayed in the first picture in Figure 5.5 with $N=3$. Drawing sufficiently many (M) points i.i.d. and uniformly at random ($M = \mathcal{O}(|\mathcal{J}_n|\log(|\mathcal{J}_n|))$) it has been shown in [LPU23] that the corresponding frame $([\psi_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}^i)]_{\boldsymbol{j},\boldsymbol{k}})_{i=1}^M$ has reasonably good frame bounds (see the second picture in Figure 5.5). In the previous experiments we only dealt with equal-norm frames. This is not given anymore in this particular frame such that we are forced to apply `PlainBSS` to extract a reasonable subframe with $b \approx 1.5$. The resulting points can be seen in the third picture of Figure 5.5.

The lower frame bound of the subsampled points can be estimated by Corollary 5.11: $A \leq \frac{0.01708(b-1)^3}{89(b+1)^2} \approx 3.84 \cdot 10^{-6}$ for $b = 1.5$. In practice we obtain a subsampled frame bound of $A = 3.21 \cdot 10^{-3}$, which indicates that the theoretical constants may be improved. Further, in the comparison to random subsampling, `PlainBSS` is better by a factor of 10 (last picture of Figure 5.5) and the upper frame bound does not differ much. Overall, this



I

$m = 192$

initial points

$M = 2400 \; (b \approx 12.5)$
$A = 0.01708$
$B = 1.01502$

BSS subsampling

$n = 288 \; (b \approx 1.5)$
$A = 0.00321$
$B = 1.05583$

random subsampling

$n = 288 \; (b \approx 1.5)$
$A = 0.00030$
$B = 1.13434$

Figure 5.5: Two-dimensional hyperbolic wavelet transform.

experiment demands for the tricky construction of Lemma 5.9 and shows its stable applicability.

Chapter 6

L_2-Marcinkiewicz-Zygmund (MZ) inequalities

In Section 2.3 we have seen that the singular values of the system matrix $\boldsymbol{L}$ are important for, e.g. the runtime of the LSQR algorithm. Now we give two equivalent characterizations for bounds of the singular values, in order to apply techniques from Chapter 5 and get a vast family of initial points. One of these characterizations goes back to J. Marcinkiewicz and A. Zygmund [MZ37] and establishes a connection between the continuous L_2-norm and point evaluations of functions. A set of points $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq D$ and weights $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n) \in [0,\infty)^{n\times n}$ fulfills an L_2-Marcinkiewicz-Zygmund (MZ) inequality with constants $0 < A \leq B < \infty$ for a finite-dimensional function space V, if

$$A\|f\|_{L_2}^2 \leq \sum_{i=1}^{n} \omega_i |f(\boldsymbol{x}^i)|^2 \leq B\|f\|_{L_2}^2 \quad \text{for all} \quad f \in V\,, \tag{6.1}$$

where, for ϱ_T a σ-finite measure and some domain $D \subseteq \mathbb{R}^d$,

$$\|f\|_{L_2} \coloneqq \sqrt{\langle f, f\rangle_{L_2}} \quad \text{with} \quad \langle f, g\rangle_{L_2} \coloneqq \int_D f(\boldsymbol{x})\overline{g(\boldsymbol{x})}\,\mathrm{d}\varrho_T(\boldsymbol{x})\,.$$

Clearly, such points together with appropriate subspaces V are good for sampling recovery in $L_2(D, \varrho_T)$. For a systematic study of MZ inequalities (also for $p \neq 2$) we refer to the recent series of papers by V. N. Temlyakov and coauthors, see for instance [Tem18, KKLT22]. We orient this chapter on [BKPU23, Section 2].

In the least squares approximation we use ansatz functions $\eta_1, \dots, \eta_{m-1}$, cf. Section 2.2. Throughout this section we assume that these ansatz functions are orthonormal with respect to the $L_2(D, \varrho_T)$ inner product. We investigate the recovery of functions $f\colon D \to \mathbb{C}$ where we assume f to be element of the function space $V = \operatorname{span}\{\eta_1, \dots, \eta_{m-1}\}$, i.e., we have the representation

$$f(\boldsymbol{x}) = \sum_{k=1}^{m-1} a_k \eta_k(\boldsymbol{x}) \quad \text{with} \quad \boldsymbol{a} = (a_k)_{k=1}^{m-1} = (\langle f, \eta_k\rangle_{L_2})_{k=1}^{m-1}\,. \tag{6.2}$$

For the recovery we have given samples $\boldsymbol{f} = (f(\boldsymbol{x}^1), \dots, f(\boldsymbol{x}^n))^\mathsf{T} \in \mathbb{C}^n$ in points $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq D$.

If the singular values of $\boldsymbol{W}^{1/2}\boldsymbol{L}$ are non-zero, we immediately obtain that the least squares approximation $S_V^{\boldsymbol{X}}$ from Section 2.2 reconstructs all functions $f \in V$. The stability of the reconstruction as well as the number of iterations of a conjugate gradient method used to solve the system of equations heavily depends on the condition number of the matrix $(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}^{1/2}$ as we have seen in Theorem 2.3.

Lemma 6.1. *[KUV21, Proposition 3.1] Let $\boldsymbol{W}^{1/2}\boldsymbol{L} \in \mathbb{C}^{n\times(m-1)}$ be a matrix with $m-1 \le n$ with non-zero singular values. Then*

$$\sigma_{\max}\Big((\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}^{1/2}\Big) = \frac{1}{\sigma_{\min}(\boldsymbol{W}^{1/2}\boldsymbol{L})}$$

and

$$\sigma_{\min}\Big((\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}^{1/2}\Big) = \frac{1}{\sigma_{\max}(\boldsymbol{W}^{1/2}\boldsymbol{L})}\,.$$

This motivates having a look at lower and upper bounds for the singular values of $\boldsymbol{W}^{1/2}\boldsymbol{L}$. The following lemma gives different characterizations for these bounds.

Lemma 6.2. *Let $\eta_1, \dots, \eta_{m-1} \colon D \to \mathbb{C}$ be an L_2-orthonormal basis of a finite-dimensional function space V, let $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq D$ be points and $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n) \in [0,\infty)^{n\times n}$ weights. Then the following statements are equivalent:*

(i) the singular values of the matrix $\boldsymbol{W}^{1/2}\boldsymbol{L} \in \mathbb{C}^{n\times(m-1)}$, where $\boldsymbol{L}$ and $\boldsymbol{W}$ are defined in (2.1)*, lie in the interval $[\sqrt{A}, \sqrt{B}]$, i.e.,*

$$A\|\boldsymbol{a}\|_2^2 \le \|\boldsymbol{W}^{1/2}\boldsymbol{L}\boldsymbol{a}\|_2^2 \le B\|\boldsymbol{a}\|_2^2 \quad \textit{for all} \quad \boldsymbol{a} \in \mathbb{C}^{m-1}\,;$$

(ii) the rows $\sqrt{\omega_i}(\eta_k(\boldsymbol{x}^i))_{k=1}^{m-1} \in \mathbb{C}^{m-1}$ for $i = 1, \dots, n$, of the matrix $\boldsymbol{W}^{1/2}\boldsymbol{L}$ form a frame with bounds A and B, i.e.,

$$A\|\boldsymbol{a}\|_2^2 \le \sum_{i=1}^{n} |\langle \boldsymbol{a}, \sqrt{\omega_i}(\eta_k(\boldsymbol{x}^i))_{k=1}^{m-1}\rangle|^2 \le B\|\boldsymbol{a}\|_2^2 \quad \textit{for all} \quad \boldsymbol{a} \in \mathbb{C}^{m-1}\,;$$

(iii) the points $\boldsymbol{X}$ and weights $\boldsymbol{W}$ form an L_2-MZ inequality for $V =$ $\operatorname{span}\{\eta_1, \dots, \eta_{m-1}\}$ with bounds A and B, i.e.,

$$A\|f\|_{L_2}^2 \le \sum_{i=1}^{n} \omega_i |f(\boldsymbol{x}^i)|^2 \le B\|f\|_{L_2}^2 \quad \textit{for all} \quad f \in V\,.$$

Proof. **Step 1.** To show the equivalence of (i) and (ii) we rewrite the matrix-vector product $\boldsymbol{W}^{1/2}\boldsymbol{L}\overline{\boldsymbol{a}}$ using the Euclidean inner products of $\boldsymbol{a}$ and the rows of $\boldsymbol{W}^{1/2}\boldsymbol{L}$ as follows

$$\begin{aligned} A\|\boldsymbol{a}\|_2^2 = A\|\overline{\boldsymbol{a}}\|_2^2 \le \|\boldsymbol{W}^{1/2}\boldsymbol{L}\overline{\boldsymbol{a}}\|_2^2 &= \sum_{i=1}^{n} |\langle \boldsymbol{a}, \sqrt{\omega_i}(\eta_k(\boldsymbol{x}^i))_{k=1}^{m-1}\rangle|^2 \\ &\le B\|\overline{\boldsymbol{a}}\|_2^2 = B\|\boldsymbol{a}\|_2^2 \,. \end{aligned}$$

This immediately shows the equivalence of the two stated conditions.

Step 2. Now we show the equivalence of (ii) and (iii). Using the series expansion (6.2) of f, we have

$$\sum_{i=1}^{n} |\langle \overline{\boldsymbol{a}}, \sqrt{\omega_i}(\eta_k(\boldsymbol{x}^i))_{k=1}^{m-1}\rangle|^2 = \sum_{i=1}^{n} \omega_i \Big| \sum_{k=1}^{m-1} a_k \eta_k(\boldsymbol{x}^i) \Big|^2 = \sum_{i=1}^{n} \omega_i |f(\boldsymbol{x}^i)|^2$$

and further using Parseval's identity, we obtain

$$\|\overline{\boldsymbol{a}}\|_2^2 = \|\boldsymbol{a}\|_2^2 = \Big\| \sum_{k=1}^{m-1} a_k \eta_k(\boldsymbol{x}^i) \Big\|_{L_2}^2 = \|f\|_{L_2}^2 \,.$$

Plugging these two formulas into the frame condition, we obtain a reformulation in terms of functions, which is an L_2-MZ inequality for the function space V with bounds A and B. ■

The L_2-MZ characterization supports the availability of points with well-behaved least squares matrices. With Lemma 6.2, using a set of points $\boldsymbol{X}$ coming from an L_2-MZ inequality with well-behaved constants, we automatically have a good point set to use in least squares approximation, and vice versa. Point sets fulfilling L_2-MZ inequalities are widely available and well-studied, cf. [MNW01, KKP07, FM11, CDL13, CM17, Tem18]. Furthermore, the next theorem shows an equivalence to an exact integration condition when the constants in the L_2-MZ inequality coincide. This widens the applicability even further, cf. [CN07, KPV15, Tre13].

Theorem 6.3. *Points $\boldsymbol{X} \subseteq D$ and weights $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n) \in [0,\infty)^{n\times n}$ obey an L_2-MZ inequality* (6.1) *on $V = \operatorname{span}\{\eta_k\}_{k=1,\dots,m-1}$ with constants $A = B$ if and only if we have exact quadrature on $V \cdot \overline{V} := \{f = \overline{g} \cdot h : g, h \in V\}$, i.e., we have*

$$\sum_{i=1}^{n} \omega_i g(\boldsymbol{x}^i)\overline{h(\boldsymbol{x}^i)} = \frac{1}{A}\int_D g(\boldsymbol{x})\overline{h(\boldsymbol{x})}\,\mathrm{d}\nu(\boldsymbol{x}) \quad \textit{for all} \quad g, h \in V \,.$$

Proof. Starting with the L_2-MZ inequality, we have

$$\sum_{i=1}^{n} \omega_i |f(\boldsymbol{x}^i)|^2 = A \int_D |f(\boldsymbol{x})|^2 \,\mathrm{d}\nu(\boldsymbol{x}) \quad \text{for all} \quad f \in V\,.$$

By the parallelogram law the corresponding inner products also coincide, which is on direction of the assertion. The reverse is achieved by using $g = h$. ■

6.1 Random L_2-MZ inequalities

An universal approach to obtain an L_2-MZ inequality is to use random points with respect to a certain measure. This was established in [CM17, Thm. 2.1], which is a key element in many theorems presented in this thesis and is widely used, cf. [NSU21, Theorem 2.3], [MU21, Theorem 5.1], [DC22b, Lemma 2.1], or [BSU23, Theorem 2.1].

Lemma 6.4. *Let $t \geq 0$, $n \in \mathbb{N}$, $\boldsymbol{x}^1, \dots, \boldsymbol{x}^n \in D$ be points drawn according to a probability measure $\mathrm{d}\varrho_S = 1/\beta \,\mathrm{d}\varrho_T$ with $\mathrm{d}\varrho_T$ a second measure and $\beta\colon D \to [0,\infty]$ the Radon-Nikodym derivative. Let further, V be an $m-1$-dimensional function space with an $L_2(D,\varrho_T)$-orthonormal basis $\eta_1, \dots, \eta_{m-1}$ with m satisfying*

$$10\|\beta(\cdot)N(V,\cdot)\|_\infty(\log(m-1)+t) \leq n\,,$$

with $N(V,\cdot)$ the Christoffel function (3.14). *Then*

$$\frac{n}{2}\|g\|^2_{L_2(D,\varrho_T)} \leq \sum_{i=1}^{n} \beta(\boldsymbol{x}^i)|g(\boldsymbol{x}^i)|^2 \leq \frac{3n}{2}\|g\|^2_{L_2(D,\varrho_T)} \quad \textit{for all} \quad g \in V\,,$$

where each inequality holds with probability exceeding $1 - \exp(-t)$.

The proof ideas go back to [CDL13, Thm. 1] and [CM17, Thm. 2.1] but for the sake of readability we state it here as well.

Proof. The result is a direct consequence of Tropp's result in Lemma 4.8. For a randomly chosen point $\boldsymbol{x}^i$ we define the random rank-one matrix $\boldsymbol{A}_i = \frac{1}{n}\beta(\boldsymbol{x}^i)(\boldsymbol{y}^i \otimes \boldsymbol{y}^i)$ with $\boldsymbol{y}^i = (\eta_1(\boldsymbol{x}^i), \dots, \eta_{m-1}(\boldsymbol{x}^i))^\mathsf{T}$. By construction, it holds

$$\sum_{i=1}^{n} \boldsymbol{A}_i = \boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L}$$

and by the orthogonality of η_k

$$\Big(\mathbb{E}(\boldsymbol{A}_i)\Big)_{k,l} = \frac{1}{n}\int_D \eta_k(\boldsymbol{x})\overline{\eta_l(\boldsymbol{x})}\beta(\boldsymbol{x})\beta^{-1}(\boldsymbol{x})\; d\varrho_T(\boldsymbol{x}) = \frac{\delta_{k,l}}{n}\,,$$

which gives $\mathbb{E}\Big(\sum_{i=1}^n \boldsymbol{A}_i\Big) = I_{m-1\times m-1}$ and, therefore, $\mu_{\max} = \mu_{\min} = 1$. Further, we have

$$\lambda_{\max}\Big(\frac{1}{n}\beta(\boldsymbol{x}^i)(\boldsymbol{y}^i \otimes \boldsymbol{y}^i)\Big) = \frac{1}{n}\beta(\boldsymbol{x}^i)\|\boldsymbol{y}^i\|_2^2 \leq \frac{1}{n}\|\beta(\cdot)N(V,\cdot)\|_\infty\,.$$

Lemma 4.8 with $t = 1/2$ then gives the lower bound

$$\mathbb{P}\Big(\lambda_{\min}\Big(\frac{1}{n}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L}\Big) \leq \frac{1}{2}\Big) \leq (m-1)\exp\Big(-\frac{n}{10}\|\beta(\cdot)N(V,\cdot)\|_\infty^{-1}\Big)\,,$$

which is smaller than $\exp(-t)$ by the assumption on m. Using Theorem 6.2 we obtain the formulation with respect to functions.

The bound for the largest eigenvalue works analogue. ■

The important part for the reconstruction of functions is the lower L_2-MZ inequality, cf. Chapter 7. For the weighted result above we have both inequalities. Aiming for an unweighted result, the next corollary states a construction of points with logarithmic oversampling and equal weights satisfying the lower L_2-MZ inequality.

Corollary 6.5. *Let V be an $m-1$-dimensional function space with an orthonormal basis $\eta_1,\dots,\eta_{m-1}$ in $L_2(D,\varrho_T)$. Let $t \geq 0$ and $n \in \mathbb{N}$ be such that*

$$20(m-1)(\log(m-1)+t) \leq n\,.$$

Then the points $\boldsymbol{x}^1,\dots,\boldsymbol{x}^n$ drawn according to a probability measure

$$\mathrm{d}\varrho_S(\boldsymbol{x}) = \frac{1}{2\varrho_T(D)} + \frac{1}{2}\frac{\sum_{k=1}^{m-1}|\eta_k(\boldsymbol{x})|^2}{m-1}\,\mathrm{d}\varrho_T$$

fulfill the unweighted lower L_2-MZ inequality with probability exceeding $1-\exp(-t)$, i.e.,

$$\|g\|_{L_2(D,\varrho_T)}^2 \leq \frac{4\varrho_T(D)}{n}\sum_{i=1}^n |g(\boldsymbol{x}^i)|^2 \quad \textit{for all} \quad g \in V\,. \tag{6.3}$$

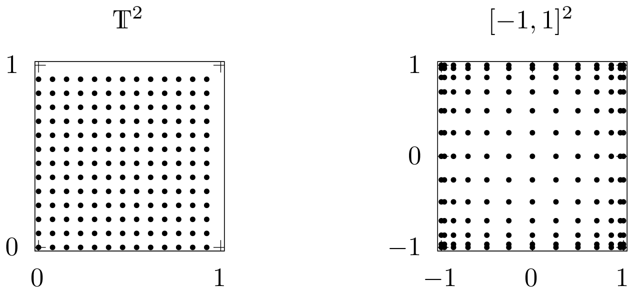


Figure 6.1: **Left:** Equispaced grid on $\mathbb{T}^2$ and, **right**, a Chebyshev grid on $[-1,1]^2$.

Proof. The assertion follows from Theorem 6.4 with

$$\beta(\boldsymbol{x}) = \Big(\frac{1}{2\varrho_T(D)} + \frac{1}{2}\frac{\sum_{k=1}^{m-1}|\eta_k(\boldsymbol{x})|^2}{m-1}\Big)^{-1} \le 2\varrho_T(D)\,. \qquad \blacksquare$$

A nice aspect of these random constructions is their universal applicability. Being able to compute the Christoffel function $N(V,\cdot) = \sum_{k=1}^{m-1}|\eta_k(\cdot)|^2$, we immediately obtain an L_2-MZ inequality with only logarithmic oversampling. On the downside the randomness is not controllable and the resulting points do not need to possess any structure.

6.2 Examples of deterministic L_2-MZ inequalities

In this section we give examples of points $\boldsymbol{X} = \{\boldsymbol{x}^1,\dots,\boldsymbol{x}^n\}$ forming exact quadrature rules for certain m-dimensional function spaces V on the torus $\mathbb{T}^d$ and the cube $[-1,1]^d$. By exact quadrature we consider points form L_2-MZ inequalities with $A = B$, cf. Lemma 6.3. Because of their structure it is possible to use fast algorithms for the matrix-vector product with the corresponding system matrix (2.1) decreasing the computational complexity to $\mathcal{O}(n\log n)$ instead of the naive $\mathcal{O}(n\cdot m)$.

Grid points. Starting with an one-dimensional quadrature rule it is easy to deduce a multivariate equivalent by simply tensorizing it. The result then has a grid-like structure. Examples for equispaced points and Chebyshev points are depicted in Figure 6.1.

The equispaced points on $\mathbb{T}^d$ are exact for the full cube of frequencies.

Lemma 6.6. *Let $n \in \mathbb{N}$ be such that $\sqrt[d]{n} \in \mathbb{N}$. The points*

$$\boldsymbol{X} = \Big\{ \frac{1}{\sqrt[d]{n}} \boldsymbol{i} : \boldsymbol{i} \in \{1, \dots, \sqrt[d]{n}\}^d \Big\}$$

fulfill an L_2-MZ inequality with equal weights $\omega_{\boldsymbol{i}} = 1/n$ and $A = B = 1$ for $V = \operatorname{span}\{\exp(2\pi \mathrm{i} \langle \boldsymbol{k}, \cdot \rangle)\}_{\boldsymbol{k} \in I}$ with the full cube of frequencies

$$I = \Big\{ -\frac{\sqrt[d]{n}}{2}, \dots, \frac{\sqrt[d]{n}}{2} - 1 \Big\}^d .$$

Proof. Let $\boldsymbol{W}$ and $\boldsymbol{L}$ be as in (2.1). By Theorem 6.2 we have to show that $\boldsymbol{W}^{1/2}\boldsymbol{L}$ is orthogonal. Because of the tensor-product structure, it is sufficient to show this in the one-dimensional case. The one-dimensional result is true by applying the formula for the geometric sum, which can be found in [PPST18, Lemma 3.10]. ■

The matrix-vector product with the Fourier-matrix given the equispaced points and the cube of frequencies is computable using the Fast Fourier Transform (FFT) with $\mathcal{O}(n \log n)$ flops, which goes back to J. W. Cooley and J. Tukey, cf. [CT65].

In order to obtain the equivalent for the cube $[-1, 1]^d$ we apply the transform $\cos(\pi \cdot)$ to every point in a component-wise fashion.

Lemma 6.7. *Let $n \in \mathbb{N}$ such that $\sqrt[d]{n-1} \in \mathbb{N}$. The points*

$$\boldsymbol{X} = \Big\{ \cos\Big(\frac{\pi}{\sqrt[d]{n-1}} \boldsymbol{i} \Big) : \boldsymbol{i} \in \{0, \dots, \sqrt[d]{n-1}\}^d \Big\}$$

fulfill an L_2-MZ inequality with weights

$$\omega_{\boldsymbol{i}} = 1/n \prod_{j=1}^{d} \begin{cases} 1/2 & \text{if} \quad i_j \in \{0, \sqrt[d]{n-1}\} \\ 1 & \text{otherwise} \end{cases}$$

and $A = B = 1$ for the Chebyshev polynomials $V = \operatorname{span}\{T_{\boldsymbol{k}}\}_{\boldsymbol{k} \in I}$, defined in Section 3.5.3, with the frequency index set

$$I = \{0, \dots, \sqrt[d]{n-1}\}^d .$$

Proof. Analogously to Theorem 6.6 using [PPST18, Lemma 3.46]. ■

It is again possible to carry out the corresponding matrix-vector product using the discrete cosine transform of type I (DCT-I) in $\mathcal{O}(n \log n)$, cf. [PPST18, Section 6.3].

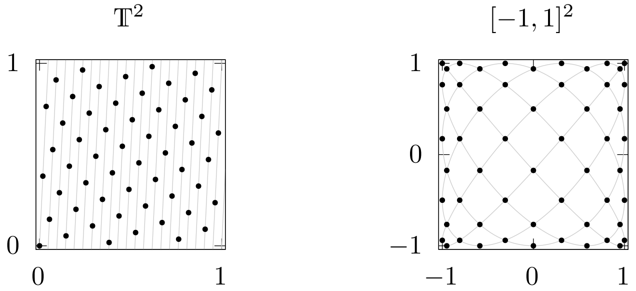


Figure 6.2: **Left:** A rank-1 lattice on $\mathbb{T}^2$ and, **right**, the transformed counterpart, a Chebyshev lattice, on $[-1,1]^2$.

L_2-MZ inequalities with this grid structure are optimal in the sense that the number of points n is equal to the dimension of the function space V. On the downside, this number grows exponentially in the dimension d where the expressivity of the full frequency grid is rarely needed for approximation. I.e., they suffer the curse of dimensionality.

Rank-1 lattices. Next, we present L_2-MZ inequalities, which work for arbitrary frequency index sets including hyperbolic crosses making them suitable for approximating functions in high-dimensions belonging to Sobolev spaces with dominating mixed smoothness, cf. Section 3.6.2. For a detailed study of rank-1 lattices see [SJ94, Kä14a, KPV15, PPST18, DKP22].

Rank-1 lattices $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq \mathbb{T}^d$ consist of equispaced points on a line which wraps around the d-dimensional torus $\mathbb{T}^d$. They are defined by a generating vector $\boldsymbol{z} \in \mathbb{N}^d$ and a lattice size $n \in \mathbb{N}$ via

$$\boldsymbol{X} := \left\{ \frac{1}{n}(i\boldsymbol{z} \bmod n\mathbb{1}) \in \mathbb{T}^d : i = 0, \dots, n-1 \right\}.$$

An example of a rank-1 lattice in dimension $d = 2$ is depicted in the left of Figure 6.2.

Because of their one-dimensional structure, a one-dimensional FFT can be used to compute the matrix-vector product with the corresponding Fourier matrix $\boldsymbol{L}$ in $\mathcal{O}(n \log n + d|I|)$ instead of the naive $\mathcal{O}(n \cdot |I|)$, where $I \subseteq \mathbb{Z}^d$ is an arbitrary frequency index set, cf. [PPST18, Algorithms 8.8 and 8.9]. For approximating functions with rank-1 lattices we suppose the following feature: We say a rank-1 lattice $\boldsymbol{X}$ has the **reconstructing property** for a frequency index set I, if

$$\frac{1}{n}\sum_{i=1}^{n}\exp(2\pi\mathrm{i}\langle\boldsymbol{k},\boldsymbol{x}^i\rangle)=\delta_{\boldsymbol{0},\boldsymbol{k}}\quad\text{for all}\quad\boldsymbol{k}\in\mathcal{D}(I)\,,\tag{6.4}$$

where $\mathcal{D}(I)=\{\boldsymbol{k}-\boldsymbol{l}:\boldsymbol{k},\boldsymbol{l}\in I\}$ is the frequency difference set. By Theorem 6.2 this is equivalent to an L_2-MZ inequality for the space $V=\operatorname{span}\{\exp(2\pi\mathrm{i}\langle\boldsymbol{k},\cdot\rangle)\}_{\boldsymbol{k}\in I}$ with $A=B=1$. There exist algorithms which, given a frequency index set I and n, compute a generating vector $\boldsymbol{z}$ such that the rank-1 lattice $\boldsymbol{X}$, as defined above, with equal weights ω_i fulfills an L_2-MZ inequality (6.1) for $A=B$. This idea was thoroughly developed in [Nuy07, CN07] and is still an active subject, cf. [Kä14a, Kä20, KMNN21]. One algorithm we will use later is a probabilistic approach recently presented in [Kä20]. On the one hand the component-by-component (CBC) rank-1 lattice construction is extremely efficient with computational costs that are linear in $|I|$ up to a few logarithmic factors. On the other hand, the sizes n of the resulting rank-1 lattices might be slightly (but only up to a factor of two with high probability) larger than those resulting from the deterministic approaches.

Note, that the reconstructing requirement (6.4) are $|\mathcal{D}(I)|\approx|I|^2$ conditions in the worst case possibly blowing up the size n of the rank-1 lattice. More detailed results on the minimal size of n are in [PV16, Theorem 2.1], which is a direct consequence of [Kä14b].

Analogously to the grid points, we may transform the rank-1 lattices via $\cos(\pi\cdot)$ to obtain points on $[-1,1]^d$. These are know as Chebyshev lattices. A prominent example are Padua points lying on Lissajois curves in dimension $d=2$. A thorough introduction is in [CP11, PC12, PV15] where we present some results from the latter.

For a generating vector $\boldsymbol{z}\in\mathbb{R}^d$ and a lattice size $n\in\mathbb{N}$, Chebyshev lattices are of the form

$$\boldsymbol{X}=\left\{\cos\left(\frac{i\pi}{n}\boldsymbol{z}\right):i=0,\dots,n\right\}.$$

An example for the special case of Padua points for $d=2$, $\boldsymbol{z}=(9,10)$, and $n=90$ is depicted in the right of Figure 6.2, where multiple points coincide. To obtain the equivalent to the reconstructing property (6.4) of rank-1 lattices one works with the mirrored index set

$$\mathcal{M}(I):=\left\{\boldsymbol{h}\in\mathbb{Z}^d:(|h_1|,\dots,|h_d|)^\mathsf{T}\in I\right\}$$

and the component-by-component algorithms which are used for rank-1 lattices. A result on the minimal size of a Chebyshev lattice is stated in [PV15, Remark IV.4] stating the existence of a reconstructing lattice for the function space $V = \operatorname{span}\{T_{\boldsymbol{k}}\}_{\boldsymbol{k}\in I}$ of size n with

$$n \le \max\left\{\frac{2}{3}\Big(|\mathcal{M}(I)|^2 - |\mathcal{M}(I)| + 8\Big),\ \max_{\boldsymbol{k}inI} 3\|\boldsymbol{k}\|_\infty\right\}.$$

The mirrored index set $\mathcal{M}(I)$ may be 2^d times bigger than I itself, increasing the lattice size and the slowing fast algorithms using the DCT-I down to $\mathcal{O}(n\log n + d2^d|I|)$ flops, cf. [PV15, Section III.A].

Multiple rank-1 lattices The drawbacks of the big lattice size of single rank-1 lattices can be overcome using the union of several rank-1 lattices known as multiple rank-1 lattices. Their analysis turns out to be tedious and resources can be found in [Kä18, GIKV21]. Again, it is possible to use Fourier algorithms to accelerate the matrix-vector product with the corresponding system matrix.

6.3 Subsampling of L_2-MZ inequalities

In Section 6.2 we have seen structured points suitable for the fast implementation of the matrix-vector product of the system matrix. But when it comes to high-dimensional approximation, the grid points are not suitable and rank-1 lattices, although they work for arbitrary frequency index sets, have a bad sampling complexity, cf. Theorem 7.11, compared to e.g. the random points in Section 6.1. In this section we will utilize subsampling techniques from Chapter 5 to find a subset of points from an L_2-MZ inequality which keeps its reconstructing properties and structure. We will go along [BKPU23, Section 3] and use random and `BSS` subsampling as illustrated in Figure 6.3.

Starting with an L_2-MZ inequality of points $\boldsymbol{X}_{\text{MZ}}$ and weights $\boldsymbol{W}_{\text{MZ}}$ exact for $\operatorname{span}\{\eta_k\}_{k\in I_{\text{MZ}}}$, the following theorem covers how to select a good L_2-MZ inequality with logarithmic oversampling.

Theorem 6.8. *Let $0 < C \le 1$, $\{\eta_k\}_{k\in I_{MZ}}$ be an L_2-orthonormal basis of the finite-dimensional function space V, let $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^M\} \subseteq D$ be points and $\boldsymbol{W}_{MZ} = \operatorname{diag}(\omega_1, \dots, \omega_M) \in [0,\infty)^{M\times M}$ weights fulfilling an L_2-MZ inequality for V with constants A and B. Further, let $t > 0$, $I \subseteq I_{MZ}$, and $n \in \mathbb{N}$ be such that*

$$n \ge \frac{12B}{AC}|I|(\log|I| + t)\,. \tag{6.5}$$

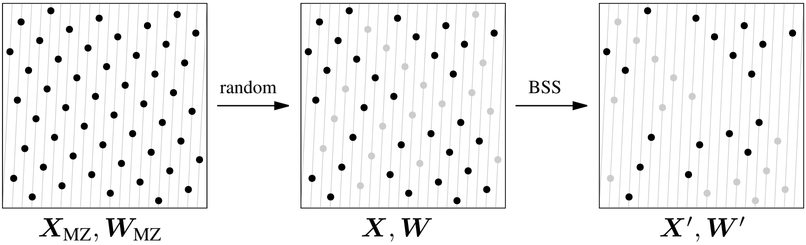

random
BSS
$\boldsymbol{X}_{\mathrm{MZ}}, \boldsymbol{W}_{\mathrm{MZ}}$
$\boldsymbol{X}, \boldsymbol{W}$
$\boldsymbol{X}', \boldsymbol{W}'$

Figure 6.3: Subsampling scheme.

We draw a set $\boldsymbol{X} = \{\boldsymbol{x}^i\}_{i\in J}$, $|J| = n$, of points i.i.d. from $\boldsymbol{X}_{MZ}$ with respect to ϱ_i (with duplicates), which are discrete probability weights ($\sum_{i=1}^n \varrho_i = 1$) fulfilling

$$\varrho_i \geq \frac{C\omega_i \sum_{k\in I} |\eta_k(\boldsymbol{x}^i)|^2}{\sum_{j=1}^M \omega_j \sum_{k\in I} |\eta_k(\boldsymbol{x}^j)|^2} \quad \textit{for} \quad i = 1, \dots, M\,. \tag{6.6}$$

Then, with probability larger than $1 - 2\exp(-t)$, there holds the subsampled L_2-MZ inequality

$$\frac{1}{2}A\|f\|_{L_2}^2 \leq \sum_{i\in J} \frac{\omega_i}{n\varrho_i}|f(\boldsymbol{x}^i)|^2 \leq \frac{3}{2}B\|f\|_{L_2}^2 \quad \textit{for all} \quad f \in \operatorname{span}\{\eta_k\}_{k\in I}\,. \tag{6.7}$$

Proof. Fixing the parameters $t := 1/2$ and $p := 2\exp(-t)$ in Theorem 5.1 yields the assertion (this is not the same t as in the present assertion). ■

Thus, we found a subset $\boldsymbol{X}$ of $\boldsymbol{X}_{\mathrm{MZ}}$ with logarithmic oversampling (independent of the original size M) which fulfills an L_2-MZ inequality with similar constants. The probability density stated in (6.6) is also used in Christoffel-weighted least squares approximation, cf. [NJZ17]. Note, that the weights ϱ_i are governed by the Christoffel function $\omega_i \sum_{k\in I} |\eta_k(\boldsymbol{x}^i)|^2$. With an upper estimate on it, the weights ϱ_i may be choosen equal.

Remark 6.9. *It is possible to apply the orthogonalization trick from [BSU23, Lemma 4.3] in order to eliminate the factor B/A in the assumption on the number of points n in* (6.5). *But this demands for setting up the matrix $\mathbf{L}$ from* (2.1) *whereas in the above formulation, we only need the evaluation of the Christoffel function.*

Next, we subsample $\boldsymbol{X}$ further to obtain merely linear oversampling.

Theorem 6.10. *Let the assumptions from Theorem 6.8 hold and $\boldsymbol{X}$ be such that* (6.7) *holds. Further, let $b > \kappa^2$ with*

$$\kappa = \frac{3B}{2A} + \frac{1}{2} + \sqrt{\Big(\frac{3B}{2A} + \frac{1}{2}\Big)^2 - 1}\,.$$

Then `BSS`*-subsampling (Algorithm 1) the points $\boldsymbol{X}$, we obtain points $\boldsymbol{X}' = \{\boldsymbol{x}^i\}_{i\in J'} \subseteq \boldsymbol{X}$ with $|\boldsymbol{X}'| \le \lceil b|I|\rceil$ and non-negative weights s_i, $i \in J'$ such that it holds the subsampled L_2-MZ inequality*

$$\frac{1}{2}A\|f\|_{L_2}^2 \le \sum_{i\in J'} \frac{\omega_i s_i}{n\varrho_i}|f(\boldsymbol{x}^i)|^2 \le \frac{3}{2}\frac{(\sqrt{b}+1)^2}{(\sqrt{b}-1)(\sqrt{b}-\kappa)}B\|f\|_{L_2}^2$$

for all $f \in \operatorname{span}\{\eta_k\}_{k\in I}$.

Proof. The result is an immediate consequence of applying Theorem 5.3 to the randomly subsampled points of Theorem 6.8. ■

The previous result is based on [BSS09] where tight frames were subsampled, which was used [LT22] for subsampling random points. This was then extended in [BSU23] for non-tight frames as well, which we use here. The `BSS`-algorithm gives no control over the weights s_i, however. One alternative would be to use Weaver-subsampling to loose the weights, but this is highly nonconstructive and spoils the involved constants, cf. [NSU21]. A clever extension of the `BSS`-algorithm makes it possible to loose the weights from the `BSS`-subsampling, regain the constructiveness, choose a smaller oversampling factor b, and save the left-hand side of the L_2-MZ inequality, which is the important one as it allows for the reconstruction from the function evaluations $f(\boldsymbol{x}^i)$.

Theorem 6.11. *Let the assumptions from Theorem 6.8 hold and $\boldsymbol{X}$ be such that* (6.7) *holds. Further, let $I \subseteq I_{MZ}$ and $b > 1 + \frac{1}{|I|}$. Then* `PlainBSS`*-subsampling (cf. Algorithm 3) the points $\boldsymbol{X}$, we obtain $\boldsymbol{X}' = (\boldsymbol{x}^i)_{i\in J'} \subseteq \boldsymbol{X}$ with $|\boldsymbol{X}'| \le \lceil b|I|\rceil$ such that it holds the subsampled left L_2-MZ inequality*

$$\frac{(b-1)^3}{178\,(b+1)^2}A\|f\|_{L_2}^2 \le \frac{1}{|I|}\sum_{i\in J'}\frac{\omega_i}{\varrho_i}|f(\boldsymbol{x}^i)|^2 \quad \textit{for all} \quad f \in V\,.$$

Proof. The result is an immediate consequence of applying Theorem 5.11 to the randomly subsampled points of Theorem 6.8. ■

Consequently, we have constructed subsets of existing L_2-MZ inequalities. This allows to use the algorithms for the initial points as we will see in the numerical experiments of Section 7.3.

Similar to Theorem 6.5 we now state the existence of an equalweighted lower L_2-MZ inequality with linear oversampling. This result uses an unstructured random draw of points for the initial L_2-MZ inequality.

Corollary 6.12. *Let V be an $m-1$-dimensional function space with an orthonormal basis $\eta_1, \dots, \eta_{m-1}$ in L_2. Further, let $\boldsymbol{X}$ denote the point set from Theorem 6.5 fulfilling* (6.3). *For $b > 1 + \frac{1}{m-1}$ we can construct an index set $J' \subseteq \{1, \dots, n\}$ with $|J'| \le \lceil b(m-1) \rceil$ for which the points $\{\boldsymbol{x}^i\}_{i \in J'}$ fulfill the unweighted lower L_2-MZ inequality with probability exceeding* $1 - \exp(-t)$, *i.e.,*

$$\|f\|^2_{L_2(D,\varrho_T)} \le 356 \frac{(b+1)^2}{(b-1)^3} \frac{\varrho_T(D)}{m-1} \sum_{i \in J'} |f(\boldsymbol{x}^i)|^2 \quad \textit{for all} \quad f \in V\,.$$

Proof. The result is an immediate consequence of applying Theorem 5.11 to the randomly subsampled points of Theorem 6.5. ■

This result allows to discretize an arbitrary $(m-1)$-dimensional function space using only point evaluations with merely linear oversampling.

Chapter 7

Least squares in the worst-case setting

The worst-case setting, also known as active learning setting, is a subfield of machine learning where we are not given the data in advance but rather the learning algorithm queries labels for specific data points. In sampling recovery this means we choose points $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq D$ in the domain $D \subseteq \mathbb{R}^d$ where we evaluate the target function $f \colon \mathbb{D} \to \mathbb{K}$ with $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. It is necessary to know a restriction in advance which will be the assumption $f \in F$ for some function class F. We are interested in worst-case error guarantees for a given class of functions F, i.e., the chosen points $\boldsymbol{X}$ will work for every function $f \in F$. The so called **(linear) sampling width** in L_2 measure the best achievable performance with $n-1$ points

$$g_n(F, L_2(D, \varrho_T)) := \inf_{\substack{\boldsymbol{x}^1,\dots,\boldsymbol{x}^{n-1} \in D \\ \varphi_1,\dots,\varphi_{n-1} \in L_2}} \sup_{\|f\|_F \le 1} \Big\| f - \sum_{i=1}^{n-1} f(\boldsymbol{x}^i)\varphi_i \Big\|_{L_2}, \quad (7.1)$$

cf. [DuTU19, (5.0.1)]. We obtain upper bounds for the sampling width by proving bounds for the **sampling width restricted to the least squares approximation**

$$g_{n,m}^{\mathrm{ls}}(F, L_2(D, \varrho_T)) := \inf_{\substack{V \subseteq \ell_\infty(D) \\ \dim V = m-1}} \inf_{\boldsymbol{X} \in D^{n-1}} \sup_{\|f\|_F \le 1} \|f - S_V^{\boldsymbol{X}} f\|_{L_2}, \quad (7.2)$$

where $S_V^{\boldsymbol{X}}$ is the least squares approximation from Chapter 2. To quantify the goodness of our bounds we further introduce the **linear width**

$$a_m(F, L_2(D, \varrho_T)) := \inf_{\substack{\ell_1,\dots,\ell_{m-1} \colon F \to \mathbb{C} \\ \varphi_1,\dots,\varphi_{m-1} \in L_2}} \sup_{\|f\|_F \le 1} \Big\| f - \sum_{k=1}^{m-1} \ell_k(f)\varphi_k \Big\|_{L_2}. \quad (7.3)$$

Since point evaluations are a special case of linear functionals, we have

$$a_m(F, L_2(D, \varrho_T)) \le g_m(F, L_2(D, \varrho_T)) \le g_{m,m}^{\mathrm{ls}}(F, L_2(D, \varrho_T))$$

and the linear width serves as a natural lower bound for the sampling width. Note, that the linear width is equal to the approximation numbers of the identity operator. If $F = H(K)$ is an RKHS and the embedding $I_K \colon H(K) \to$

L_2 has singular values σ_k and we obtain $a_m(H(K), L_2(D, \varrho_T)) = \sigma_m$ by Theorem 3.12.

As we will see, least squares approximation already achieves a sharp bound for the sampling width.

7.1 Sampling recovery in spaces of finite measure

In this section we consider the worst-case function approximation in the space $L_2(D, \varrho_T)$ for a finite measure ϱ_T, which is based on [BSU23]. It turns out the existence of a non-weighted lower L_2-MZ inequality is sufficient to prove error bounds:

Lemma 7.1. *Let ϱ_T be a finite measure and $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$ fulfill a lower L_2-MZ inequality for the function space V with constant A and all weights $\omega_i = 1$.*

Then we have for $S_V^{\boldsymbol{X}} \boldsymbol{f}$ the plain least squares approximation defined in Section 2.2

$$\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \le 4\Big(\varrho_T(D) + \frac{n}{A}\Big) \inf_{g \in V} \|f - g\|_\infty^2 \,.$$

Proof. For any $h \in V$ we have by triangle inequality

$$\|f - S_V^{\boldsymbol{X}}\|_{L_2}^2 \le 2\|f - h\|_{L_2}^2 + 2\|h - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \,.$$

Since we have finite measure, the first summand is bounded by $\|f - h\|_{L_2}^2 \le \varrho_T(D)\|f - h\|_\infty^2$.

For the second summand we use the invariance of $S_V^{\boldsymbol{X}}$ to functions in V, Lemmata 6.1, and 6.2

$$\begin{aligned}\|h - S_V^{\boldsymbol{X}} f\|_{L_2}^2 &= \|S_V^{\boldsymbol{X}}(h - f)\|_{L_2}^2 \\ &\le \|(\boldsymbol{L}^*\boldsymbol{L})^{-1}\boldsymbol{L}^*\|_{2\to 2}^2 \sum_{i=1}^n |(h - f)(\boldsymbol{x}^i)|^2 \\ &\le \frac{n}{A}\|h - f\|_\infty^2 \,.\end{aligned}$$

Overall, we obtain

$$\|f - S_V^{\boldsymbol{X}}\|_{L_2}^2 \le 2\Big(\varrho_T(D) + \frac{n}{A}\Big)\|f - h\|_\infty^2 \,.$$

It is left to choose $h \in V$ such that $\|f - h\|_\infty^2 \le 2\inf_{g\in V} \|f - g\|_\infty^2$. ■

Next, we use the constructions of points for L_2-MZ inequalities from Chapter 6. We start with the random points from Theorem 6.5.

Theorem 7.2. *Let V be an $m-1$-dimensional function space with an orthonormal basis $\eta_1, \dots, \eta_{m-1}$ in $L_2(D, \varrho_T)$. Let $t \geq 0$ and $n \in \mathbb{N}$ such that*

$$20(m-1)(\log(m-1)+t) \leq n$$

and let the points $\boldsymbol{x}^1, \dots, \boldsymbol{x}^n$ be drawn according to the following probability measure

$$\mathrm{d}\varrho_S(\boldsymbol{x}) = \frac{1}{2\varrho_T(D)} + \frac{1}{2}\frac{\sum_{k=1}^{m-1}|\eta_k(\boldsymbol{x})|^2}{m-1}\,\mathrm{d}\varrho_T\,.$$

Then we have for $S_V^{\boldsymbol{X}}\boldsymbol{f}$ the plain least squares approximation defined in Section 2.2 with probability exceeding $1-\exp(-t)$

$$\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \leq 20\varrho_T(D)\inf_{g\in V}\|f-g\|_\infty^2\,.$$

Proof. The assertion follows from Theorem 7.1 and Theorem 6.5. ■

Inequalities of this type with logarithmic oversampling have been first established by A. Cohen and G. Migliorati [CM17]. This has been improved by V. N. Temlyakov [Tem21] to $n = \mathcal{O}(m)$ samples with unspecified constants. The mentioned results rely on the weighted least squares approximation and V. N. Temlyakov posed the question in [Tem21], if also classical plain least squares approximation could be used. The next theorem gives an affirmative answer and even displays the dependence of the constant on the oversampling factor b.

Theorem 7.3. *Let V be an $m-1$-dimensional function space with an orthonormal basis $\eta_1, \dots, \eta_{m-1}$ in $L_2(D, \varrho_T)$. Further, let $b > 1 + \frac{1}{m-1}$ and $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$ with $|\boldsymbol{X}| \leq \lceil b(m-1)\rceil$ be the points from Theorem 6.12.*

Then we have for $S_V^{\boldsymbol{X}}\boldsymbol{f}$ the plain least squares approximation defined in Section 2.2

$$\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \leq 714\,\varrho_T(D)\Big(\frac{b+1}{b-1}\Big)^3\inf_{g\in V}\|f-g\|_\infty^2\,.$$

Proof. The assertion follows from Theorem 7.1 where we use the L_2-MZ inequality from Theorem 6.12 and the estimate

$$\frac{\lceil b(m-1)\rceil}{m-1} \leq 2\frac{b(m-1)}{m-1} \leq 2(b+1)\,.$$ ■

This result yields a relation of the sampling width restricted to the least squares approximation and the **Kolmogorov width** in $\ell_\infty(D)$

$$d_m(F, \ell_\infty(D)) := \inf_{\substack{V \subseteq F \\ \dim(V)=m-1}} \sup_{\|f\|_F \le 1} \inf_{g \in V} \Big\| f - g \Big\|_\infty ,$$

which includes non-linear approximations in contrast to the linear width. Note, that these coincide when the error is measured in a Hilbert spaces norm. When optimizing over all $m-1$-dimensional reconstruction spaces, we obtain the following estimate.

Corollary 7.4. *Let F be a class of functions in $\ell_\infty(D)$. Then for $m \in \mathbb{N}$ and $b > 1 + \frac{1}{m}$ there exists a constant C_{ϱ_T} such that*

$$g^{\mathrm{ls}}_{\lceil bm \rceil, m}(F, L_2(D, \varrho_T)) \le C_{\varrho_T} \Big(\frac{b+1}{b-1} \Big)^{3/2} d_m(F, \ell_\infty(D)) . \tag{7.4}$$

Since the quantities on the left-hand side of (7.4) are in general larger than the standard sampling width (7.1), where there are no such restrictions on the recovery, this slightly improves on recent results by V. N. Temlyakov, Theorems 1.1 and 1.2 in [Tem21] as well as [LT22, Thm. 3.4], the latter joint work with I. Limonova. Interestingly, the b-dependent constant may be improved to $(b-1)^{-1}$ when allowing weighted least squares approximations, cf. [LT22, Thm. 1.7] (or the original [DPS$^+$21, Thm. 6.3]) for the case of real functions (an extension to the complex case has been given in [LT22, Rem. 3.2] but only for $b > 2$). In our case a distinction between real and complex $L_2(D, \varrho_T)$ in (7.4) as in [LT22] is unnecessary due to the validity of Theorem 5.3 in the complex setting. Note that the right-hand side in (7.4) is of particular importance if the linear widths in L_2 are not square-summable [TU21, TU22].

7.2 Sampling recovery in RKHSs

7.2.1 Optimal sampling complexity with unstructured points

In this section, we investigate the same problem - the recovery of functions in $L_2(D, \varrho_T)$ - but under the assumption of functions being element of an RKHS $H(K)$, where ϱ_T is allowed to be infinite. This scenario is investigated in the recent papers [DKU23, KU21a, KUV21, MU21, NSU21], where this section is based on [BSU23].

We start with some basic assumptions

Assumption 7.5. *Let $H(K)$ be a separable RKHS with finite trace* (3.4) *and let σ_k^2 and η_k denote the eigenvalues and $L_2(D, \varrho_T)$-orthonormal eigenfunctions of the integral operator T_K from* (3.2). *Further, we assume $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq D$ to be i.i.d. random points with respect to the measure* $\mathrm{d}\varrho_S(\boldsymbol{x}) = 1/\beta(\boldsymbol{x})\, \mathrm{d}\varrho_T$, *where*

$$\frac{1}{\beta(\boldsymbol{x})} = \frac{1}{2}\frac{\sum_{k=1}^{m-1}|\eta_k(\boldsymbol{x})|^2}{m-1} + \frac{1}{2}\frac{\sum_{k=m}^{\infty}|e_k(\boldsymbol{x})|^2}{\sum_{k=m}^{\infty}\sigma_k^2}$$

for $m, n \in \mathbb{N}$, which will be specified later.

The use of this density is based on an idea first used in [KU21a]. Before proving the actual bounds we need an auxiliary lemma, which will be used to bound the discrete truncation error where the coarse $\ell_\infty(D)$-estimate was used in Section 7.1.

Lemma 7.6. *Let Theorem 7.5 hold, $n \geq 3$, and let*

$$\boldsymbol{\Phi} = [e_k(\boldsymbol{x}^i)]_{i=1,\dots,n;k=m,m+1,\dots} \in \mathbb{C}^{n\times\infty}$$

and $$\boldsymbol{W} = \operatorname{diag}(\beta(\boldsymbol{x}^1), \dots, \beta(\boldsymbol{x}^n)) \in [0,\infty)^{n\times n}\,.$$

Then we have with probability exceeding $1 - 2^{3/4}\exp(-t)$

$$\frac{1}{n}\|\boldsymbol{W}^{1/2}\boldsymbol{\Phi}\|_{2\to2}^2 \leq \frac{42(\log n + t)}{n}\sum_{k=m}^{\infty}\sigma_k^2 + 2\sup_{k\geq m}\sigma_k^2\,.$$

Proof. We apply Theorem 4.9 for $\boldsymbol{u}^i = \sqrt{\beta(\boldsymbol{x}^i)}(e_k(\boldsymbol{x}^i))_{k=m}^\infty$ using

$$\begin{aligned}\|\boldsymbol{u}^i\|_2^2 &= \frac{2}{\frac{\sum_{k=1}^{m-1}|\eta_k(\boldsymbol{x}^i)|^2}{m-1} + \frac{\sum_{k=m}^{\infty}|e_k(\boldsymbol{x}^i)|^2}{\sum_{k=m}^{\infty}\sigma_k^2}}\sum_{k=m}^{\infty}|e_k(\boldsymbol{x}^i)|^2 \\ &\leq 2\sum_{k=m}^{\infty}\sigma_k^2\end{aligned}$$

and plug in the identity

$$\mathbb{E}(\boldsymbol{u}^i \otimes \boldsymbol{u}^i) = \operatorname{diag}(\sigma_m^2, \sigma_{m+1}^2, \dots)\,. \qquad ■$$

With that, we state a bound on the worst-case error of the least squares approximation using random points with logarithmic oversampling.

Theorem 7.7. *Let Theorem 7.5 hold with $m \geq 3$ and*

$$n \coloneqq \lceil 20(m-1)(\log(m-1)+t)\rceil \leq (m-1)^r\,,$$

for some $t > 0$ and $r > 1$.

Then, for $S_V^{\boldsymbol{X}} \boldsymbol{f}$ the weighted least squares approximation defined in Section 2.2 with $\omega_i = \beta(\boldsymbol{x}^i)$, we have with probability exceeding $1 - 4\exp(-t)$

$$\sup_{\|f\|_{H(K)}\leq 1} \|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \leq 3 \sup_{k\geq m} \sigma_k^2 + \frac{5r}{m-1} \sum_{k=m}^{\infty} \sigma_k^2\,.$$

Proof. We begin by defining two events. The first one bounds the spectral norm of the least squares matrix

$$S_1 \coloneqq \Big\{ \boldsymbol{X} \subseteq D^n : \Big\| (\boldsymbol{L}^* \boldsymbol{W} \boldsymbol{L})^{-1} \boldsymbol{L} \boldsymbol{W}^{1/2} \Big\|_{2\to 2}^2 \leq \frac{2}{n} \Big\}\,.$$

Using Theorem 6.2, we have that this is equivalent to the lower bound of the singular values of $\boldsymbol{W}^{1/2}\boldsymbol{L}$. By the definition of β, we have

$$10\Big(\|\beta(\cdot)N(V,\cdot)\|_\infty\Big)(\log(m-1)+t) \leq 20(m-1)(\log(m-1)+t) \leq n$$

and apply Theorem 6.4 to obtain that event S_1 holds with probability at least $1 - \exp(-t)$.

Event S_2 is the inequality of Theorem 7.6 which holds with probability at least $1 - 2^{3/4}\exp(-t)$. Together we obtain the desired probability

$$\mathbb{P}(S_1 \cap S_2) \geq 1 - \mathbb{P}S_1^{\complement} - \mathbb{P}S_2^{\complement} \geq 1 - 4\exp(-t)\,.$$

It remains to show the bound based on events S_1 and S_2. On that account we decompose the error using orthogonality

$$\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 = \|f - P_V f\|_{L_2}^2 + \|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \tag{7.5}$$

and estimate each summand individually.

Step 1. We estimate the first summand of (7.5) by

$$\begin{aligned}\|f - P_V f\|_{L_2}^2 &= \Big\| \sum_{k\geq m} \langle f, e_k\rangle_{H(K)} e_k \Big\|_{L_2}^2 = \sum_{k\geq m} \sigma_k^2 |\langle f, e_k\rangle_{H(K)}|^2 \\ &\leq \|f\|_{H(K)}^2 \sup_{k\geq m} \sigma_k^2\,.\end{aligned}$$

Step 2. For the second summand we use the invariance of $S_V^{\boldsymbol{X}}$ to functions in V

$$\begin{aligned}\|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 &= \|S_V^{\boldsymbol{X}}(f - P_V f)\|_{L_2}^2\\ &= \|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}((f - P_V f)(\boldsymbol{x}^i))_{i=1}^n\|_{L_2}^2\\ &\le \|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}^{1/2}\|_{2\to 2}^2 \sum_{i=1}^n \beta(\boldsymbol{x}^i)|(f - P_V f)(\boldsymbol{x}^i)|^2 .\end{aligned}$$

We estimate the first factor by the means of event S_1 and write the latter in terms of its coefficients in order to apply event S_2

$$\begin{aligned}\|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 &\le \frac{2}{n}\Big\|\boldsymbol{W}^{1/2}\boldsymbol{\Phi}\Big(\langle f, e_k\rangle_{H(K)}\Big)_{k\ge m}\Big\|_2^2\\ &\le \frac{2}{n}\|\boldsymbol{W}^{1/2}\boldsymbol{\Phi}\|_{2\to 2}^2 \sum_{k\ge m} |\langle f, e_k\rangle_{H(K)}|^2\\ &\le \Big(\frac{84(\log n + t)}{n}\sum_{k=m}^\infty \sigma_k^2 + 2\sup_{k\ge m}\sigma_k^2\Big)\|f\|_{H(K)}^2 .\end{aligned}$$

Overall we obtain

$$\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \le \Big(3\sup_{k\ge m}\sigma_k^2 + \frac{84(\log n + t)}{n}\sum_{k=m}^\infty \sigma_k^2\Big)\|f\|_{H(K)}^2 .$$

By the assumption on m, we have

$$\frac{84(\log n + t)}{n} \le \frac{84(\log n + t)}{20(m-1)(\log(m-1) + t)} \le \frac{84}{20}\frac{r}{m-1} \le \frac{5r}{m-1}$$

and obtain the assertion. ∎

The logarithmic oversampling in Theorem 7.7 is improved to linear oversampling in the next theorem with the drawback of loosing a logarithmic term in the bound itself. However, this gives a bound with better overall sampling complexity.

Theorem 7.8. *Let Theorem 7.5 hold and let t, m, r, and $\boldsymbol{X}$ be as in Theorem 7.7. For $b > 1 + \frac{1}{m-1}$, we construct $J' \subseteq \{1, \dots, n\}$ with $|J'| \le \lceil b(m-1)\rceil$ by applying `PlainBSS` Algorithm 3 to $(\sqrt{\beta(\boldsymbol{x}^i)}\eta_k(\boldsymbol{x}^i))_{k=1}^{m-1}$.*

Then, for $S_V^{\boldsymbol{X}'}\boldsymbol{f}$ the weighted least squares approximation defined in Section 2.2 with $\boldsymbol{X}' = \{\boldsymbol{x}^i\}_{i\in J'}$ and $\omega_i = \beta(\boldsymbol{x}^i)$, we have with probability exceeding $1 - 4\exp(-t)$

$$\begin{aligned}&\sup_{\|f\|_{H(K)}\le 1} \|f - S_V^{\boldsymbol{X}'} f\|_{L_2}^2\\ &\quad\le 9049\frac{(b+1)^2}{(b-1)^3}(\log(m-1)+t)\Big(\sup_{k\ge m}\sigma_k^2 + \frac{r}{m-1}\sum_{k=m}^{\infty}\sigma_k^2\Big)\,.\end{aligned}$$

Proof. We use the events S_1 and S_2 with the desired probability from Theorem 7.7 to show the result. By Theorem 5.11 and S_1we have for the subsampled least squares matrix $\boldsymbol{L}'$ and weight matrix $\boldsymbol{W}'$

$$\begin{aligned}&\|((\boldsymbol{L}')^*\boldsymbol{W}'\boldsymbol{L}')^{-1}(\boldsymbol{L}')^*(\boldsymbol{W}')^{1/2}\|_{2\to 2}^2 &(7.6)\\ &\le 89\frac{(b+1)^2}{(b-1)^3}\frac{n}{m-1}\|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}^{1/2}\|_{2\to 2}^2 &(7.7)\\ &\le 89\frac{(b+1)^2}{(b-1)^3}\frac{2}{m-1}\,. &(7.8)\end{aligned}$$

Analogous to Theorem 7.7, we decompose the error and estimate the first summand

$$\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 = \sup_{k\ge m}\sigma_k^2\|f\|_{H(K)}^2 + \|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2\,.$$

For the second summand we use the invariance of $S_V^{\boldsymbol{X}}$ to functions in V

$$\begin{aligned}\|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 &= \|S_V^{\boldsymbol{X}}(f - P_V f)\|_{L_2}^2\\ &= \Big\|((\boldsymbol{L}')^*\boldsymbol{W}'\boldsymbol{L}')^{-1}(\boldsymbol{L}')^*\boldsymbol{W}'((f-P_V f)(\boldsymbol{x}^i))_{i\in J'}\Big\|_{L_2}^2\\ &\le \Big\|((\boldsymbol{L}')^*\boldsymbol{W}'\boldsymbol{L}')^{-1}(\boldsymbol{L}')^*(\boldsymbol{W}')^{1/2}\Big\|_{2\to 2}^2\sum_{i\in J'}\beta(\boldsymbol{x}^i)|(f-P_V f)(\boldsymbol{x}^i)|^2\,.\end{aligned}$$

We estimate the first factor by the means of (7.6) and extend the latter to the points $\boldsymbol{X}\supset\boldsymbol{X}'$

$$\|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \le 89\frac{(b+1)^2}{(b-1)^3}\frac{n}{m-1}\frac{2}{n}\sum_{i=1}^{n}\beta(\boldsymbol{x}^i)|(f-P_V f)(\boldsymbol{x}^i)|^2\,.$$

Analogously to Step 2 of Theorem 7.7, we obtain

$$\begin{aligned}&\|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \\ &\quad\le 89 \frac{(b+1)^2}{(b-1)^3} \frac{n}{m-1} \Big(2 \sup_{k \ge m} \sigma_k^2 + \frac{5r}{m-1} \sum_{k=m}^{\infty} \sigma_k^2 \Big) \|f\|_{H(K)}^2 .\end{aligned}$$

Overall, this gives

$$\begin{aligned}&\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \\ &\quad\le 445 \frac{(b+1)^2}{(b-1)^3} \frac{n}{m-1} \Big(\sup_{k \ge m} \sigma_k^2 + \frac{r}{m-1} \sum_{k=m}^{\infty} \sigma_k^2 \Big) \|f\|_{H(K)}^2 .\end{aligned}$$

Using

$$\frac{n}{m-1} = \frac{\lceil 20(m-1)(\log(m-1)+t) \rceil}{m-1} \le \frac{61}{3} (\log(m-1)+t) ,$$

we obtain the assertion. ■

The performance of Theorem 7.8 is near-optimal as in [NSU21], i.e., we obtain for the sampling width (7.1)

$$g_n^2 \le \frac{C \log n}{n} \sum_{k \ge \lfloor cn \rfloor} \sigma_k^2 .$$

The latter reference is the first which used the Weaver subsampling technique based on the Kadison-Singer Theorem for the sampling recovery problem. However, it does not achieve the optimal rate. By a further refinement of the technique, established recently in [DKU23] by M. Dolbeault, D. Krieg, and M. Ullrich, the optimal rate (without additional log-term) has been found. In contrast to [DKU23, NSU21] we have a semi-constructive method to generate the sampling points (offline step) that does not need the Kadison-Singer theorem in terms of the Weaver subsampling. In addition, the dependence on the oversampling factor b is displayed. An **open question** remains: Although in many relevant cases (like periodic Sobolev spaces with mixed smoothness) the recovery operator turns out to be a canonical **plain** least squares approximation operator $S_V^{\boldsymbol{X}}$ with equal weights (acting on the hyperbolic cross frequency subspace with points displayed in Figure 5.2) we do not know whether loosing the weights is possible in general.

7.2.2 Sampling recovery with L_2-MZ points

The points in Section 7.2.1 stem from an initial random draw. Therefore, there is missing structure which may be useful for applying fast algorithms. In Section 6.2 we considered deterministic sets of points, which fulfill an L_2-MZ inequality, for which the corresponding least squares approximation is fast in the sense that Fourier-like algorithms can be used for the matrix-vector product with the corresponding least squares matrix (2.1). Based on [BKPU23], in this section we bound the worst-case error of the least squares approximation utilizing samples in L_2-MZ points. Error bounds for L_2-MZ points can also be found in [Grö20].

In the framework of this section, we consider RKHS $H(K)$ with bounded kernels, i.e.,

$$\|K\|_\infty := \sqrt{\sup_{\boldsymbol{x}\in D} K(\boldsymbol{x},\boldsymbol{x})} < \infty\,. \tag{7.9}$$

With $\|\cdot\|^2_{H(K)} = \langle\cdot,\cdot\rangle_{H_K}$, this condition implies $\|f\|_\infty \le \|K\|_\infty \cdot \|f\|_{H(K)}$, which, in other words, is the continuous embedding of $H(K)$ into the space of essentially bounded functions $\ell_\infty(D)$.

Further, we consider the following set of assumptions on the initial points $\boldsymbol{X}_{\mathrm{MZ}}$ and weights $\boldsymbol{W}_{\mathrm{MZ}}$ for the theorems of this section.

Assumption 7.9. *Let $H(K)$ be a separable RKHS with finite trace* (3.4)*, bounded* (7.9) *kernel, and let σ_k^2 and η_k denote the eigenvalues and the $L_2(D,\varrho_T)$-orthonormal eigenfunctions of the integral operator T_K* (3.2)*.*

For $V = \operatorname{span}\{\eta_k\}_{k\in I_{MZ}} \subseteq H(K)$ a finite-dimensional function space, we assume the points $\boldsymbol{X}_{MZ} \subseteq D$ and weights $\boldsymbol{W}_{MZ} \in [0,\infty)^{M\times M}$ fulfill an L_2-MZ inequality for V with constants A and B.

Under these assumptions, we state a rather general result.

Theorem 7.10. *Let Assumptions 7.9 hold. Then, for $S^{\boldsymbol{X}_{MZ}}_{V_{MZ}} f$ the weighted least squares approximation defined in Section 2.2 with weights $\boldsymbol{W}_{MZ}$, we have the exact reconstruction for function in V_{MZ} and otherwise*

$$\begin{aligned}
&\sup_{\|f\|_{H(K)}\le 1} \left\| f - S^{\boldsymbol{X}_{MZ}}_{V_{MZ}} f \right\|^2_{L_2} \\
&\le \sup_{k\notin I_{MZ}} \sigma_k^2 + \frac{\sum_{i=1}^M \omega_i}{A} \sup_{\|f\|_{H(K)}\le 1} \|f - P_{V_{MZ}} f\|^2_\infty\,.
\end{aligned}$$

Proof. The reconstructing property for functions in V_{MZ} is immediate. For general functions $f \in H(K)$ we use orthogonality of the projection to obtain

$$\|f - S_{V_{\mathrm{MZ}}}^{\boldsymbol{X}_{\mathrm{MZ}}} f\|_{L_2}^2 = \|f - P_{V_{\mathrm{MZ}}} f\|_{L_2}^2 + \|P_{V_{\mathrm{MZ}}} f - S_{V_{\mathrm{MZ}}}^{\boldsymbol{X}_{\mathrm{MZ}}} f\|_{L_2}^2 \tag{7.10}$$

of which we now estimate each summand individually.

Step 1. We bound the first summand of (7.10) by

$$\begin{aligned}\|f - P_{V_{\mathrm{MZ}}} f\|_{L_2}^2 &= \Big\| \sum_{k \notin I_{\mathrm{MZ}}} \langle f, e_k \rangle_{H(K)} e_k \Big\|_{L_2}^2 = \sum_{k \notin I_{\mathrm{MZ}}} \sigma_k^2 \Big| \langle f, e_k \rangle_{H(K)} \Big|^2 \\ &\le \|f\|_{H(K)}^2 \sup_{k \notin I_{\mathrm{MZ}}} \sigma_k^2 \,.\end{aligned}$$

Step 2. For the second summand of (7.10) we use the reconstructing property of $S_{V_{\mathrm{MZ}}}^{\boldsymbol{X}_{\mathrm{MZ}}}$ for functions in V_{MZ}

$$\begin{aligned}\|P_{V_{\mathrm{MZ}}} f - S_{V_{\mathrm{MZ}}}^{\boldsymbol{X}_{\mathrm{MZ}}} f\|_{L_2}^2 &= \|S_{V_{\mathrm{MZ}}}^{\boldsymbol{X}_{\mathrm{MZ}}} (P_{V_{\mathrm{MZ}}} f - f)\|_{L_2}^2 \\ &\le \Big\| ((\boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}})^* \boldsymbol{W}_{\mathrm{MZ}} \boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}})^{-1} (\boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}})^* \boldsymbol{W}_{\mathrm{MZ}}^{1/2} \Big\|_{2\to 2}^2 \cdot \\ &\quad \Big\| \boldsymbol{W}_{\mathrm{MZ}}^{1/2} \Big((P_{V_{\mathrm{MZ}}} f - f)(\boldsymbol{x}^i) \Big)_{i=1}^M \Big\|_2^2 \,.\end{aligned}$$

By Lemmata 6.1 and 6.2 we have for the first factor

$$\Big\| ((\boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}})^* \boldsymbol{W}_{\mathrm{MZ}} \boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}})^{-1} (\boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}})^* \boldsymbol{W}_{\mathrm{MZ}}^{1/2} \Big\|_{2\to 2}^2 \le \frac{1}{A} \,.$$

Finally, we obtain

$$\begin{aligned}&\sup_{\|f\|_{H(K)} \le 1} \|P_{V_{\mathrm{MZ}}} f - S_{V_{\mathrm{MZ}}}^{\boldsymbol{X}_{\mathrm{MZ}}} f\|_{L_2}^2 \\ &\le \frac{1}{A} \sup_{\|f\|_{H(K)} \le 1} \sum_{i=1}^M \omega_i |(P_{V_{\mathrm{MZ}}} f - f)(\boldsymbol{x}^i)|^2 \\ &\le \frac{\sum_{i=1}^M \omega_i}{A} \sup_{\|f\|_{H(K)} \le 1} \|P_{V_{\mathrm{MZ}}} f - f\|_\infty^2 \,. \qquad \blacksquare\end{aligned}$$

The bounds in Theorem 7.10 are rather basic and hold in a general setting. In special cases like the exponential functions, Chebyshev polynomials, and the half-period cosine functions this topic is examined in detail in [KMNN21] for rank-1 lattices, cf. Section 6.2. In particular for function spaces with mixed smoothness there exist almost tight error bounds in [BKUV17] which improve on Theorem 7.10 which we will present next. For simplicity we omit constants in the following. For two sequences $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty} \subseteq \mathbb{R}$ we write $a_n \lesssim b_n$ if there exists a constant $c > 0$ such that $a_n \leq c b_n$ for all n. We will write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

The following theorem is a collection of results from [KSU15, BDSU16, BKUV17].

Theorem 7.11 (Approximation with rank-1 lattices in H^s_{mix})**.** *Let $I_{MZ} \subseteq \mathbb{N}^d$ be a hyperbolic cross frequency index set with radius $R \in (1, \infty)$, see Section 3.6.2 and $\boldsymbol{X}_{MZ}$ a corresponding reconstructing rank-1 lattice with M points, cf.* (6.4)*.*

(i) The size of a hyperbolic cross is asymptotically $|I_{MZ}| \asymp R(\log R)^{d-1}$.

(ii) The best possible error of functions in H^s_{mix} behaves as follows

$$\sup_{\|f\|_{H^s_{\mathrm{mix}}} \leq 1} \|f - P_{V_{MZ}} f\|_{L_2}^2 \asymp |I_{MZ}|^{-2s} (\log |I_{MZ}|)^{2(d-1)s}$$

and

$$\sup_{\|f\|_{H^s_{\mathrm{mix}}} \leq 1} \|f - P_{V_{MZ}} f\|_{\infty}^2 \asymp |I_{MZ}|^{-2s+1/2} (\log |I_{MZ}|)^{2(d-1)s} \,.$$

(iii) The size of a reconstructing rank-1 lattice for I_{MZ} is bounded by

$$|I_{MZ}|^2 (\log |I_{MZ}|)^{-2(d-1)} \lesssim M \lesssim |I_{MZ}|^2 (\log |I_{MZ}|)^{-d} \,,$$

where the lower inequality holds for all reconstructing rank-1 lattices and there exists a rank-1 lattice fulfilling the upper one.

(iv) The error of the least squares approximation operator using rank-1 lattices and $V_{I_{MZ}} = \mathrm{span}\{\exp(2\pi \mathrm{i} \langle \boldsymbol{k}, \cdot \rangle)\}_{\boldsymbol{k} \in I_{MZ}}$ is bounded as follows:

$$\begin{aligned} M^{-s} &\lesssim \sup_{\|f\|_{H^s_{\mathrm{mix}}} \leq 1} \|f - S^{\boldsymbol{X}_{MZ}}_{V_{I_{MZ}}} f\|_{L_2}^2 \\ &\lesssim \begin{cases} |I_{MZ}|^{-2s} (\log |I_{MZ}|)^{(d-1)(2s+1)} \\ M^{-s} (\log M)^{(d-2)s+d-1} \,, \end{cases} \end{aligned}$$

where the lower inequality holds for all reconstructing rank-1 lattices and there exists a rank-1 lattice fulfilling the upper one.

Proof. There are different definitions for the hyperbolic cross and the Sobolev spaces with mixed smoothness. Up to constants the considered quantities coincide, cf. [KSU15]. The assertion (i) is given in [BKUV17, Lemma 2]. To show (ii), we use $\sup_{\|f\|_{H^s_{\text{mix}}}\le 1} \|f - P_{V_{\text{MZ}}} f\|_{L_2}^2 = R^{-s}$ and (i). The second part of (ii) is given in [BDSU16, Theorem 6.11 (iii)].

Assertion (iii) follows from Lemmata 2 and 3 in [BKUV17]. To show (iv), we use [BKUV17, Theorem 2] with $\alpha = s$, $\beta = \gamma = 0$ and Lemmata 2 and 3 from [BKUV17] again.

Note, that the Fourier coefficients of the approximation in [BKUV17] (formula 2.3) are computed by applying the adjoint of the Fourier matrix. Because of the reconstructing property, this coincides with the least squares approximation presented here, i.e., with $\boldsymbol{W} = \operatorname{diag}(1/M, \dots, 1/M)$ we have

$$S_{V_{\text{MZ}}}^{\boldsymbol{X}_{\text{MZ}}} f = \sum_{\boldsymbol{k}\in I_{\text{MZ}}} a_{\boldsymbol{k}} \exp(2\pi \mathrm{i}\langle \boldsymbol{k}, \cdot\rangle)$$

with $\boldsymbol{a} = (\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{f} = \frac{1}{M}\boldsymbol{L}^*\boldsymbol{f}$. ■

Similar bounds were obtained for the kernel method operating on rank-1 lattice points, cf. [KKK$^+$21a]. For random points one has the rate M^{-2s} and additional logarithmic terms, however. So, using full rank-1 lattices we lose half the rate of convergence in the main order in comparison to the results from Section 7.2.1. In the next section we will present a mixture of both approaches fixing that drawback.

7.2.3 Sampling recovery with subsampled L_2-MZ points

The next step is to subsample the L_2-MZ points like in Section 6.3 whilst paying attention to the approximation error. With that we obtain an error behavior similar to Section 7.2.1 and are still able to use the fast algorithms associated to the L_2-MZ points. This strategy is already published in [BKPU23].

In contrast to Section 6.3, we subsample two frames and one Bessel sequence simultaneously from the points $\boldsymbol{X}_{\text{MZ}}$ coming from an initial L_2-MZ inequality. In order to control the subsampling we use a change of measure via a convex combination of the respective probability densities. Each point

$\boldsymbol{x}^i \in \boldsymbol{X}_{\mathrm{MZ}}$ will be drawn with respect to the following probability density:

$$\varrho_i = \frac{\omega_i \sum_{k\in I} |\eta_k(\boldsymbol{x}^i)|^2}{3\sum_{j=1}^{M} \omega_j \sum_{k\in I} |\eta_k(\boldsymbol{x}^j)|^2} + \frac{\omega_i \sum_{k\in I_{\mathrm{MZ}}\setminus I} |e_k(\boldsymbol{x}^i)|^2}{3\sum_{j=1}^{M} \omega_j \sum_{k\in I_{\mathrm{MZ}}\setminus I} |e_k(\boldsymbol{x}^j)|^2} + \frac{\omega_i}{3}\,. \tag{7.11}$$

Since $\varrho_i \geq 0$ and $\sum_{i=1}^{M} \varrho_i = 1$, these are proper density weights. In general, the spectral decomposition has to be know in order to compute these weights. For specific examples, like Sobolev spaces on the torus $H^s(\mathbb{T})$, cf. Section 3.5.1, the basis are the exponential functions simplifying the above density: they have absolute value one resulting in the above density to be constant and the resulting measure being uniform.

For the later analysis we need some preparations, starting with a concentration inequality for random infinite matrices similar to Theorem 7.6.

Lemma 7.12. *Let Assumptions 7.9 hold and let $\boldsymbol{X} = \{\boldsymbol{x}^i\}_{i\in J}$, $|J| = n$, be points drawn i.i.d. from $\boldsymbol{X}_{MZ}$ with respect to the discrete density weights ϱ_i, defined in* (7.11). *Further, for $I \subseteq I_{MZ}$, we define the matrices*

$$\boldsymbol{W} = \operatorname{diag}(\omega_i/\varrho_i)_{i\in J} \quad \textit{and} \quad \boldsymbol{\Phi} = (e_k(\boldsymbol{x}^i))_{i\in J, k\in I_{MZ}\setminus I}\,.$$

Then we have with probability exceeding $1 - 2^{3/4}\exp(-t)$

$$\frac{1}{n}\left\|\boldsymbol{W}^{1/2}\boldsymbol{\Phi}\right\|_{2\to 2}^2 \leq \frac{63(\log(n)+t)}{n} B \sum_{k\in I_{MZ}\setminus I} \sigma_k^2 + 2B \sup_{k\notin I} \sigma_k^2\,.$$

Proof. Since points with probability $\varrho_i = 0$ will almost surly not be drawn, we assume $\varrho_i > 0$. Let

$$\boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}\setminus I} = \Big(\eta_k(\boldsymbol{x}^i)\Big)_{\boldsymbol{x}^i\in \boldsymbol{X}_{\mathrm{MZ}}, k\in I_{\mathrm{MZ}}\setminus I} = \boldsymbol{\Phi}\operatorname{diag}(\sigma_k^{-1})_{k\in I_{\mathrm{MZ}}\setminus I}.$$

Our objective is to apply Theorem 6.1. To this end we define the vectors $\boldsymbol{u}^i = \sqrt{\omega_i/\varrho_i}(e_k(\boldsymbol{x}^i))_{k\in I_{\mathrm{MZ}}\setminus I}$. Note,

$$\sum_{i\in J} \boldsymbol{u}^i \otimes \boldsymbol{u}^i = (\boldsymbol{\Phi})^* \boldsymbol{W}\boldsymbol{\Phi}\,.$$

Now we estimate $\|\boldsymbol{u}^i\|_2^2$ and $\|\mathbb{E}\boldsymbol{u}^i \otimes \boldsymbol{u}^i\|_{2\to 2}$. We bound $\|\boldsymbol{u}^i\|_2$ by using the L_2-MZ inequality and $\|e_k\|_{L_2}^2 = \sigma_k^2$

$$\begin{aligned}\|\boldsymbol{u}^i\|_2^2 &\leq 3 \sum_{k\in I_{\mathrm{MZ}}\setminus I} \sum_{j=1}^{M} \omega_j |e_k(\boldsymbol{x}^j)|^2 \\ &\leq 3B \sum_{k\in I_{\mathrm{MZ}}\setminus I} \|e_k\|_{L_2}^2 \\ &= 3B \sum_{k\in I_{\mathrm{MZ}}\setminus I} \sigma_k^2 \,.\end{aligned}$$

We further have by the compatibility of the spectral norm and Lemma 6.2

$$\begin{aligned}\left\|\mathbb{E}\boldsymbol{u}^i \otimes \boldsymbol{u}^i\right\|_{2\to 2} &= \left\|\boldsymbol{W}_{\mathrm{MZ}}^{1/2} \boldsymbol{\Phi}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}\setminus I}\right\|_{2\to 2}^2 \\ &\leq \left\|\boldsymbol{W}_{\mathrm{MZ}}^{1/2} \boldsymbol{L}_{\boldsymbol{X}_{\mathrm{MZ}}}^{I_{\mathrm{MZ}}\setminus I}\right\|_{2\to 2}^2 \left\| \operatorname{diag}(\sigma_k)_{I_{\mathrm{MZ}}\setminus I}\right\|_{2\to 2}^2 \\ &\leq B \sup_{k\in I_{\mathrm{MZ}}\setminus I} \sigma_k^2 \,,\end{aligned}$$

which gives the assertion after applying Proposition 4.9. ■

We now formulate a central result of this section, which bounds the worst-case reconstruction error for the least squares approximation where the points are drawn randomly from a discrete set of points fulfilling an L_2-MZ inequality, cf. middle of Figure 6.3.

Theorem 7.13. *Let Assumption 7.9 hold and $I \subseteq I_{MZ} \subseteq \mathbb{N}$, $|I| \geq 3$, be an index set. For $n \in \mathbb{N}$, $t > 0$, and $r \geq 1$ such that*

$$n := \left\lceil \frac{36B}{A} |I| (\log |I| + t) \right\rceil \leq |I|^r,$$

let $\boldsymbol{X} = (\boldsymbol{x}^i)_{i\in J}$, $|J| = n$, be points drawn i.i.d.from $\boldsymbol{X}_{MZ}$ with respect to the discrete density weights ϱ_i. Then we have with probability exceeding $1 - 4\exp(-t)$

$$\begin{aligned}\sup_{\|f\|_{H(K)}\leq 1} \|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \leq \frac{9B}{A} \sup_{k\notin I} \sigma_k^2 + \frac{7r}{|I|} \sum_{k\in I_{MZ}\setminus I} \sigma_k^2 \\ + \frac{12}{A} \sup_{\|f\|_{H(K)}\leq 1} \|f - P_{V_{MZ}} f\|_\infty^2 \,,\end{aligned}$$

where $\boldsymbol{W} = \operatorname{diag}(\omega_i/\varrho_i)_{i\in J}$, $V = \operatorname{span}\{\eta_k\}_{k\in I}$, and $S_V^{\boldsymbol{X}} = S_V^{\boldsymbol{X}}(\boldsymbol{W})$.

Proof. Without loss of generality let $\varrho_i > 0$. We begin by defining two events. The first one bounds the spectral norm of the least squares matrix

$$S_1 = \Big\{ \boldsymbol{X} \subseteq \boldsymbol{X}_{\mathrm{MZ}} : \Big\| \Big(\boldsymbol{L}^* \boldsymbol{W} \boldsymbol{L} \Big)^{-1} \boldsymbol{L}^* \boldsymbol{W}^{1/2} \Big\|_{2\to2}^2 \le \frac{2}{An} \Big\}.$$

Using Theorem 6.1, we have, that this is equivalent to a lower bound of the singular values of $\boldsymbol{W}^{1/2}\boldsymbol{L}$. Via Lemma 6.2 we use the L_2-MZ characterization and the assumption on n in order to apply Theorem 6.8 with $C = 1/3$ and obtain that the above holds with probability $1 - 2\exp(-t)$.

We define the matrix

$$\boldsymbol{\Phi} = \Big(e_k(\boldsymbol{x}^i) \Big)_{\boldsymbol{x}^i \in \boldsymbol{Y}, k \in I_{\mathrm{MZ}} \setminus I} .$$

Event S_2 is the inequalities of Lemma 7.12 which holds with probability at least $1 - 2^{3/4}\exp(-t)$. Together we obtain the desired probability

$$\mathbb{P}(S_1 \cap S_2) \ge 1 - \mathbb{P}S_1^{\complement} - \mathbb{P}S_2^{\complement} \ge 1 - 4\exp(-t) .$$

It remains to show the bound based on the events S_1 and S_2. On that account we decompose the recovery error using triangle inequality

$$\begin{aligned} \|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 &\le \|f - P_V f\|_{L_2}^2 + \|P_V f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 \\ &\le \|f - P_V f\|_{L_2}^2 + 2\|P_V f - S_V^{\boldsymbol{X}} P_{V_{\mathrm{MZ}}} f\|_{L_2}^2 \\ &\quad + 2\|S_V^{\boldsymbol{X}} P_{V_{\mathrm{MZ}}} f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 . \end{aligned} \tag{7.12}$$

In the following we estimate each of the three summands individually.

Step 1. We estimate the first summand of (7.12) in the same manner as in the first step of the proof of Theorem 7.10:

$$\|f - P_V f\|_{L_2}^2 \le \|f\|_{H(K)}^2 \sup_{k \notin I} \sigma_k^2 .$$

Step 2. Using the invariance of $S_V^{\boldsymbol{X}}$ on $P_V f$, we estimate the second summand of (7.12) by

$$\begin{aligned} &\Big\| P_V f - S_V^{\boldsymbol{X}} P_{V_{\mathrm{MZ}}} f \Big\|_{L_2}^2 = \Big\| S_V^{\boldsymbol{X}} (P_{V_{\mathrm{MZ}}} f - P_V f) \Big\|_{L_2}^2 \\ &= \Big\| (\boldsymbol{L}^* \boldsymbol{W} \boldsymbol{L})^{-1} \boldsymbol{L}^* \boldsymbol{W} ((P_{V_{\mathrm{MZ}}} f - P_V f)(\boldsymbol{x}^i))_{i \in J} \Big\|_2^2 \\ &\le \Big\| (\boldsymbol{L}^* \boldsymbol{W} \boldsymbol{L})^{-1} \boldsymbol{L}^* \boldsymbol{W}^{1/2} \Big\|_{2\to2} \sum_{i \in J} \frac{\omega_i}{\varrho_i} |(P_{V_{\mathrm{MZ}}} f - P_V f)(\boldsymbol{x}^i)|^2 . \end{aligned}$$

We estimate the first factor by the means of event S_1 and write the latter in terms of its coefficients in order to apply event S_2

$$\begin{aligned}\Big\|S_V^{\boldsymbol{X}} P_{V_{\mathrm{MZ}}} f - S_V^{\boldsymbol{X}} f\Big\|_{L_2}^2 &\le \frac{2}{An}\Big\|\boldsymbol{W}^{1/2}\boldsymbol{\Phi}\Big(\langle f, e_k\rangle_{H(K)}\Big)_{k\in I_{\mathrm{MZ}}\setminus I}\Big\|_2^2 \\ &\le \frac{2}{An}\Big\|\boldsymbol{W}^{1/2}\boldsymbol{\Phi}\Big\|_{2\to 2}^2 \sum_{k\in I_{\mathrm{MZ}}\setminus I} |\langle f, e_k\rangle_{H(K)}|^2 \\ &\le \Big(\frac{126B(\log(n)+t)}{An}\sum_{k\in I_{\mathrm{MZ}}\setminus I}\sigma_k^2 + \frac{4B}{A}\sup_{k\notin I}\sigma_k^2\Big)\|f\|_{H(K)}^2 .\end{aligned}$$

Step 3. We start estimating the third summand of (7.12) analogously to Step 2

$$\|P_V f - S_V^{\boldsymbol{X}} P_{V_{\mathrm{MZ}}} f\|_{L_2}^2 \le \frac{2}{An}\sum_{i\in J}\frac{\omega_i}{\varrho_i}|(f - P_{V_{\mathrm{MZ}}} f)(\boldsymbol{x}^i)|^2 .$$

Now we use the third part of ϱ_i, to obtain

$$\begin{aligned}\|P_V f - S_V^{\boldsymbol{X}} P_{V_{\mathrm{MZ}}} f\|_{L_2}^2 &\le \frac{6}{An}\sum_{i\in J}|(f - P_{V_{\mathrm{MZ}}} f)(\boldsymbol{x}^i)|^2 \\ &\le \frac{6}{A}\|f - P_{V_{\mathrm{MZ}}} f\|_\infty^2 .\end{aligned}$$

Overall, applying the estimates 1. – 3. in (7.12), we obtain

$$\begin{aligned}\sup_{\|f\|_{H(K)}\le 1}\|f - S_V^{\boldsymbol{X}} f\|_{L_2}^2 &\le \sup_{k\notin I}\sigma_k^2 + \frac{252B(\log(n)+t)}{An}\sum_{k\in I_{\mathrm{MZ}}\setminus I}\sigma_k^2 \\ &+ \frac{8B}{A}\sup_{k\notin I}\sigma_k^2 + \frac{12}{A}\sup_{\|f\|_{H(K)}\le 1}\|f - P_{V_{\mathrm{MZ}}} f\|_\infty^2 .\end{aligned}$$

By the assumption on $|I|$ we have

$$\frac{252B(\log(n)+t)}{An} \le \frac{7(\log(n)+t)}{|I|(\log(|I|)+t)} \le \frac{7r(\log(|I|)+t)}{|I|(\log(|I|)+t)} = \frac{7r}{|I|} \le 3 ,$$

and obtain the assertion. ■

Remark 7.14. *(i) Given a bounded orthonormal system, i.e., $\|\eta_k\|_\infty = C < \infty$ for all k we can alter events S_1 and S_2 such that we will not need to sample with respect to the density weights ϱ_i and sample with respect to the weights ω_i coming from the L_2-MZ inequality directly.*

(ii) *Whenever the Christoffel-function $\sum_{k\in I}|\eta_k(\boldsymbol{x})|^2$ is constant, independent of the underlying index set I, the three summands in the density weights ϱ_i in* (7.11) *coincide and the number of random points may be divided by three whilst achieving the same error bound.*

(iii) *Up to the quotient of the constants for the L_2-MZ inequality, the first two summands of Theorem 7.13 are the same as in Theorem 7.8 or [KUV21] where points were drawn with respect to a continuous probability measure. The difference is in the latter summand, which only depends on the initial point set. By choosing a suitable initial point set, i.e., a point set satisfying a L_2-MZ inequality for large enough I_{MZ}, we can make this as small as needed. In particular smaller than the first two terms, which, therefore, determine the error decay behavior.*

Next, we use the unweighted frame subsampling from [BSU23] to prove the main result which lowers the number of points to be linear in $|I|$, cf. right of Figure 6.3.

Theorem 7.15. *Let Assumption 7.9 hold and $I \subseteq I_{MZ} \subseteq \mathbb{N}$, $|I| \geq 3$ be an index set. For $n \in \mathbb{N}$, $t > 0$, and $r \geq 1$ such that*

$$n := \left\lceil \frac{36B}{A}|I|(\log|I| + t)\right\rceil \leq |I|^r,$$

let $\boldsymbol{X} = \{\boldsymbol{x}^i\}_{i\in J}$, $|J| = n$ be points drawn i.i.d.from $\boldsymbol{X}_{MZ}$ with respect to the discrete density weights ϱ_i, defined in (7.11). *For $b > 1 + \frac{1}{|I|}$, `PlainBSS` subsampling (cf. [BSU23, Algorithm 3]) the points $\boldsymbol{X}$, we obtain points $\boldsymbol{X}' = \{\boldsymbol{x}^i\}_{i\in J'} \subseteq \boldsymbol{X}$ with $|\boldsymbol{X}'| \leq \lceil b|I|\rceil$ such that we have with probability exceeding* $1 - 4\exp(-t)$

$$\begin{aligned}&\sup_{\|f\|_{H(K)}\leq 1} \|f - S_V^{\boldsymbol{X}'}f\|_{L_2}^2 \\ &\leq 39\,808\,\frac{B}{A}\frac{(b+1)^2}{(b-1)^3}(\log|I| + t)\Bigg(\frac{B}{A}\sup_{k\notin I}\sigma_k^2 + \frac{r}{|I|}\sum_{k\in I_{MZ}\setminus I}\sigma_k^2 \\ &+ \frac{1}{A}\sup_{\|f\|_{H(K)}\leq 1}\|f - P_{V_{MZ}}f\|_\infty^2\Bigg)\end{aligned}$$

with weights $\boldsymbol{W}' = \operatorname{diag}(\omega_i/\varrho_i)_{i\in J'}$ and the least squares approximation $S_V^{\boldsymbol{X}'}f = S_V^{\boldsymbol{X}'}(\boldsymbol{W}')f$ defined in Section 2.2.

Proof. For $g \in V$ we have

$$\|S_V^{\boldsymbol{X}'} g\|_{L_2}^2 \le \left\|((\boldsymbol{L}')^* \boldsymbol{W}' \boldsymbol{L}')^{-1} (\boldsymbol{L}')^* (\boldsymbol{W}')^{1/2}\right\|_{2\to 2}^2 \sum_{i \in J'} \frac{\omega_i}{\varrho_i} |g(\boldsymbol{x}^i)|^2 .$$

By Theorem 6.11 and Theorem 6.1 we have

$$\begin{aligned}\|S_V^{\boldsymbol{X}'} g\|_{L_2}^2 &\le \frac{89(b+1)^2}{(b-1)^3} \frac{n}{|I|} \frac{2}{An} \sum_{i \in J'} \frac{\omega_i}{\varrho_i} |g(\boldsymbol{x}^i)|^2 \\ &\le \frac{3234(b+1)^2 B}{(b-1)^3 A} (\log(|I|) + t) \Big(\frac{2}{An} \sum_{i \in J'} \frac{\omega_i}{\varrho_i} |g(\boldsymbol{x}^i)|^2 \Big) ,\end{aligned}$$

where $|I| \ge 3$ was used in

$$n = \left\lceil \frac{36B}{A} |I| (\log |I| + t) \right\rceil \le \frac{109B}{3A} |I| (\log |I| + t) .$$

Using this estimate in step 2 and 3 of the proof of Theorem 7.13 we obtain the assertion. ■

We payed the logarithmic factor in the bound to work with linearly many points whilst achieving the same error bound as in [NSU21]. Recent progress has shown that the logarithmic factor can be avoided but this utilizes the Kadison-Singer theorem and is not constructive, cf. [DKU23].

7.3 Application on the torus $\mathbb{T}^d$ with rank-1 lattices

In this section we apply our general theory from Section 7.2.3 to Sobolev function spaces with dominating mixed smoothness on the d-dimensional torus $H^s_{\text{mix}}(\mathbb{T}^d)$, see Section 3.6.2. This gives a feel for the general theory and shows that a subset of a rank-1 lattice is able to achieve the good sampling complexity in contrast to the full rank-1 lattice.

Corollary 7.16. *Let $H^s_{\text{mix}}(\mathbb{T}^d)$, $s > 1/2$, be the Sobolev space with dominating mixed smoothness on the d-torus, $I \subseteq I_{MZ} \subseteq \mathbb{Z}^d$, $|I| \ge 3$, be the hyperbolic cross frequency index sets, and $\boldsymbol{X}_{MZ}$ a reconstructing rank-1 lattice for I_{MZ} with M points, cf.* (6.4). *For $b > 1 + \frac{1}{|I|}$ we construct a subset of points $\boldsymbol{X}' = \{\boldsymbol{x}^i\}_{i \in J'} \subseteq \boldsymbol{X}_{MZ}$ with $|\boldsymbol{X}'| \le \lceil b|I| \rceil$ such that we have the following* *bound for the plain least squares approximation*

$$\sup_{\|f\|_{H^s_{\text{mix}}}\le 1} \|f - S_V^{\boldsymbol{X}'} f\|_{L_2}^2 \le C_{d,s}\Big(\frac{b}{b-1}\Big)^3 \log|I|\Big(|I|^{-2s}(\log|I|)^{(d-1)2s} \tag{7.13}$$
$$+ |I_{MZ}|^{-2s+1}(\log|I_{MZ}|)^{2(d-1)s}\Big),$$

with probability $1-4/|I|$ *and* $C_{d,s}$ *a constant depending on* d *and* s*.*

Proof. Since we are dealing with a reconstructing lattice, we have a tight frame $A=B$. We need to apply Theorem 7.15 and Remark 7.14 and use the following inequality

$$\sum_{\boldsymbol{k}\notin I}\sigma_{\boldsymbol{k}}^2 = \sum_{\boldsymbol{k}\notin I}\prod_{j=1}^{d}(1+(2\pi|k_j|)^{2s})^{-1} \lesssim |I|^{-2s+1}(\log|I|)^{2(d-1)s}$$

from [DuTU18, (2.3.2)]. Using Hölder's inequality and the above inequality again, we obtain

$$\begin{aligned}\|f-P_{V_{\text{MZ}}}f\|_\infty^2 &\le \Big|\sum_{k\notin I_{\text{MZ}}}\sigma_{\boldsymbol{k}}^{-2}\sigma_{\boldsymbol{k}}^2|\hat{f}_{\boldsymbol{k}}|\Big|^2\\ &\le \sum_{k\notin I_{\text{MZ}}}\sigma_{\boldsymbol{k}}^{-4}\sum_{k\notin I_{\text{MZ}}}\sigma_{\boldsymbol{k}}^2|\hat{f}_{\boldsymbol{k}}|^2\\ &\lesssim |I_{\text{MZ}}|^{-2s+1}(\log|I_{\text{MZ}}|)^{2(d-1)s}\|f\|_{H^s_{\text{mix}}}^2 .\end{aligned}$$ ■

Remark 7.17 (Optimality for rank-1 lattices)**.** *With the initial hyperbolic cross* I_{MZ} *in Corollary 7.16 slightly bigger than* I *we recover the optimal error bound* $n^{-s}(\log n)^{(d-1)s}$*, cf. [DuTU18, (2.3.2)], up to a logarithmic factor. Thus, the phrase "This main rate (without logarithmic factors) is half the optimal main rate [...] and turns out to be the best possible among all algorithms taking samples on [rank-1] lattices" from [BKUV17] has to be stated more precisely to "…samples on* ***full*** *[rank-1] lattices". We conjecture that the non-constructive approach in [DKU23] (based on Kadison-Singer and Weaver subsampling [NSU21]) may lead to a bound without additional logarithm.*

Next, we support our findings with numerical experiments.

Because of the one-dimensional structure of rank-1 lattices, the matrix-vector product with the system matrix $\boldsymbol{L}_{\boldsymbol{X}_{\text{MZ}}}$ can be carried out using a one-dimensional Fast Fourier Transform (FFT) in $\mathcal{O}(M \log M)$. We may use this algorithm for the subsampled points $\boldsymbol{X} = \{\boldsymbol{x}^i\}_{i \in J}$ and system matrix $\boldsymbol{L}$ too:

$$\boldsymbol{L} = \boldsymbol{P}\boldsymbol{L}_{\boldsymbol{X}_{\text{MZ}}} \quad \text{where} \quad \boldsymbol{P} = \begin{array}{cc} \left(\begin{array}{ccccccc} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ & \vdots & & & & \vdots & \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{array}\right) & \begin{array}{c} 1 \\ 2 \\ \vdots \\ n \end{array} \\ \begin{array}{ccccccc} j_1 & & j_2 & & & j_n & \end{array} & \end{array} . \tag{7.14}$$

Here, we use hyperbolic crosses for the frequency index sets with size $M \lesssim |I_{\text{MZ}}|^2 (\log |I_{\text{MZ}}|)^{-d}$, cf. Theorem 7.11 (iii). Using the FFT this yields the following complexity for the matrix-vector product with the full rank-1 lattice and, subsequently, the subsampled one:

$$\mathcal{O}(M \log M) = \mathcal{O}(|I_{\text{MZ}}|^2 (\log |I_{\text{MZ}}|)^{1-d})$$

with the same memory usage. In contrast, if we naively set up the matrix with the $n \sim |I_{\text{MZ}}| \log |I_{\text{MZ}}|$ random points we have a complexity of $\mathcal{O}(|I_{\text{MZ}}|^2 \log |I_{\text{MZ}}|)$ for the number of arithmetic operations and memory usage of the matrix-vector product. If we further use `BSS` subsampling this reduces to $\mathcal{O}(|I_{\text{MZ}}|^2)$, which is still slower than using the FFT for $d \geq 2$. In general, whenever we start with a rank-1 lattice with fewer than quadratic points, we gain computation speed. For hyperbolic cross frequency index sets this is a logarithmic gain, cf. Theorem 7.11 (iii), and may be more relevant for other frequency sets.

Experiment 1. For the numerical experiments, use the five-dimensional bump function from [BKUV17, KUV21]

$$f(\boldsymbol{x}) = f((x_1, \dots, x_d)^{\mathsf{T}}) = \prod_{j=1}^{5} \Big(\frac{5^{3/4} 15}{4\sqrt{3}} \max \Big\{ \frac{1}{5} - \Big(x_j - \frac{1}{2} \Big)^2, 0 \Big\} \Big)$$

of which we know the exact Fourier coefficients. We remark that $\|f\|_{L_2} = 1$ and $f \in H^{3/2-\varepsilon}_{\text{mix}}(\mathbb{T}^5)$ for $\varepsilon > 0$.

For frequency sets we use hyperbolic crosses suitable for approximating functions with dominating mixed smoothness, cf. Section 3.6.2. They are parametrized by an radius R determinating the size and a shape parameter γ

$$I = I_{\text{MZ}} = \Big\{ \boldsymbol{k} \in \mathbb{Z}^d : \prod_{j=1}^{d} \max \Big\{ 1, \frac{|k_j|}{\gamma} \Big\} \leq R \Big\} ,$$

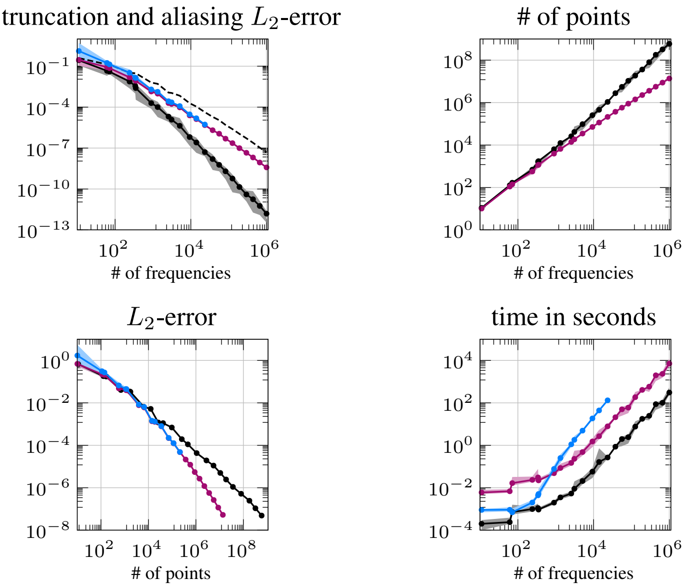


Figure 7.1: Five-dimensional experiment 1 on the torus for different point sets and algorithms. Black: full lattice, magenta: randomly subsampled rank-1 lattice, azure: continuously random points, and dashed in black the truncation error.

where we choose the shape parameter $\gamma = 1/2$ thinning the hyperbolic cross even more without drastically altering the space. Using the probabilistic algorithm from [Kä20] we computed a reconstructing rank-1 lattice $\boldsymbol{X}_{\mathrm{MZ}}$ for I_{MZ}. We used three different techniques for approximation:

- We solved the least squares system for the full rank-1 lattice, which has a direct solution in this case. Using Lemma 6.3 for $\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L} = \boldsymbol{I}$, we omit the inverse matrix all together. The special structure of the points was used to perform the matrix-vector product using a one-dimensional fast Fourier transform in $\mathcal{O}(M \log M)$.

- We randomly subsampled the full rank-1 lattices according to Theorem 7.13 (all points have equal probability $\varrho_i = 1/M$) such that we have $n = \lceil |I| \log |I| \rceil$ points , cf. Figure 6.3. We solved the least squares system iteratively using the same one-dimensional fast Fourier transform as described in (7.14) and set an iteration limit of ten.

- We drew random samples with respect to the Lebesgue measure as suggested in Theorem 7.8 or [KUV21] with the same number of points $n = \lceil |I| \log |I| \rceil$. We set up the system matrix and solved the least squares system with the n random points iteratively. We limit the iterations to at most ten as before.

We computed the L_2-error decomposed into truncation error

$$\|f - P_V f\|_{L_2}^2 = \|f\|_{L_2}^2 - \sum_{k \in I} |\hat{f}_k|^2$$

and the aliasing error

$$\|P_V f - S_V^X f\|_{L_2}^2 = \sum_{k \in I} |\hat{f}_k - \hat{g}_k|^2$$

with $\hat{g}_k$ the Fourier coefficients of the approximation. Further we measured the elapsed time for the computations and stopped the computations when more then 100 gigabytes of memory were used. We repeated this experiment ten times and the minimal, maximal and average results can be seen in Figure 7.1.

- In the upper left figure, we see that for all proposed methods the aliasing error is below the truncation error (dashed line). We emphasize at this point that we have used the minimal choice of I_{MZ}, namely $I = I_{\mathrm{MZ}}$. A larger I_{MZ} would only affect the aliasing error which is dominated by the truncation error which is the bottleneck and cannot be prevented. Differences in the sets I and I_{MZ} might only have a positive effect on the aliasing error, which seems to be superfluous in our setting.

- In the upper right figure, we see that the full rank-1 lattice has more points and is worth subsampling. In particular for 10^6 frequencies we have an oversampling factor of 489 for the full lattice whereas the oversampling for the random choices is $\log(10^6) \leq 14$.

- In the lower left we see the slower error decay of the full rank-1 lattice and the faster error decay of the continuously random points with respect to the number of points. The error of the subsampled rank-1 lattice is the minimum of these two. In particular, for increasing number of points we obtain the better error decay for random points as well as for the randomly subsampled rank-1 lattice.

- In the lower right we see the computation time of the respective methods. The full rank-1 lattice and the subsampled rank-1 lattice differ by a factor of ten. This is due to the same underlying algorithm and the iteration count of ten for the subsampled version. Longer than both took the continuous random points as there the full matrix has to be used in contrast to using the highly tuned FFT. As we showed at the beginning of this section this is slower in regards to the complexity ($\mathcal{O}(|I|^2(\log|I|)^{1-d})$ versus $\mathcal{O}(|I|\log|I|)$). For $23\,483$ frequencies the computations took 258 seconds for the full matrix, whereas merely 4 seconds for the subsampled rank-1 lattice. Furthermore storing the matrix needs $23\,483 \cdot 212\,432 \cdot 16\,\text{bytes} \approx 80\,\text{gigabytes}$, where $16\,\text{bytes}$ is the size of a complex floating point number with double precision (which is why the experiment for the continuously random points stopped early). In contrast to that the rank-1 lattice uses around $1\,000\,000$ points and therefore $1\,000\,000 \cdot 16\,\text{bytes} = 16\,\text{megabytes}$ of memory. Even the largest considered rank-1 lattice uses merely $8\,\text{gigabytes}$.

  Note, that a detailed investigation of the computation time need to consider the evaluation time for the function as well. When function evaluations are cheap, the full rank-1 lattice would benefit. On the other hand, when every function evaluation would correspond to the solution of a partial differential equation the subsampled rank-1 lattices would be benefitial.

Experiment 2. We repeat Experiment 1 but additionally `BSS` subsample the randomly subsampled rank-1 lattice further to an oversampling factor of $b = 2$. We need to stop earlier as the `BSS` algorithm is (so far) not suitable for arbitrarily large matrices. The results can be seen in Figure 7.2.

As the random subsampling step was already evaluated in the experiment above, we focus on the `BSS` subsampling. We obtain an even smaller number of points ($|\boldsymbol{X}'| = 2|I|$) while still having the aliasing error slightly smaller than the truncation error. This results in a faster error decay with respect to the number of points, cf. lower left of Figure 7.2. We also see that the `BSS` algorithm takes way more time, cf. lower right of Figure 7.2. But this is a precomputation step and only has to be done once with the actual iterative solver for the solution not suffering from this.

Overall, we see that the theory is applicable for approximating functions from samples in subsampled rank-1 lattices. The numerical experiments show the practicality of this method, which is even better than the theory suggests as the choice $I = I_{\mathrm{MZ}}$ was possible without deteriorating the error.

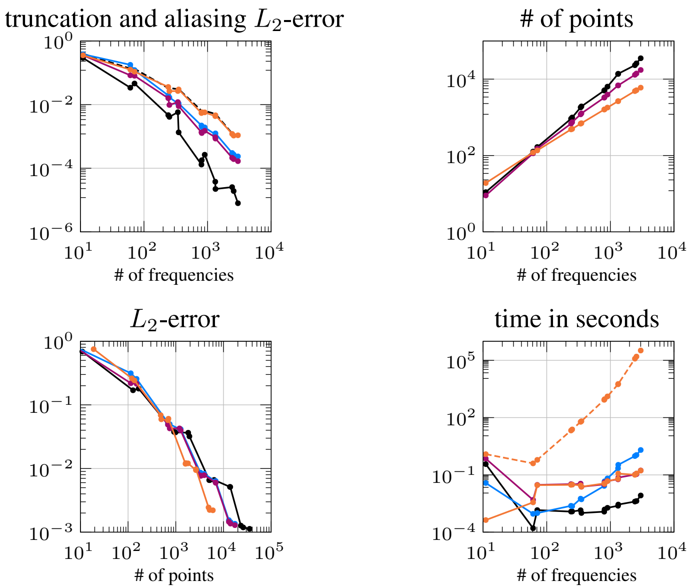


Figure 7.2: Five-dimensional experiment 2 on the torus for different point sets and algorithms. Black: full lattice, magenta: randomly subsampled lattice, orange: random and BSS subsampled lattice (dashed in orange: time of the BSS precomputation step) azure: continuously random points, and dashed in black the truncation error.

7.4 Comparison to the kernel method

We have seen that the least squares approximation achieves the best possible error in the worst-case setting while being stable without the need for further regularization other than the inherent oversampling. It is still worth mentioning that there are also other methods used in practice. Particularly interesting is the **kernel method** $S^{\boldsymbol{X}}$ directly associated to the RKHS setting

$$(S^{\boldsymbol{X}} f)(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}^i) \quad \text{with} \quad \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^\mathsf{T} = \boldsymbol{K}_{\boldsymbol{X}}^{-1} \boldsymbol{f} \tag{7.15}$$

where $(\boldsymbol{K}_{\boldsymbol{X}})_{i,j=1}^{n} = K(\boldsymbol{x}^i, \boldsymbol{x}^j)$ is the kernel matrix.

Note, the kernel matrix from above may be ill-conditioned but is always invertible, cf. [Wen05b, Section 12.2]. With that we immediately have the interpolation property of the kernel method. Based on that, we repeat on the optimality of the kernel method in the worst-case setting of the respective space.

Lemma 7.18. *Let* $\|\cdot\|_{\mathrm{W}}$ *be some error norm of a function space* W *where* $H(K) \hookrightarrow (W, \|\cdot\|_{\mathrm{W}})$. *In terms of the worst-case error, the kernel method* $S^{\boldsymbol{X}}$ *from* (7.15) *beats any other method based on the same samples. In particular, for our least squares approximation* $S_V^{\boldsymbol{X}}$ *from Chapter 2,* $\mathrm{W} = L_2$, *and some function space* V *this reads as*

$$\sup_{\|f\|_{H(K)}\le 1} \left\|f - S^{\boldsymbol{X}} f\right\|_{L_2} \le \sup_{\|f\|_{H(K)}\le 1} \left\|f - S_V^{\boldsymbol{X}} f\right\|_{L_2} .$$

The proof is in [Wen05b, Section 13.2] and goes back to [MR77]. For the sake of readability we will give a proof here as well. We show the statement only for the least squares approximation but it works the same for arbitrary methods.

Proof. Let $g \in H(K)$ be such that $\|g\|_{H(K)} \le 1$ and $g(\boldsymbol{x}^i) = 0$ for $1 \le i \le n$. Then $S_V^{\boldsymbol{X}}(g) = S_V^{\boldsymbol{X}}(-g)$ and we obtain

$$\begin{aligned} \|g\|_{\mathrm{W}} &\le \frac{1}{2}\left(\left\|g - S_V^{\boldsymbol{X}}(g)\right\|_{\mathrm{W}} + \left\|g + S_V^{\boldsymbol{X}}(-g)\right\|_{\mathrm{W}}\right) \\ &\le \max\left\{\left\|g - S_V^{\boldsymbol{X}}(g)\right\|_{\mathrm{W}}, \left\|-g - S_V^{\boldsymbol{X}}(-g)\right\|_{\mathrm{W}}\right\} \\ &\le \sup_{\|h\|_{H(K)}\le 1} \left\|h - S_V^{\boldsymbol{X}} h\right\|_{\mathrm{W}} . \end{aligned}$$

Since the kernel method $S^{\boldsymbol{X}}$ is an orthogonal projection onto $H(K)$, we have that the residual $f - S^{\boldsymbol{X}} f$ is orthogonal to the kernel approximation $S^{\boldsymbol{X}} f$ in $H(K)$. Thus, for a general $f \in H(K)$ with $\|f\|_{H(K)} \le 1$ we obtain $\|f - S^{\boldsymbol{X}} f\|_{H(K)}^2 = \|f\|_{H(K)}^2 - \|S^{\boldsymbol{X}}(f)\|_{H(K)}^2 \le 1$. With the interpolation property of the kernel method we additionally have $(f - S^{\boldsymbol{X}}(f))(\boldsymbol{x}^i) = 0$ for $i = 1, \dots, n$. Thus, we may use $f - S^{\boldsymbol{X}} f$ in place of g. Taking the supremum we obtain the assertion. ■

Thus, using the point constructions from the preceding section in the kernel approximation, we achieve an error as good as the least squares approximation itself. But, as least squares is already optimal, there is no gain. E.g. similar to the approximation bounds for rank-1 lattices in Theorem 7.11 there are results from the kernel perspective in [KKK$^+$21b, Section 3]. One difference is that all computations are done in spatial domain with the following consequences:

- With the kernel method we do not have to know the spectral functions η_k, for which in some scenarious lots of computations are needed, cf. Chapter 3. Instead, we may start with a kernel without thinking about the underlying space. Especially when approximating in high dimensions and kernels having small support the corresponding kernel matrix $\boldsymbol{K}_{\boldsymbol{X}}$ is expected to be sparse, which speeds up matrix-vector multiplications.

- On the downside, when having two close points $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$, the kernel matrix $\boldsymbol{K}_{\boldsymbol{X}}$ will have two similar columns/rows yielding a high condition number. In practice some sort of regularization is often used to regain numerical stability. However, this introduces additional errors.

  Further, when the kernel is global the way of speeding up matrix-vector multiplications is by using the Mercer representation of the kernel

  $$K(x,y) \coloneqq \sum_{k=1}^{\infty} \sigma_k^2 \overline{\eta_k(y)} \eta_k(x)\,,$$

  which is then truncated and used for a matrix decomposition in order to apply fast algorithms, cf. [APSV18]. But truncating the kernel, one ends up with a similar approximation as with the least squares approximation with additional possibilities for numerical errors.

In the end, there is no algorithm to solve all problems and it comes down to the specific setting at hand. For instance, when working with PDEs the series representation attained from the least squares approximation allows to compute derivatives or norms of the approximation immediately when derivatives or norms of the basis functions are known due to linearity. This is often called spectral method as it is based on computing eigenfunctions of the partial derivative operator similar to how we introduced the Sobolev spaces in Section 3.5.

Chapter 8

Least squares in statistical learning

Statistical learning deals with approximating individual functions based on data samples which may be perturbed by noise. This is the everyday task in data mining, machine learning, or nonparametric statistics. In the learning framework, the data samples $\boldsymbol{z} = (z_1, \dots, z_n)$ with $z_i = (\boldsymbol{x}^i, y_i) \in D \times \mathbb{K}$ are independent and identically distributed (i.i.d.) according to some unknown **source measure** p on $D \times \mathbb{K}$. Because of the random nature of the problem, tools from statistics and probability theory are essential next to mathematical/numerical analysis and computer science. For in-depth introductions to the topic we refer to [Vap00, CS02, CZ07, SC08, HTF09]. In contrast to Chapter 7 there is only one function to approximate and points are given rather than chosen.

The theory we present in this chapter also covers a field, which attracts rising attention in the recent years: the covariate shift setting, a subfield of transfer learning. The general goal of learning is to find a function $g\colon D \to \mathbb{K}$ in some hypothesis space $\mathcal{H}$ minimizing the **risk**

$$\mathcal{E}(g) := \int_{D\times\mathbb{K}} |g(\boldsymbol{x}) - y|^2 \, \mathrm{d}q(\boldsymbol{x}, y) \tag{8.1}$$

with respect to some **target measure** q on $D \times \mathbb{K}$, where we almost exclusively use the squared loss function $|\cdot|^2$ but other options are possible.

- In the **classical learning setting** the source and target measure coincide, i.e., $p = q$. E.g. training a self-driving car to drive in the city with driving data of that car in that city.

- In **transfer learning** the knowledge from solving one problem is used to solve a related problem. Formally, that means the source and target measure may differ: $p \neq q$. E.g. training a self-driving car to drive in the city with driving data from a bicycle on a training course.

- Because the general transfer learning problem may be arbitrarily difficult, we rely on the so-called **covariate shift assumption**, cf. [Shi00, HGB$^+$06, GMM$^+$22]. Here, only the probabilities of inputs in the source and the target domains (marginal probabilities) $\varrho_S(x)$ and $\varrho_T(x)$ differ, while the conditional probability $p(y|x)$ is the same under both the source and the target measure. This means that the joint probabilities $p(x, y)$, $q(x, y)$ can be factorized as the following products

$$p(x,y) = p(y|x)\varrho_S(x) \quad \text{and} \quad q(x,y) = p(y|x)\varrho_T(x)\,. \tag{8.2}$$

This scenario occurs e.g. when training a self-driving car to drive in the city with driving data from a car on a training course.

We follow [HGB$^+$06] and assume that there is a function $\beta : D \to \mathbb{R}_+$ such that

$$\mathrm{d}\varrho_T(x) = \beta(x)\mathrm{d}\varrho_S(x).$$

Then β is the **Radon-Nikodym derivative** $\frac{\mathrm{d}\varrho_T}{\mathrm{d}\varrho_S}$ of the target measure ϱ_T with respect to the source measure ϱ_S. In our analysis we assume to know β. If this is not the case it is possible to approximate it from sampled points, cf. [GMM$^+$22].

8.1 Basic concepts in statistical learning

In this section we introduce the standard vocabulary of learning theory following [SC08].

We start with defining the optimal approximation also known as **regression function**

$$f_q(\boldsymbol{x}) := \int_{\mathbb{K}} y \;\mathrm{d}q(y|\boldsymbol{x})\,, \tag{8.3}$$

which is the target quantity but utilizes the underlying distribution q which we do not know in practice. The **variance** σ_q^2 of a learning problem is given by

$$\sigma_q^2(\boldsymbol{x}) = \int_{\mathbb{K}} |y - f_q(\boldsymbol{x})|^2 \;\mathrm{d}q(y|\boldsymbol{x}) \quad \text{and} \quad \sigma_q^2 = \int_D \sigma^2(\boldsymbol{x}) \;\mathrm{d}\varrho_T(\boldsymbol{x})\,. \tag{8.4}$$

The regression function f_q is optimal in the sense that it minimizes the risk (8.1) to which the variance σ_q^2 is a lower bound as stated in the next lemma.

Lemma 8.1. *For a learning problem with probability measure $q(x, y) = q(y|x)\varrho_T(x)$ on $D \times \mathbb{K}$, f_q the regression function* (8.3)*, and $g \colon D \to \mathbb{K}$, we have*

$$\mathcal{E}(g) = \|g - f_q\|^2_{L_2(D,\varrho_T)} + \sigma_q^2\,.$$

Proof. The binomial formula gives

$$\begin{aligned}\mathcal{E}(g) &= \mathcal{E}(g - f_q + f_q)\\ &= \int_{D\times\mathbb{K}} |g(\boldsymbol{x}) - f_q(\boldsymbol{x})|^2 \,\mathrm{d}q(\boldsymbol{x}, y) + \int_{D\times\mathbb{K}} |f_q(\boldsymbol{x}) - y|^2 \,\mathrm{d}q(\boldsymbol{x}, y)\\ &\quad + 2\int_{D\times\mathbb{K}} |g(\boldsymbol{x}) - f_q(\boldsymbol{x})||f_q(\boldsymbol{x}) - y| \,\mathrm{d}q(\boldsymbol{x}, y)\,.\end{aligned}$$

The first term does not depend on y and, thus, we omit the integration over $\mathbb{K}$. By definition, the second term is equal σ_q^2. The third term is zero by the definition of the regression function f_q. ■

For an algorithm we cannot access the underlying distributions p or q but rather have to work with a discretized version of p using a set of training data $\boldsymbol{z}$. This is then known as **empirical risk minimization**: For a hypothesis space $\mathcal{H}$ it is given by

$$f_{\boldsymbol{z}} := \min_{g\in\mathcal{H}} \mathcal{E}_{\boldsymbol{z}}(g) \quad \text{where} \quad \mathcal{E}_{\boldsymbol{z}} = \sum_{i=1}^{n} \omega_i |f(\boldsymbol{x}^i) - y_i|^2\,.$$

Usually $\omega_i = 1/n$ but for the considered covariate shift setting other choices will prove to be useful. Choosing a finite dimensional hypothesis space $\mathcal{H} = V = \mathrm{span}\{\eta_1, \dots, \eta_{m-1}\}$ this coincides with the weighted least squares approximation from Chapter 2:

$$f_{\boldsymbol{z}} = S_V^{\boldsymbol{X}} \boldsymbol{y} = S_V^{\boldsymbol{X}}(\omega_1, \dots, \omega_n)\boldsymbol{y}\,.$$

Except when $\sigma_q^2 = 0$, which is the noiseless case, we have to deal with a trade-off:

- Having a “rich” hypothesis space $\mathcal{H}$, we will have a smaller error on the training data but possibly big variance on the unseen data;
- Having $\mathcal{H}$ “small”, the variance is small but we may not be able to represent the training data.

Example 8.2. *Let* $x^1, \dots, x^n \in [0, 1]$ *be uniformly distributed on the interval and* $y_i = f(x^i) + \varepsilon_i$ *where* $f\colon [0, 1] \to \mathbb{R}, x \mapsto x^2$ *and* ε_i *is Gaussian noise with variance* 0.1*. As hypothesis space we use polynomials* $\mathcal{H} = V = \mathrm{span}\{1, x, \dots, x^{m-1}\}$*. The least squares approximations for polynomial degree* 0*,* 2*, and* 14 *are depicted in Figure 8.1. Already in this example*

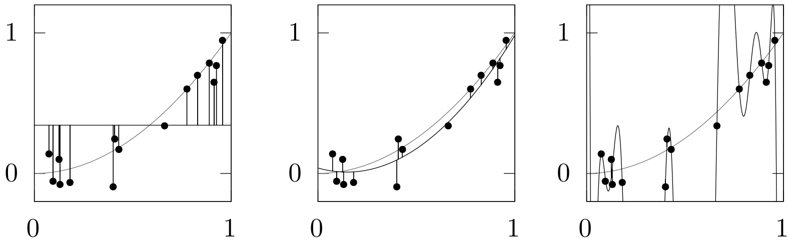


Figure 8.1: Fitting noisy data from a function with polynomials of degree 0, 2, and 14.

the typical over- and underfitting behavior takes effect: Choosing a "small" hypothesis space $\mathcal{H}$ we may not be able to represent the training data but choosing a too "rich" hypothesis space $\mathcal{H}$ the error on the training set is small but the variance on the unseen data may be big.

8.2 Error guarantees for least squares approximation

In this section we bound the quantity $\|f_q - S_V^{\boldsymbol{X}} \boldsymbol{y}\|_{L_2(D,\varrho_T)}$, which is the same as bounding the risk $\mathcal{E}(S_V^{\boldsymbol{X}} \boldsymbol{y})$ up to the additive constant σ_q^2, cf. Theorem 8.1. For individual function approximation the majority of L_2-error bounds are stated in expectation, cf. [Bar02, Theorem 1.1] for penalized least squares, [CDL13, Theorem 3] for plain least squares or, [HNP22, Theorem 4.1], and [KUV21, Theorem 6.1] for weighted least squares approximation. Bounds holding with high probability are known for polynomial approximation, cf. [MNvST14, Theorem 3], wavelet approximation, cf. [LPU23, Theorems 3.20 & 3.21], or in a more general setting including noise in [CM17, Theorem 4.3] with the coarser L_∞-norm instead of the natural L_2-norm in the estimate. Further, in [CM17, Theorem 4.1] an error bound with the natural L_2-norm estimate is presented in expectation with the same behavior as we will present with high probability. The results enable to give performance guarantees for model selection strategies like the balancing principle [PL13, LMP20] or cross-validation [BHP20, BH22] in Chapter 9.

Introducing a benchmark is difficult in the individual function context, as an algorithm predicting the target function is best for that particular function but would fail for others. To still have a benchmark for least squares approximation, we use the underlying approximation space. Given an $m-1$-dimensional function space $V = \operatorname{span}\{\eta_1, \ldots, \eta_{m-1}\} \subseteq L_2$, the best possible

L_2-approximation to $f\colon D \to \mathbb{K}$ in V is given by the projection

$$P_V f_q = \arg\min_{g\in V} \|f_q - g\|_{L_2(D,\varrho_T)} .$$

Note, since V is finite-dimensional the minimum is actually attained. Let $\boldsymbol{f} = (f_q(\boldsymbol{x}^1), \dots, f_q(\boldsymbol{x}^n))^\mathsf{T}$ and $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{f}$. Because of its linearity, the approximation error $\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}$ splits as follows:

$$\begin{aligned}\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 &= \|f_q - P_V f_q\|_{L_2}^2 + \|P_V f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2\\ &\le \underbrace{\|f_q - P_V f_q\|_{L_2}^2}_{\text{truncation error}} + 2\underbrace{\|P_V f_q - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2}_{\text{discretization error}} + 2\underbrace{\|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_{L_2}^2}_{\text{noise error}} .\end{aligned}$$

For a fixed number of points n, we have a look at the behavior with respect to m, the dimension of the approximation space V. The truncation error is the best possible benchmark and depends on the decay of the coefficients $\langle f_q, \eta_k\rangle$, which is usually polynomially m^{-s} for some rate $s \ge 1$ depending on f_q and the choice of V.

To investigate of the discretization error, we first have a look at the noiseless case, i.e., $\sigma_q^2 = 0$. The next result shows, that the discretization error obeys the same decay as the truncation error and is heavily based on [LPU23, Theorem 3.20] which extends to a more general setting. It was originally stated in [Bar23, Theorem 3.2].

Theorem 8.3. *Let* $\mathrm{d}\varrho_S = 1/\beta\,\mathrm{d}\varrho_T$ *be probability measures with* β *their respective Radon-Nikodym derivative. Further let* $\sigma_q^2 = 0$*, i.e., no noise and the conditional probability* $p(y|x) = \delta_{f_q(x)}$ *with* f_q *the regression function* (8.3)*, and let* $\boldsymbol{z} = \{(\boldsymbol{x}^1, y_1), \dots, (\boldsymbol{x}^n, y_n)\}$ *be samples drawn according to* $p(x,y) = p(y|x)\varrho_T(x)$*. Further, let* $t \ge 0$ *and* V *be an* $m-1$*-dimensional function space with an orthonormal basis* $\eta_1, \dots, \eta_{m-1}$ *satisfying*

$$10\|\beta(\cdot)N(V,\cdot)\|_\infty(\log(m-1) + t) \le n .$$

Then, for $S_V^{\boldsymbol{X}}\boldsymbol{f}$ *the weighted least squares approximation defined in Section 2.2 with* $\omega_i = \beta(\boldsymbol{x}^i)$ *and exact data* $\boldsymbol{f} = (y_1, \dots, y_n)^\mathsf{T}$*, we have*

$$\begin{aligned}\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2 &\le 8\Big(\|f_q - P_V f_q\|_{L_2} + \sqrt{\frac{t}{n}}\|f_q - P_V f_q\|_{L_\infty}\Big)^2\\ &\le 8\Big(1 + \sqrt{\frac{N(V)}{\|\beta(\cdot)N(V,\cdot)\|_\infty}}\Big)^2 \|f_q - P_V f_q\|_{L_2}^2 ,\end{aligned}$$

with probability exceeding $1 - 2\exp(-t)$ *where* $L_2 = L_2(D, \varrho_T)$ *and* $L_\infty = L_\infty(D, \varrho_T)$*.*

Proof. For abbreviation, we use $e_2 = \|f_q - P_V f_q\|_{L_2}$ and $e_\infty = \|f_q - P_V f_q\|_{L_\infty}$. Further, for $\boldsymbol{L}$, $\boldsymbol{W}$ the least squares and weight matrix as in Section 2.2, we define the event

$$A \coloneqq \Big\{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n \in D : \frac{n}{2} \le \|\boldsymbol{W}^{1/2}\boldsymbol{L}\|_{2\to 2}^2\Big\} \tag{8.5}$$

which has probability $\mathbb{P}(A) \ge 1 - \exp(-t)$ by Lemma 6.4 and the assumption on V. We split the approximation error

$$\|f - S_V^{\boldsymbol{X}} \boldsymbol{f}\|_{L_2}^2 = e_2^2 + \|P_V f_q - S_V^{\boldsymbol{X}} \boldsymbol{f}\|_{L_2}^2 \,.$$

Due to the invariance of $S_V^{\boldsymbol{X}}$ to functions in V, we pull it in front and use compatibility of the operator norm to obtain

$$\|f - S_V^{\boldsymbol{X}} \boldsymbol{f}\|_{L_2}^2 \le e_2^2 + \|S_V^{\boldsymbol{X}}\|_{2\to L_2}^2 \sum_{i=1}^n \beta(\boldsymbol{x}^i)|(f - P_V f_q)(\boldsymbol{x}^i)|^2 \,.$$

By Theorem 6.1 and the event (8.5), we have $\|S_V^{\boldsymbol{X}}\|_{2\to L_2}^2 = \|\boldsymbol{W}^{1/2}\boldsymbol{L}\|_{2\to 2}^{-1} \le 2/n$. Thus,

$$\|f - S_V^{\boldsymbol{X}} \boldsymbol{f}\|_{L_2}^2 \le 3e_2^2 + \frac{2}{n}\sum_{i=1}^n \Big|\beta(\boldsymbol{x}^i)|(f - P_V f_q)(\boldsymbol{x}^i)|^2 - e_2^2\Big| \,.$$

It remains to estimate the latter summand. To this end we define

$$\xi_i = \beta(\boldsymbol{x}^i)|(f - P_V f_q)(\boldsymbol{x}^i)|^2 - e_2^2,$$

which is mean-zero since we sample with respect to the distribution ϱ_S. Further, we have

$$\begin{aligned}\mathbb{E}(\xi_i^2) &= \mathbb{E}\Big((\beta(\boldsymbol{x}^i))^2|(f - P_V f_q)(\boldsymbol{x}^i)|^4\Big) - e_2^4 \\ &\le \|f - P_V f_q\|_{L_\infty}^2 e_2^2 - e_2^4 \\ &\le e_2^2(e_2 + e_\infty)^2 \,,\end{aligned}$$

and

$$\|\xi_i\|_\infty \le \sup_{x\in D}\Big|\beta(\boldsymbol{x})|(f - P_V f_q)(\boldsymbol{x})|^2 - e_2^2\Big| \le e_\infty^2 + e_2^2 \,.$$

Thus, the conditions in order to apply Bernstein, cf. Theorem 4.1, are fulfilled:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \xi_i &\le \frac{2t}{3n}\Big(e_2^2 + e_\infty^2\Big) + \sqrt{\frac{2t}{n}}\Big(e_\infty e_2 + e_2^2\Big) \\ &\le \Big(\frac{2}{3} + \sqrt{2}\Big)e_2^2 + \sqrt{\frac{2t}{n}}e_\infty e_2 + \frac{2t}{3n}e_\infty^2 \end{aligned} \tag{8.6}$$

with probability $1-\exp(-t)$, where $t \le n$ was used in the last inequality. Thus,

$$\begin{aligned}\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2 &\le \Big(\frac{13}{3}+2\sqrt{2}\Big)e_2^2 + \sqrt{\frac{8t}{n}}e_\infty e_2 + \frac{4t}{3n}e_\infty^2 \\ &\le \Big(\frac{13}{3}+2\sqrt{2}\Big)\Big(e_2 + \sqrt{\frac{t}{n}}e_\infty\Big)^2 .\end{aligned}$$

By union bound we obtain the overall probability exceeding the sum of the probabilities of events given by (8.5) and (8.6).

The **second bound** is achieved in the following way: For any function $g = \sum_{k=1}^{m-1}\langle g, \eta_k\rangle_{L_2}\eta_k \in V$ the Hölder's inequality gives an estimate on the L_∞-norm in terms of the L_2-norm:

$$\begin{aligned}\|g\|_{L_\infty} &= \Big\|\sum_{k=1}^{m-1}\langle g,\eta_k\rangle_{L_2}\eta_k\Big\|_{L_\infty} \\ &\le \Big\|\Big(\sum_{k=1}^{m-1}|\langle g,\eta_k\rangle_{L_2}|^2\Big)^{1/2}\Big(\sum_{k=1}^{m-1}|\eta_k|^2\Big)^{1/2}\Big\|_{L_\infty} \\ &\le \sqrt{N(V)}\|g\|_{L_2} . \end{aligned} \tag{8.7}$$

Using the assumption on V, we have

$$\begin{aligned}\sqrt{\frac{t}{n}}e_\infty &\le \sqrt{\frac{tN(V)}{n}}e_2 \\ &\le \sqrt{\frac{t}{10(\log(m)+t)}\frac{N(V)}{\|\beta(\cdot)N(V,\cdot)\|_\infty}}e_2 \\ &\le \sqrt{\frac{N(V)}{\|\beta(\cdot)N(V,\cdot)\|_\infty}}e_2 . \qquad \blacksquare\end{aligned}$$

Provided $N(V)/\|\beta(\cdot)N(V,\cdot)\|_\infty$ is finite and given the oversampling condition, Theorem 8.3 states, that the least squares approximation from a finite-dimensional function space V has the same error as the L_2-projection up to a multiplicative constant with high probability. Thus, the discretization error $\|P_V f_q - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2$ (error without noise) only differs by a multiplicative constant in comparison to the error of the L_2-projection $\|f_q - P_V f_q\|_{L_2}^2$.

This improves on [CM17, Theorem 2.1] where the same bound was shown in expectation or bounded by the L_∞-error without the prefactor $\sqrt{t/n}$ with high probability. Note, there exists a distribution ϱ_S such that linear oversampling $m \in \mathcal{O}(n)$ achieves the optimal error in expectation, but this is not constructive, cf. [DC22a].

Next, we investigate the noise error. We have shown in [Bar23, Theorem 1.1] that, next to the truncation error $\|f_q - P_V f_q\|_{L_2}^2$ and discretization error $\|P_V f_q - S_V^{\boldsymbol{X}} \boldsymbol{f}\|_{L_2}^2$, we obtain an additional summand increasing linear in the dimension of V.

Theorem 8.4. *Let* $\mathrm{d}\varrho_S = 1/\beta \, \mathrm{d}\varrho_T$ *and let* $\boldsymbol{z} = \{(\boldsymbol{x}^1, y_1), \dots, (\boldsymbol{x}^n, y_n)\}$ *be a sample drawn according to* $p(x,y) = p(y|x)\varrho_T(x)$ *on* $D \times \{y \in \mathbb{C} : |y| \le K\}$. *Further, let* $t \ge 0$, V *be an* m*-dimensional function space with an orthonormal basis* $\eta_1, \dots, \eta_{m-1}$ *satisfying*

$$10\|\beta(\cdot)N(V,\cdot)\|_\infty(\log(m-1)+t) \le n \,. \tag{8.8}$$

Then, for $S_V^{\boldsymbol{X}} \boldsymbol{y}$ *the weighted least squares approximation defined in Section 2.2 with* $\omega_i = \beta(\boldsymbol{x}^i)$ *and* f_q *the regression function* (8.3), *we have*

$$\begin{aligned}\|f_q - S_V^{\boldsymbol{X}} \boldsymbol{y}\|_{L_2}^2 \le 14\Big(\|f_q - P_V f_q\|_{L_2} + \sqrt{\frac{t}{n}}\|f_q - P_V f_q\|_\infty\Big)^2 \\ + 4\|\beta\|_\infty \|\sigma_q^2\|_\infty \frac{m-1}{n} + \frac{2048Kt\|\beta\|_\infty}{n} \,, \end{aligned} \tag{8.9}$$

with probability exceeding $1 - 3\exp(-t)$ *where* $L_2 = L_2(D, \varrho_T)$ *and* $L_\infty = L_\infty(D, \varrho_T)$.

Proof. Let $\boldsymbol{f} = (f_q(\boldsymbol{x}^1), \dots, f_q(\boldsymbol{x}^n))^\mathsf{T}$ and $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{f}$. We split the approximation error

$$\|f_q - S_V^{\boldsymbol{X}} \boldsymbol{y}\|_{L_2}^2 \le \|f_q - P_V f_q\|_{L_2}^2 + 2\|f_q - S_V^{\boldsymbol{X}} \boldsymbol{f}\|_{L_2}^2 + 2\|S_V^{\boldsymbol{X}} \boldsymbol{\varepsilon}\|_{L_2}^2$$

and bound the first two summands as in the proof of Theorem 8.3 with the events given by (8.5) and (8.6). Note, the constant changes from $13/3 + 2\sqrt{2}$ to $23/3 + 4\sqrt{2} \le 14$. Now, we focus on the third summand. Applying the Hanson-Wright inequality from Corollary 4.5 gives

$$\begin{aligned}\|S_V^{\boldsymbol{X}} \boldsymbol{\varepsilon}\|_{L_2}^2 &= \|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{\varepsilon}\|_2^2 \\ &\le (2(m-1)\|\sigma_q^2\|_\infty + 1024Kt)\|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\|_{2\to2}^2 \end{aligned} \tag{8.10}$$

with probability $1 - \exp(-t)$. Now we apply $\|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\|_{2\to2}^2 \le \|\beta\|_\infty \|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}^{1/2}\|_{2\to2}^2$, Theorem 6.1, and event (8.5) to obtain

$$\begin{aligned}\|S_V^{\boldsymbol{X}} \boldsymbol{\varepsilon}\|_{L_2}^2 &= \|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{\varepsilon}\|_2^2 \\ &\le 4\|\beta\|_\infty\|\sigma_q^2\|_\infty \frac{m-1}{n} + \frac{2048Kt\|\beta\|_\infty}{n} \,. \end{aligned}$$

By union bound we obtain the overall probability exceeding the sum of the probabilities of the events given by (8.5), (8.6), and (8.10). ■

The first line of (8.9) corresponds to the truncation error $\|f_q - P_V f_q\|_{L_2}^2$ and discretization error $\|P_V f_q - S_V^{\boldsymbol{X}} f_q\|_{L_2}^2$, decaying in m. Note, that the L_∞-term with the prefactor $n^{-1/2}$ behaves as the L_2-term whenever β is bounded from below, cf. Theorem 8.3. The second line of (8.9) is the error due to noise, increasing in m, cf. Figure 8.1. This linear behavior in m is approved by [LMP20, Theorem 4.9] (by using the regularization $g_\lambda(\sigma) = 1/(\lambda + \sigma)$ with $\lambda = 0$). This resembles the well-known bias-variance trade off modeling the over- and undersmoothing effects which one wants to balance, cf. Section 8.1 or [GKKW02, PL13].

The behavior of our bound (8.9) is similar to [CM17, Theorem 4.1], which is stated only in expectation. The estimation of the noise error is using a Hanson-Wright concentration inequality, which can be found using different assumptions. Thus, we can replace the noise model by general Bernstein conditions, cf. Lemma 4.5, or sub-Gaussian noise, cf. [RV13].

The Radon-Nikodym derivative $\beta = \frac{\mathrm{d}\varrho_T}{\mathrm{d}\varrho_S}$ and the Christoffel function $N(V, \cdot)$ affect the maximal size of V in the assumption and the amplification of the noise in bound. There are two extremal cases:

(i) Having $\beta(\boldsymbol{x}) = m/N(V, \boldsymbol{x})$, as it was presented in [HD15, NJZ17, CM17, KUV21], we obtain the assumption

$$10\|\beta(\cdot)N(V,\cdot)\|_\infty(\log(m) + t) = 10m(\log(m) + t) \le n\,,$$

which allows for the biggest choice of m in our bound, i.e, logarithmic oversampling. But this spoils $\|\beta\|_\infty$ in the error bound when the Christoffel function attains small values.

(ii) For domains D with bounded measure, we may choose $\beta(\boldsymbol{x}) = \varrho_T(D)$, as it was done in [CDL13, CM17, LPU23]. As all weights $\omega_i = \varrho_T(D)$, $S_V^{\boldsymbol{X}}$ becomes the plain least squares approximation. In this case, $\|\beta\|_\infty$ is minimal and noise is least amplified. But this choice spoils the assumption on the choice of m when the Christoffel function $N(V, \cdot)$ attains big values. This effect is controllable, for instance, when working with a bounded orthonormal system (BOS) ($\|\eta_k\|_\infty \le B$ for some $B > 0$ and all k). Then

$$N(V) \le \sum_{k=0}^{m-1} \|\eta_k\|_\infty^2 \le mB^2$$

and the assumption on the size of V can be replaced by

$$10\|\beta(\cdot)N(V,\cdot)\|_\infty(\log(m)+t) \le 10\varrho_T(D)Bm(\log(m)+t) \le n\,.$$

We explore these intricacies of the covariate shift setting for different ϱ_S and ϱ_T setting on the unit interval $[0,1]$ in Section 8.3.

Next, we prove a bound for the approximation error of least squares approximation in the uniform norm from [Bar23, Theorem 1.2].

Theorem 8.5 (L_∞-error bound with noise)**.** *Let the assumptions of Theorem 8.4 hold and let $P_V^\infty f_q = \arg\min_{g\in V}\|f_q - g\|_{L_\infty}$.*

Then, for $S_V^{\boldsymbol{X}}\boldsymbol{y}$ the weighted least squares approximation defined in Section 2.2 with $\omega_i = \beta(\boldsymbol{x}^i)$, we have with probability exceeding $1-3\exp(-t)$:

$$\begin{aligned}\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_\infty} &\le 4\sqrt{N(V)}\Big(\|f_q - P_V^\infty f_q\|_{L_\infty} + \sqrt{\frac{t}{n}}\|f_q - P_V^\infty f_q\|_{L_2}\Big)\\ &\qquad + 2\sqrt{\|\beta\|_\infty N(V)\|\sigma_q^2\|_\infty\frac{m-1}{n}} + 46\sqrt{\frac{Kt\|\beta\|_\infty}{n}}\,.\end{aligned}$$

with probability exceeding $1-3\exp(-t)$ where $L_2 = L_2(D,\varrho_T)$ and $L_\infty = L_\infty(D,\varrho_T)$.

Proof. For abbreviation, we use $e_2 = \|f_q - P_V^\infty f_q\|_{L_2}$ and $e_\infty = \|f_q - P_V^\infty f_q\|_{L_\infty}$. Using (8.7) we reduce the L_∞-case to the L_2-case which we already covered. We split the approximation error

$$\begin{aligned}\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_\infty} &\le \|f_q - P_V^\infty f\|_{L_\infty} + \|P_V^\infty f - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_\infty} + \|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_\infty\\ &\le e_\infty + \sqrt{N(V)}\|P(f,V,L_\infty) - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2} + \sqrt{N(V)}\|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_{L_2}\,.\end{aligned}$$

Analogously to (8.6) we obtain

$$\begin{aligned}\|P_V^\infty f - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2 &\le \Big(\frac{2}{3}+\sqrt{2}\Big)e_\infty^2 + \sqrt{\frac{2t}{n}}e_\infty e_2 + \frac{2t}{3n}e_2^2\\ &\le \Big(\frac{2}{3}+\sqrt{2}\Big)\Big(e_\infty + \sqrt{\frac{t}{n}}e_2\Big)^2,\end{aligned}$$

where the last inequality follows from $t \le n$. Thus,

$$\begin{aligned}&\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_\infty}\\ &\le \Big(1+\sqrt{\frac{4+6\sqrt{2}}{3}N(V)}\Big)\Big(e_\infty + \sqrt{\frac{t}{n}}e_2\Big) + \sqrt{N(V)}\|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_{L_2}\\ &\le 4\sqrt{N(V)}\Big(e_\infty + \sqrt{\frac{t}{n}}e_2\Big) + \sqrt{N(V)}\|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_{L_2}\,.\end{aligned}$$

Using the same bound as in Theorem 8.4 for $\|S_V^{\boldsymbol{X}}\varepsilon\|_{L_2}$ we obtain the assertion. ■

The bound is similar to [LPU23, Theorem 3.21] in the wavelet setting without noise but we use the best approximation with respect to the more natural L_∞ instead of L_2. In addition to the error of the best approximation we now have the additional factor $N(V)$ due to using the norm estimate $\|g\|_{L_\infty} \le \sqrt{N(V)}\|g\|_{L_2}$ for functions $g \in V$. The same factor appears when approximating the worst-case error where it is known from various examples, e.g. [Tem93b, Thm .1.1], [PU22, Sec. 7]. Optimally, this factor evaluates to $\sqrt{N(V)} = \sqrt{m}$ but may be worse as we will see in the next section. In [KPUU23] a technique was used to improve the factor to $\sqrt{m}$ independent of the Christoffel function.

8.3 Application on the unit interval $[0, 1]$

An interesting example, where different effects of the general theory can be investigated, is the approximation of functions on the unit interval $D = [0, 1]$ from samples given in uniformly random points. In this section we have a look at different scenarios occurring when approximating on the interval $[0, 1]$ and using different measures for ϱ_S, ϱ_T, and different ansatz functions in V. For ϱ_T will use either the Lebesgue measure $\mathrm{d}x$ or the Chebyshev measure $\pi(1-(2x-1)^2)^{-1/2}\,\mathrm{d}x$. As ansatz functions we will consider monomials, which, when orthogonalized with respect to to these measures, are the Legendre polynomials $P_k(x) = \frac{1}{2^k k!}\frac{\mathrm{d}^k}{\mathrm{d}x^k}((2x-1)^2-1)^k$ or Chebyshev polynomials $T_k(x) = \cos(k \arccos(2x-1))$, respectively. We will further consider the bases of the Sobolev spaces $H^1((0,1),\mathrm{d}x)$ and $H^2((0,1),\mathrm{d}x)$ from Theorem 3.26 and Theorem 3.26.

We start by having a look at what the theory predicts and will validate this with numerical experiments afterwards.

8.3.1 Analysis

As for the approximation rates of these four bases the following gives answer depending on the smoothness of the regression function f_q, cf. (8.3).

- Let $s \in \mathbb{N}$, $f_q, \dots, f_q^{(s-1)} \colon [0,1] \to K$ be absolute continuous, and $f_q^{(s)}$ of bounded variation. Using Theorem 3.37, this implies Sobolev regularity $f_q \in H^{s+1/2-\varepsilon}([0,1],\mathrm{d}x)$ for all $\varepsilon > 0$ and by Theorem 3.36

we obtain for the polynomial approximation by Legendre polynomials

$$\begin{aligned}&\|f_q - P_{\operatorname{span}\{P_0,\dots,P_{m-2}\}} f_q\|_{L_2((0,1),\mathrm{d}x)}\\ &= \sqrt{\sum_{k=m-1}^{\infty} \left|\left\langle f, \frac{P_k}{\|P_k\|_{L_2}}\right\rangle_{L_2((0,1),\mathrm{d}x)}\right|^2}\\ &\lesssim \sqrt{\sum_{k=m-1}^{\infty} k^{-(2s+2)}} \lesssim m^{-(s+1/2)}\end{aligned}$$

and analog for Chebyshev polynomials

$$\|f_q - P_{\operatorname{span}\{T_0,\dots,T_{m-2}\}} f_q\|_{L_2((0,1),\pi(1-(2x-1)^2)^{-1/2}\mathrm{d}x)} \lesssim m^{-(s+1/2)}\,.$$

- Let $s \in \{1,2\}$ and V the span of the first $m-1$ basis functions of $H^s([0,1])$ from Theorem 3.26 or Theorem 3.27. Using Theorem 3.12 we obtain by using the orthonormal basis in H^s that the error of the projection is given by the singular m-th singular value. Consequently, we have for $f_q \in H^s$
$$\|f_q - P_V f_q\|_{L_2((0,1),\mathrm{d}x)} \lesssim m^{-s}\,.$$

From this, approximation with Legendre or Chebyshev polynomials seems advantageous in comparison to the $H^s([0,1])$ basis from Theorem 3.26 or Theorem 3.27. Next, we will see what happens when we draw uniform samples, i.e., $\mathrm{d}\varrho_S = \mathrm{d}x$. The deterministic equivalent to uniform sampling are equispaced points. When using these for polynomial interpolation, C. Runge already knew in 1901, that higher degree polynomials lead to oscillatory behavior towards the border which spoil the approximation error. Even though, we do not interpolate, we will observe similar behavior using Legendre and Chebyshev polynomials.

- When using the Legendre polynomials and the Lebesgue error measure $\mathrm{d}\varrho_T = \mathrm{d}x$ we have $\beta \equiv 1$. Since $\|P_k\|^2_{L_2((0,1),\mathrm{d}x)} = 2k+1$ and $P_k(0) = 1$, we have for the Christoffel function given in (3.14)
$$N(V,0) = \sum_{k=0}^{m-1} \frac{|P_k(0)|^2}{\|P_k\|^2_{L_2}} = \sum_{k=0}^{m-1}(2k+1) = m^2\,. \tag{8.11}$$
Plugging this into the assumption (8.8), we require $m \le \sqrt{n}$, i.e., quadratic oversampling is required. The same phenomenon was also observed in [MNvST14].

- When using the Chebyshev target measure, we have $\mathrm{d}\varrho_T = \pi(1-(2x-1)^2)^{-1/2}\,\mathrm{d}x$ and $\beta(x) = \frac{\pi}{4}(1-(2x-1)^2)^{-1/2}$. The Chebyshev polynomials are a BOS, but the distribution β spoils both the assumption on m and the error bound, since β diverges at the border. In this way the Theorem 8.4 has no statement at all. This effect can be circumvented using a padding technique at the border, cf. [PS22].
- Using the basis of the Sobolev-space $H^1[0, 1]$ from Theorem 3.26 or $H^2[0, 1]$ from Theorem 3.27 we have an BOS, i.e., $\|\eta_k\|_\infty \le B$ for $B = \sqrt{6}$ and all k. Thus, with $\beta \equiv 1$ and $N(V) \le Bm$, these basis are suitable for approximation in uniform random points on $D = [0, 1]$ using plain least squares approximation and only logarithmic oversampling by (8.8).

Note, the approximation with polynomials can be saved when sampling with respect to the Chebyshev measure $\mathrm{d}\varrho_S(x) = \pi(1-(2x-1)^2)^{-1/2}\,\mathrm{d}x$:

- For the Chebyshev polynomials we have $\beta \equiv 1$. Since the Chebyshev polynomials are a BOS, this does not spoil our bounds.
- Using the Legendre polynomials ($\mathrm{d}\varrho_T = \mathrm{d}x$) we have $\beta(x) = \pi(1-(2x-1)^2)^{1/2}$. Further, we use [RW12, Lemma 5.1]:

  $$\sqrt{1-(2x-1)^2}|P_k(x)|^2 \le \frac{2}{\pi}\Big(2+\frac{1}{k}\Big)$$

  for $k \ge 1$. Thus, $\|\beta(\cdot)N(V,\cdot)\|_\infty$ and $\|\beta(\cdot)\|_\infty$ are bounded and do not spoil the choice of polynomial degree m, making logarithmic oversampling $m \log m \lesssim n$ possible, nor the error bound.

It turns out, that there is a measure for every kind of Jacobi polynomial such that $\|\beta(\cdot)N(V,\cdot)\|_\infty$ is well-behaved, cf. [Nev79, Section 6.3, Lemma 5].

8.3.2 Numerical Experiments

To support our findings from Section 8.3.1, we give numerical examples. As a test function we use

$$f_q(x) = B_2^{\text{cut}}(x) \quad \text{with} \quad B_2^{\text{cut}}(x) = \begin{cases} -x^2 + 3/4 & \text{for } x \in [0, 1/2] \\ x^2/2 - 3/2x + 9/8 & \text{for } x \in [1/2, 1] \end{cases} \tag{8.12}$$

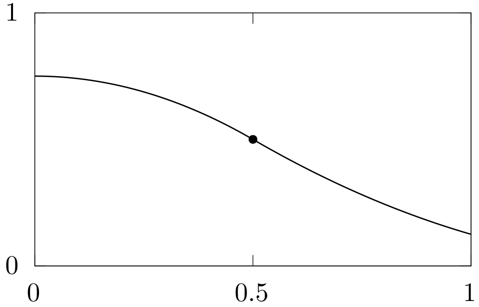


Figure 8.2: Cutout of the B-spline of order two given in (8.12).

which was already considered in [PV15, NP22]. The function B_2^{cut} is shown in Figure 8.2 and is a cutout of the B-spline of order two. It and its first derivative are absolute continuous and the second derivative is of bounded variation. Thus, from Section 8.3.1, we expect the rate $5/2$ for the polynomial basis and 1 and 2 for the H^1 and H^2 bases.

We sample f in $10\,000$ uniformly random points and add $0.1\%M$ Gaussian noise to obtain $\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{\varepsilon}$, where $M = \max_{x\in[0,1]} f(x) - \min_{x\in[0,1]} f(x) = 5/8$. For V we consider the four choices from above: Chebyshev polynomials, Legendre polynomials, the H^1 basis, and the H^2 basis. For $m = \dim(V)$ up to $1\,000$ we do the following:

(i) Compute the minimal and maximal singular values of $1/\sqrt{n}\boldsymbol{W}^{1/2}\boldsymbol{L}$, with $\boldsymbol{W} = \operatorname{diag}(\beta(\boldsymbol{x}^1), \dots, \beta(\boldsymbol{x}^n))$ and $\boldsymbol{L}$ given in (2.1).

(ii) We use least squares approximation with 20 iterations to obtain the approximation $S_V^{\boldsymbol{X}}\boldsymbol{y} = \sum_{k=0}^{m-1} \hat{g}_k\eta_k$, defined in (2.1).

(iii) We compute the L_2-error by using Parseval's equality:

$$\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 = \|f_q\|_{L_2}^2 - \sum_{k=1}^{m-1} |\hat{f}_k|^2 + \sum_{k=1}^{m-1} |\hat{f}_k - \hat{g}_k|^2 \,,$$

where the coefficients $\hat{f}_k = \langle f_q, \eta_k\rangle_{L_2}$ are computed analytically.

(iv) We compute the split approximation error:

$$\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 \le 2\|f - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2 + 2\|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_{L_2}^2 \,,$$

where we compute both quantities separately, again, using Parseval's equality.

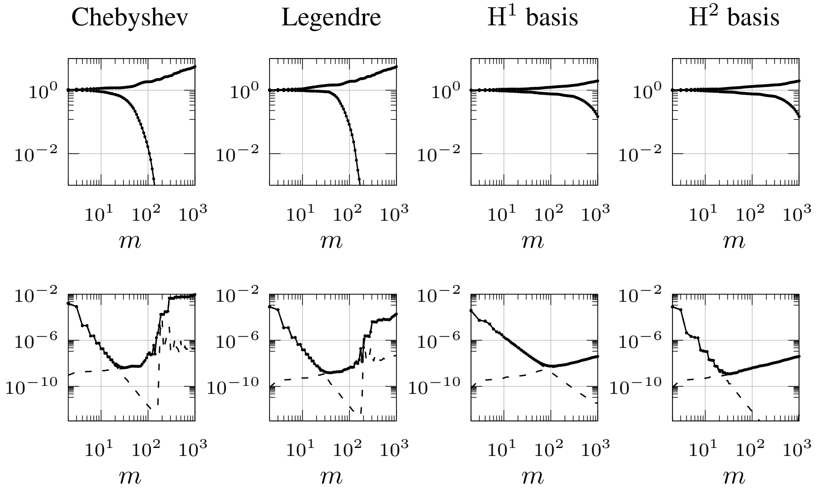


Figure 8.3: One-dimensional experiment for different choices of V. Top row: minimal and maximal singular value of $1/\sqrt{n}\boldsymbol{W}^{1/2}\boldsymbol{L}$. Bottom row: the L_2-approximation error $\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2$ (solid line) split into the error for exact function values $\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2$ and the noise error $\|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_{L_2}^2$ (dashed lines) with respect to m.

The results are depicted in Figure 8.3.

- The smallest singular values for the Chebyshev polynomials and the Legendre polynomials decay rapidly for bigger m. This coincides with the violation of the assumption in Lemma 6.4 for small m:

$$10\|\beta(\cdot)N(V,\cdot)\|_\infty(\log(m) + t) \leq n,$$

  where $\|\beta(\cdot)N(V,\cdot)\|_\infty$ is unbounded in the Chebyshev case and grows quadratic in the Legendre case, cf. (8.11). In this experiment, for $m = 1\,000$ the condition number $\sigma_{\max}(\boldsymbol{W}^{1/2}\boldsymbol{L})/\sigma_{\min}(\boldsymbol{W}^{1/2}\boldsymbol{L})$ exceeded 10^{29} for the algebraic polynomials and was below 14 for the H^s basis.

- The error for exact function values $\|f - S_V^{\boldsymbol{X}}\boldsymbol{f}\|_{L_2}^2$ has decay $3/2$ for H^1 and $5/2$ for the other bases. This conforms with the theory for the polynomial bases. For the H^1 and H^2 bases the theory predicted only decay rate 1 and 2.

- For the noise error $\|S_V^{\boldsymbol{X}}\boldsymbol{\varepsilon}\|_{L_2}^2$ we observe linear growth in $m = \dim(V)$ as predicted in Theorem 8.4. Furthermore, this error is bigger by a factor of around 40 in the Chebyshev case compared to the others. The maximal weight $\|\boldsymbol{W}\|_\infty$ in this case is around 40 as well. The error due to noise in our bound has the factor $\|\beta\|_\infty$ which can be replaced by $\|\boldsymbol{W}\|_\infty$ to sharpen the bound and explain this effect.

This numerical experiment and the theoretical discussion in Section 8.3.1 shows, that the H^1 and the H^2 bases are suitable for approximating functions on the unit interval given in uniform random samples. They are numerically stable in contrast to polynomial approximation with Chebyshev or Legendre polynomials. This becomes clear when looking at the error without noise (dashed line) in Figure 8.3. The error decay of the polynomial basis stops approximately at polynomial degree 100, where the system matrix $\boldsymbol{L}\boldsymbol{W}^{1/2}$ becomes too bad conditioned for the lsqr algorithm to find a good solution with limited iterations. In contrast to that, the error using the H^1- and H^2-bases keeps decaying beyond degree 1 000. In particular, the least squares matrix is well-conditioned and we can limit the iterations when using an iterative solver, cf. [Gre97, Theorem 3.1.1] or Theorem 2.3.

8.4 Numerical experiment on the unit cube $[0,1]^d$

To put the theory to test in a higher-dimensional setting, we repeat the one-dimensional experiment scaled to five dimensions. We do this using Sobolev spaces of dominating mixed smoothness, cf. Section 3.6.2, with the tensorized H^2 basis. For our test function we tensorize the B-Spline cutout

$$f(\boldsymbol{x}) = \prod_{j=1}^{5} B_2^{\text{cut}}(x_j)$$

where B_2^{cut} is defined in (8.12). We increase the number of samples to be 1 000 000 and add Gaussian noise with the following three different variances $\sigma^2 \in \{0.00,\, 0.01M,\, 0.03M\}$ where M is the range of f, i.e., $M = \max_{\boldsymbol{x}\in[0,1]^d} f(\boldsymbol{x}) - \min_{\boldsymbol{x}\in[0,1]^d} f(\boldsymbol{x}) = 5/8$.

This function is in the Sobolev space of dominating mixed smoothness $5/2 - \varepsilon$ for $\varepsilon > 0$. Accordingly, we use the tensorized H^2 basis $\eta_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^{5} \eta_{k_j}(x_j)$ with frequencies in a hyperbolic cross

$$I_R(H^2_{\text{mix}}) \coloneqq \Big\{\boldsymbol{k} \in \mathbb{N}^d : \prod_{j=1}^{5} \sigma_{k_j}^{-2} \le R\Big\}\,.$$

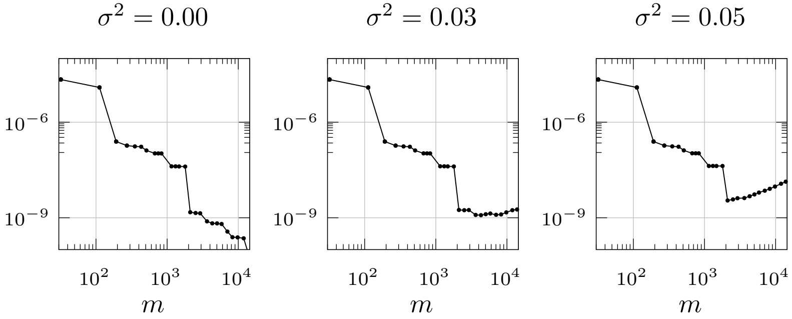


Figure 8.4: Five-dimensional experiment for H^2_{mix}. The lines represent the L_2-error $\|f - S^{\boldsymbol{X}}_V \boldsymbol{y}\|^2_{L_2}$.

Let $V = \operatorname{span}\{\eta_{\boldsymbol{k}} : \boldsymbol{k} \in I_R(H^2_{\text{mix}})\}$ of size m. Since the H^2_{mix} basis is a BOS, we obtain

$$\frac{N(V)}{m} \leq 6\,.$$

With $t = 6$, we satisfy the assumptions of Theorem 8.4 for $m \leq 12\,250$ and obtain a probability exceeding 0.99 for the error bound in Theorem 8.4. For $m = \dim(V)$ up to $10\,000$ ($R \approx 7 \cdot 10^{11}$) we do the following:

(i) We use plain least squares approximation with 20 iterations to obtain the approximation $S^{\boldsymbol{X}}_V \boldsymbol{y} = \sum_{k=1}^{m-1} \hat{g}_k \eta_k$, defined in (2.1).

(ii) We compute the L_2-error by using Parseval's equality analog to the one-dimensional case.

The results are depicted in Figure 8.4.

- The theoretical bounds capture the error behavior well having the decay of the projection and a linear increasing function depending on the noise. We did not plot them as the involved constants deteriorate the bound, especially in the experiments with noise. Here, improving constants in the Hanson-Wright inequality in Theorem 4.5 could be a starting point.

- Furthermore, this experiment shows, that the H^2 basis is easily suitable for high-dimensional approximation as well.

8.5 Comparison to other methods

In this section we compare our theory to other methods from the literature similar to Section 7.4 in the worst-case setting. There it made sense to optimize over all algorithms working with n samples. In the learning setting we always must take the method at hand into consideration, since an algorithm just predicting the target function f_q ignoring all samples is always best but useless otherwise. In the learning setting there is the bias-variance trade-off which one tries to balance by using some sort of regularization. With least squares approximation we regularize by using a finite-dimensional ansatz space V with, e.g. different polynomial degrees. To make different methods comparable, we assume that the optimal regularization parameter is used. Note, in practice this is usually not possible and methods like the cross-validation or the balancing principle have to be applied, cf. Chapter 9. To get a first feeling, we show how to optimally choose the number of ansatz functions and give the resulting approximation rates for a simple example on the one-dimensional torus $\mathbb{T}$. Note, our theory also directly applies to general domains like the multi-dimensional torus $\mathbb{T}^d$ or the cube $[0,1]^d$.

Example 8.6. *We consider the learning problem on the one-dimensional torus $\mathbb{T}$ with i.i.d. uniformly distributed points $\boldsymbol{X}$ and data $\boldsymbol{y}$ such that $|y_i| \le K$ for some constant $K > 0$. We use the least squares approximation $S_V^{\boldsymbol{X}}\boldsymbol{y}$ defined in Section 2.2 with $V = \operatorname{span}\{\exp(2\pi\mathrm{i}(-1)^k\lfloor k/2\rfloor x)\}_{k=1}^{m-1}$ for some even m such that $10m(\log m + t) \le n$. By Theorem 8.4, we have the error bound*

$$\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 \le 56\|f_q - P_V f_q\|_{L_2}^2 + 4\|\sigma_q^2\|_\infty \frac{m}{n} + \frac{2048Kt}{n} \tag{8.13}$$

with probability exceeding $1 - 3\exp(-t)$ where $P_V f_q$ is the projection of f_q onto V and σ_q^2 is the variance (8.4)*.*

Assuming Sobolev smoothness $f_q \in H^s(\mathbb{T})$ for some $s > 1/2$, cf. Section 3.5.1, we have by Theorem 3.24, the polynomial decay $s + 1/2$ of the Fourier coefficients, i.e.,

$$|\langle f_q, \exp(2\pi\mathrm{i}(-1)^k\lfloor k/2\rfloor x)\rangle_{L_2}|^2 \lesssim k^{-2s-1}\,.$$

Thus the error of the projection decays with the rate m^{-s}

$$\|f_q - P_V f_q\|_{L_2}^2 \lesssim \sum_{k=m}^{\infty} k^{-2s} \lesssim m^{-2s}\,.$$

Plugging this into (8.13)*, the error behavior is*

$$\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 \lesssim m^{-2s} + \frac{m}{n}\,,$$

which resembles a bias variance trade-off. Without noise there would only be the decreasing term and we would choose the polynomial degree as big as possible as in Theorem 8.3 or Chapter 7 to achieve the smallest error. With noise, we choose the optimal polynomial degree $m \sim n^{1/(2s+1)}$ minimizing the above to achieve an error rate of $\|f_q - S_V^{\boldsymbol{X}} \boldsymbol{y}\|_{L_2}^2 \lesssim n^{-2s/(2s+1)}$.

If we have an analytic function, the Fourier coefficients decay exponentially and the optimal polynomial degree is $m \sim \log(n)$ giving the error rate $\log(n)/n$. This is optimal for all learning algorithms in a minmax sense, cf. [LMP20, Theorem 5.3].

To put our method in the general context, we use [DVPR10] where general regularization schemes are used in the non-covariate shift setting. In order to do that, we need further notation. We consider the ansatz space $\mathcal{H}$ to be an possibly infinite-dimensional RKHS. For $I_{\mathcal{H}} \colon \mathcal{H} \to L_2$, the inclusion operator, we define the integral operator $T_{\mathcal{H}} = I_{\mathcal{H}}^* I_{\mathcal{H}} \colon \mathcal{H} \to \mathcal{H}$, cf. (3.2).

The smoothness is defined by requiring **general source conditions**, cf. [DVPR10, Definition 3.2], i.e., we assume

$$P_{\mathcal{H}} f_q \in \{f \in \mathcal{H} : f = \varphi(T_{\mathcal{H}})g, \|g\|_{\mathcal{H}} \leq 1\}$$

for a continuous non-decreasing **index function** $\varphi\colon [0,\kappa] \to [0,\infty)$ with $\varphi(0) = 0$. This resembles the coefficient decay of our target function in the considered ansatz space similar to the cosine spaces defined in Theorem 3.34.

Further, we need the **effective dimension** $\mathcal{N}(\lambda) := \operatorname{trace}((\lambda I + T_{\mathcal{H}})^{-1} T_{\mathcal{H}})$ for $\lambda > 0$, which will model the bias of the respective method and is usually assumed to have polynomial $\mathcal{N}(\lambda) \lesssim \lambda^{-\beta}$ or logarithmic decay $\mathcal{N}(\lambda) \lesssim \log(1/\lambda)$.

Then in [DVPR10, Theorem 4.9] the following bound for the error of a general regularization scheme $S_{\boldsymbol{X}} \boldsymbol{y}$ was proven to hold under certain conditions with probability exceeding $1 - 6\exp(-t)$

$$\begin{aligned}\|f_q - S_{\boldsymbol{X}} \boldsymbol{y}\|_{L_2} &\leq \|f_q - P_{\mathcal{H}} f_q\|_{L_2} + \|P_{\mathcal{H}} f_q - S_{\boldsymbol{X}} \boldsymbol{y}\|_{L_2} \\ &\leq \|f_q - P_{\mathcal{H}} f_q\|_{L_2} + Ct^3\Big(\sqrt{\lambda}\varphi(\lambda) + \sqrt{\frac{\mathcal{N}(\lambda)}{n}}\Big).\end{aligned}$$

We omit the details since we are merely interested in what is possible to achieve with other methods and do not focus on the assumptions.

To find the least squares approximation used in Theorem 8.6 in there, we use $\lambda = 0, \mathcal{N}(\lambda) = N(V) = m$, and the error of the projection polynomially decaying. This results in the same behavior we have proven above

In [DVPR10, Corollaries 5.1 and 5.2] different cases of the above quantities were considered which cover general regularization schemes. There the same rates appear as we have for least squares approximation with the lower bound from [DVPR10, Theorem 5.3] reasoning for their optimality. Note, there the error of the approximation is usually considered to be zero and the bias affects the index function φ.

This suggests, that least squares approximation is on a par with other methods in terms of the approximation rates. However, one always has to tailor the method to the problem at hand/ which is done in least squares approximation by choosing different ansatz functions and weights.

Chapter 9

Cross-validation

In this chapter we deal with different aspects of cross-validation, a parameter choice strategy introduced in [GHW79]. In Chapter 8 we inspected the error behavior of the least squares approximation with respect to the ansatz functions. In particular, in Theorem 8.6 we investigated the optimal choice of polynomial degree m when approximating on the one-dimensional torus $\mathbb{T}$. There we used the exact Fourier coefficients as well as the information about the noise to estimate the L_2-error. However, in practice this is not feasible as it relies on the underlying data distribution. In this chapter we have a look at a method estimating the L_2-error only relying on the data at hand. A basic idea is to split the data into a training set and a validation set for estimating the error. Doing this multiple times we obtain a reasonable estimator for the L_2-error functional known as cross-validation score, which is widely used in learning, cf. [TW96, BS02, MS00, Ros09, LMP20]. A special case is where the partitionings seclude single points, then the training sets become $\boldsymbol{z}_{-i} := (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n)$ and the validation sets $\{z_i\}$. This leads to the so called leave-one-out cross-validation score, which we will investigate.

Definition 9.1. *Let $S_V^{\boldsymbol{X}}\boldsymbol{y}$ be an approximation based on the data samples $\boldsymbol{z} = \{(\boldsymbol{x}^i, y_i), \boldsymbol{x}^i \in D, y_i \in \mathbb{K}, i = 1, \dots, n\}$. Further let $S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i}$ be the same method applied to the samples with the i-th sample omitted and $\beta\colon D \to \mathbb{R}_+$ be a function. The* ***importance weighted cross-validation score*** *is defined via*

$$\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} \beta(\boldsymbol{x}^i)\left|S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i}(\boldsymbol{x}^i) - y_i\right|^2 .$$

Note, that we used a weighted version making cross-validation applicable to the domain adaptation setting discussed in Chapter 8.

Even though the cross-validation score is constructed in a way to estimate the risk, we are interested in theoretical validation thereof. One has propositions about the goodness of the cross-validation score in asymptotic cases, cf. [Li86, GKKW02, Luk06, Gu13], on average, cf. [GHW79, BR08, Bec11], or by restriction of noise, cf. [KN08, KPP18]. In Section 9.1 we discuss what the cross-validation score estimates and give novel concentration inequalities on how well it does that for the least squares approximation.

An immediate drawback of the cross-validation score is the numerical complexity of having to compute the n approximations $S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i}$. However, this can be circumvented in many cases with ideas including Monte Carlo approximations [DG91], matrix decomposition methods [Wei07, SWB08], or Krylow space methods [LdHA10]. In Section 9.2 we present results from [BHP20] tackling this problem for least squares approximation.

9.1 Concentration of the cross-validation score

In this section we investigate the statistical properties of the cross-validation score. We start with its expected value to show what it actually estimates.

Theorem 9.2. *Let $\boldsymbol{z} \in (D \times \mathbb{K})^n$ be a sample distributed according to a source measure ϱ_S, ϱ_T a target measure, and $\beta = \frac{\mathrm{d}\varrho_T}{\mathrm{d}\varrho_S}$ the Radon-Nikodym derivative. Then the importance weighted cross-validation score CV_β from Theorem 9.1 estimates the expected error with respect to the target measure ϱ_T for $n-1$ samples plus the variance σ_q^2 of the learning problem, i.e., for a general learning method $S_V^{\boldsymbol{X}}$ we have*

$$\mathbb{E}_{\boldsymbol{z}}\Big(\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})\Big) = \mathbb{E}_{\boldsymbol{z}_{-1}}\Big(\|S_V^{\boldsymbol{X}_{-1}}\boldsymbol{y}_{-1} - f_q\|^2_{L_2(D,\varrho_T)}\Big) + \sigma_q^2\,.$$

Proof. By the definition of the cross-validation score and linearity we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{z}}\Big(\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})\Big)\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\boldsymbol{z}_{-i}}\bigg(\mathbb{E}_{\boldsymbol{x}^i,y_i}\Big(\beta(\boldsymbol{x}^i)\Big|(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i})(\boldsymbol{x}^i) - y_i\Big|^2\Big)\bigg)\,.
\end{aligned}$$

We simplify the inner expected value as follows

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{x}^i,y_i}\Big(\beta(\boldsymbol{x}^i)\Big|(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i})(\boldsymbol{x}^i) - y_i\Big|^2\Big)\\
&= \mathbb{E}_{\boldsymbol{x}^i,y_i}\Big(\beta(\boldsymbol{x}^i)\Big|(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i})(\boldsymbol{x}^i) - f_q(\boldsymbol{x}^i) + f_q(\boldsymbol{x}^i) - y_i\Big|^2\Big)\\
&= \mathbb{E}_{\boldsymbol{x}^i}\Big(\beta(\boldsymbol{x}^i)\Big|(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i})(\boldsymbol{x}^i) - f_q(\boldsymbol{x}^i)\Big|^2\Big) + \mathbb{E}_{\boldsymbol{x}^i}\Big(\beta(\boldsymbol{x}^i)\sigma^2(\boldsymbol{x})\Big)\\
&\quad + 2\mathbb{E}_{\boldsymbol{x}^i}\Big(\beta(\boldsymbol{x}^i)\Big|(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i})(\boldsymbol{x}^i) - f_q(\boldsymbol{x}^i)\Big|\underbrace{|f_q(\boldsymbol{x}^i) - \mathbb{E}_{y_i}(y_i)|}_{=0}\Big)\\
&= \int_D |(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i})(\boldsymbol{x}) - f_q(\boldsymbol{x})|^2\,\varrho_T(\boldsymbol{x}) + \sigma_q^2\,.
\end{aligned}$$

Thus, by the independence of the data samples, we obtain the assertion:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{z}}\Big(\operatorname{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})\Big) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\boldsymbol{z}_{-i}}\Big(\|(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i}) - f_q\|^2_{L_2(D,\varrho_T)} + \sigma_q^2\Big)\\ &= \mathbb{E}_{\boldsymbol{z}_{-1}}\Big(\|S_V^{\boldsymbol{X}_{-1}}\boldsymbol{y}_{-1} - f_q\|^2_{L_2(D,\varrho_T)}\Big) + \sigma_q^2\,. \qquad \blacksquare\end{aligned}$$

It was shown in [BHT23] that the cross-validation score estimates the expected risk on the unseen samples rather than the error of the model at hand by investigating the respective variances.

In expectation Theorem 9.2 implies, that instead of minimizing the L_2-error over a hypothesis space, which requires the underlying distribution, we may minimize the cross-validation score and obtain a good hypothesis.

Remark 9.3. *There are scenarios, where we do not know β exactly but rather an approximation $\tilde\beta$. Assuming we have a bound on the error $\|\beta - \tilde\beta\|_\infty \le \varepsilon$ like in [GMM$^+$22] and our approximations $S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i}$ and sample values y_i are bounded by some constant $K > 0$, we have*

$$\begin{aligned}&\Big|\operatorname{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \operatorname{CV}_{\tilde\beta}(S_V^{\boldsymbol{X}}\boldsymbol{y})\Big|\\ &\le \frac{1}{n}\sum_{i=1}^n |\beta(\boldsymbol{x}^i) - \tilde\beta(\boldsymbol{x}^i)|\Big|(S_V^{\boldsymbol{X}_{-i}}\boldsymbol{y}_{-i})(\boldsymbol{x}^i) - y_i\Big|^2\\ &\le \frac{4K^2}{n}\sum_{i=1}^n |\beta(\boldsymbol{x}^i) - \tilde\beta(\boldsymbol{x}^i)|\\ &\le 4K^2\varepsilon\,.\end{aligned}$$

Thus, an error in β only enters linearly in the cross-validation score.

To show how good the cross-validation score estimates the L_2-error, we will state concentration results. In [BE02] this was done using a robustness concept of an algorithm in combination with McDiarmid's concentration inequality, which is not directly applicable to the least squares approximation. We combine this with concepts of [BH22] using an extension of McDiarmid's concentration result to overcome this relying only on robustness with high probability instead of everywhere.

To work towards the main result we need several lemmas starting with an updated L_2-MZ inequality for omitting points.

Lemma 9.4. *Assume that points $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq D$ and weights $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n) \in [0,\infty)^{n\times n}$ fulfill an L_2-MZ inequality* (6.1) *with* *constants A and B for a function space V. Then omitting $k \in \mathbb{N}$ points, we have an L_2-MZ inequality with constants $A - k\|\boldsymbol{W}\|_\infty N(V)$ and B, where $N(V)$ is the Christoffel function from* (3.14).

Proof. Subtracting the j-th term in the L_2-MZ inequality, we obtain

$$\Big(A - \omega_j \frac{|f(\boldsymbol{x}^j)|^2}{\|f\|_{L_2}^2}\Big)\|f\|_{L_2}^2 \le \sum_{i\neq j} \omega_i |f(\boldsymbol{x}^i)|\,.$$

And by (3.15)

$$\omega_j \frac{|f(\boldsymbol{x}^j)|^2}{\|f\|_{L_2}^2} \le \|\boldsymbol{W}\|_\infty \sup_{g\in V} \frac{|g(\boldsymbol{x}^j)|^2}{\|g\|_{L_2}^2}\|f\|_{L_2}^2 \le \|\boldsymbol{W}\|_\infty N(V)\,.$$

Iterating this k times, we obtain the assertion. ■

The next lemma states a coarse bound on the possible deviation of the least squares approximation.

Lemma 9.5. *Let $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$ be points and $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n)$ weights fulfilling a lower L_2-MZ inequality* (6.1) *with constant A for a function space V. Further, let $\boldsymbol{y} = (y_1, \dots, y_n)^\mathsf{T} \in \mathbb{K}^n$ such that $|y_i| \le K$ for some $K > 0$.*

Then the least squares approximation $S_V^{\boldsymbol{X}}\boldsymbol{y}$ from Section 2.2 is bounded

$$\|S_V^{\boldsymbol{X}}\boldsymbol{y}\|_\infty \le \sqrt{\frac{N(V)\sum_{i=1}^n \omega_i}{A}}K\,.$$

Proof. Let $\hat{\boldsymbol{g}}$ be the coefficients of $S_V^{\boldsymbol{X}}\boldsymbol{y}$. We first use (8.7) to reduce the $\|\cdot\|_\infty$ to the $\|\cdot\|_{L_2}$ norm: $\|S_V^{\boldsymbol{X}}\boldsymbol{y}\|_\infty \le \sqrt{N(V)}\|\hat{\boldsymbol{g}}\|_2$. By Theorem 6.2, we obtain

$$\|\hat{\boldsymbol{g}}\|_2^2 \le \|(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}^{1/2}\|_{2\to 2}^2 \|\boldsymbol{W}^{1/2}\boldsymbol{y}\|_2^2 \le \frac{K^2\sum_{i=1}^n \omega_i}{A}\,. \quad ■$$

As omitting single points is at the core of cross-validation, we have to know the behavior of the least squares approximation in this case. Next, we show different identities which will serve as central tools for the later proof.

Lemma 9.6. *Let $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$ be points and $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n)$ weights, and V a function space. Let $S_V^{\boldsymbol{X}}\boldsymbol{y}$ be the least squares approximation defined in Section 2.2 with $\hat{\boldsymbol{g}}$ its Fourier coefficients and $S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j}$, $\hat{\boldsymbol{g}}_{-j}$ the* *same for the j-th sample omitted. Further, let $h_{j,j}$ be the j-th diagonal entry of matrix $\boldsymbol{L}(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}$ with $\boldsymbol{L}$ and $\boldsymbol{W}$ as in* (2.1). *Then*

(i) $$(S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j})(\boldsymbol{x}^j) - y_j = \frac{(S_V^{\boldsymbol{X}}\boldsymbol{y})(\boldsymbol{x}^j) - y_j}{1-h_{j,j}}\,,$$

(ii) $$\hat{\boldsymbol{g}}_{-j} - \hat{\boldsymbol{g}} = \omega_j(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}_j^*\Big((S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j})(\boldsymbol{x}^j) - y_j\Big)\,.$$

(iii) If we further have a lower L_2-MZ inequality for the function space V with constants A and $|y_i| \leq K$ for some $K > 0$, then

$$\|S_V^{\boldsymbol{X}}\boldsymbol{y} - S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j}\|_{L_2}^2 \leq \frac{4K^2\|\boldsymbol{W}\|_1\|\boldsymbol{W}\|_\infty^2 N^2(V)}{(A - \|\boldsymbol{W}\|_\infty N(V))^3}\,.$$

Proof. Let $\boldsymbol{L}_{-j}$, $\boldsymbol{W}_{-j}$, and $\boldsymbol{y}_{-j}$ be the respective quantities with the j-th column, column and row, and entry omitted, respectively.

Assertions (i) and (ii). The coefficients of the approximations $S_V^{\boldsymbol{X}}\boldsymbol{y}$ and $S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j}$ are given by

$$\begin{aligned}\hat{\boldsymbol{g}} &= (\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{y},\\ \hat{\boldsymbol{g}}_{-j} &= (\boldsymbol{L}_{-j}^*\boldsymbol{W}_{-j}\boldsymbol{L}_{-j})^{-1}\boldsymbol{L}_{-j}^*\boldsymbol{W}_{-j}\boldsymbol{y}_{-j},\end{aligned}$$

respectively. Using the Sherman-Morrison formula for the inverse of rank-1 updates, see e.g. [Har97], we obtain

$$\begin{aligned}\hat{\boldsymbol{g}}_{-j} - \hat{\boldsymbol{g}} &= (\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L} - \boldsymbol{L}_j^*\omega_j\boldsymbol{L}_j)^{-1}(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{y} - \boldsymbol{L}_j^*\omega_j y_j) - \hat{\boldsymbol{g}}\\ &= \Big((\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1} + \frac{(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}_j\omega_j\boldsymbol{L}_j^*(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}}{1-\omega_j\boldsymbol{L}_j^*(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}_j}\Big)\cdot\\ &\quad(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{y} - \boldsymbol{L}_j^*\omega_j y_j) - \hat{\boldsymbol{g}}\\ &= (\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}_j\omega_j\frac{(S_V^{\boldsymbol{X}}\boldsymbol{y})(\boldsymbol{x}^j) - y_j}{1-h_{j,j}}\,. \end{aligned}\tag{9.1}$$

Thus,

$$\begin{aligned}(S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j})(\boldsymbol{x}^j) - (S_V^{\boldsymbol{X}}\boldsymbol{y})(\boldsymbol{x}^j) &= \boldsymbol{L}_j(\hat{\boldsymbol{g}}_{-j} - \hat{\boldsymbol{g}})\\ &= h_{j,j}\frac{(S_V^{\boldsymbol{X}}\boldsymbol{y})(\boldsymbol{x}^j) - y_j}{1-h_{j,j}}\,,\end{aligned}$$

which shows (i). Plugging (i) into (9.1) we obtain (ii).

Assertion (iii). We use the lower L_2-MZ inequality with Theorem 6.2, Parseval's identity, and (ii) to obtain

$$\begin{aligned}\Big\|S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j}-S_V^{\boldsymbol{X}}\boldsymbol{y}\Big\|_{L_2}^2 &= \Big\|\omega_j(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}_j^*((S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j})(\boldsymbol{x}^j)-y_j)\Big\|_2^2\\ &\le \frac{\|\boldsymbol{W}\|_\infty^2 N(V)}{A^2}\Big|(S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j})(\boldsymbol{x}^j)-y_j\Big|^2 .\end{aligned}$$

Using Lemmata 9.4 and 9.5 we obtain

$$\|S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j}-S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 \le 2K^2\frac{\|\boldsymbol{W}\|_\infty^2 N(V)}{A^2}\Big(\frac{N(V)\|\boldsymbol{W}\|_1}{A-\|\boldsymbol{W}\|_\infty N(V)}+1\Big).$$

We replace the added one in the bracketed term by a factor of two using

$$A-\|\boldsymbol{W}\|_\infty N(V)\le A\le \frac{1}{m}\sum_{i=1}^n\omega_i\Big|\sum_{k=0}^{m-1}\eta_k(\boldsymbol{x}^i)\Big|^2\le\|\boldsymbol{W}\|_1 N(V)$$

to finally obtain

$$\|S_V^{\boldsymbol{X}_{-j}}\boldsymbol{y}_{-j}-S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 \le 4K^2\frac{\|\boldsymbol{W}\|_\infty^2 N(V)}{A^2}\frac{N(V)\|\boldsymbol{W}\|_1}{A-\|\boldsymbol{W}\|_\infty N(V)} . \qquad \blacksquare$$

With these auxiliary lemmas accomplished, we work towards applying McDiarmid's concentration inequality Theorem 4.7. For that we will use the following assumptions

Assumption 9.7. *Let ϱ_S, ϱ_T be measures, $\beta=\frac{\mathrm{d}\varrho_T}{\mathrm{d}\varrho_S}$ the Radon-Nikodym derivative, and V be an $m-1$-dimensional function space with $m\ge 2$ and an orthonormal basis $\eta_1,\dots,\eta_{m-1}$ in $L_2(D,\varrho_T)$ satisfying*

$$20\|\beta\|_\infty N(V)\le\frac{n}{\log n} .$$

Further, let $\boldsymbol{y}=(y_1,\dots,y_n)^\mathsf{T}\in\mathbb{K}^n$ with $|y_i|,|\tilde{y}_i|\le K$ for some $K>0$ and we define the set of points where the lower L_2-MZ inequality holds

$$\Xi := \Big\{\{\boldsymbol{x}^1,\dots,\boldsymbol{x}^n\}\in D^n : \frac{n}{2}\|f\|_{L_2(D,\varrho_T)}^2\le\sum_{i=1}^n\beta(\boldsymbol{x}^i)|f(\boldsymbol{x}^i)|^2\ \forall f\in V\Big\}.$$

Under these assumptions, we will show the $\boldsymbol{c}$-boundedness of the cross-validation score and the L_2-error and, thereafter, their concentration by using McDiarmids concentration inequality. Note, by Theorem 6.4 we have that a random draw of point with respect to ϱ_S is element of Ξ with probability exceeding $1-1/n$.

Lemma 9.8. *Let Theorem 9.7 hold. Then for all $\boldsymbol{x} \in D$ and all $i = 1, \dots, n$ the weighted least squares approximation $(S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i})(\boldsymbol{x})$ defined in Section 2.2 with $\omega_i = \beta(\boldsymbol{x}^i)$ is $\boldsymbol{c}$-bounded in Ξ with*

$$\boldsymbol{c} = \frac{13KN^{3/2}(V)\|\beta\|_\infty^{3/2}}{n}\mathbb{1}\,.$$

Proof. Let $\boldsymbol{X} \in \Xi$ and $\boldsymbol{y} \in \mathbb{K}^n$ with $|y_i| \le K$. To show $\boldsymbol{c}$-boundedness we assume $\tilde{\boldsymbol{X}} \in \Xi$, $\tilde{\boldsymbol{y}} \in \mathbb{K}^n$, $|\tilde{y}_i| \le K$, be copies of $\boldsymbol{X}$, $\boldsymbol{y}$ differing in the first component without loss of generality. By using the usual $\|\cdot\|_\infty$ to $\|\cdot\|_{L_2}$ estimate (8.7) and triangle inequality, we have

$$\begin{aligned}
&\left\| S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i} - S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}}_{-i} \right\|_\infty \\
&\le \sqrt{N(V)} \left\| S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i} - S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}}_{-i} \right\|_{L_2(D,\varrho_T)} \\
&\le \sqrt{N(V)} \Big(\left\| S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i} - S_V^{\boldsymbol{X}_{-i-1}} \boldsymbol{y}_{-i-1} \right\|_{L_2(D,\varrho_T)} \\
&\quad + \left\| S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}}_{-i} - S_V^{\tilde{\boldsymbol{X}}_{-i-1}} \tilde{\boldsymbol{y}}_{-i-1} \right\|_{L_2(D,\varrho_T)} \Big)\,.
\end{aligned}$$

Applying Theorem 9.6 (iii) with $A = n/2$ the construction of Ξ gives

$$\left\| S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y} - S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}} \right\|_\infty \le \frac{2^{7/2}K}{n} \Big(\frac{nN(V)\|\beta\|_\infty}{n - 4\|\beta\|_\infty N(V)} \Big)^{3/2}\,.$$

Since $N(V) \ge m-1 \ge 1$ and $\|\beta\|_\infty \ge 1$, the assumptions on n gives $n \ge 20$ and

$$2^{7/2} \Big(\frac{n}{n - 4\|\beta\|_\infty N(V)} \Big)^{3/2} \le 2^{7/2} \Big(\frac{1}{1 - 1/(5\log n)} \Big)^{3/2} \le 13\,.$$

Thus, we have bounded the effect of changing one point:

$$\left\| S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y} - S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}} \right\|_\infty \le \frac{13KN^{3/2}(V)\|\beta\|_\infty^{3/2}}{n}\,. \qquad ■$$

With that, we show the $\boldsymbol{c}$-boundedness of the cross-validation score and the L_2-error. We make a fine adjustment to our approximation method by truncating the absolute value of our least squares approximation by some $K > 0$ via the operator $T_K \colon \mathbb{K}^D \to \mathbb{K}$,

$$(T_K f)(\boldsymbol{x}) = \arg(f(\boldsymbol{x})) \max\{|f(\boldsymbol{x})|, K\}\,. \tag{9.2}$$

This is easy to implement and improves the bound.

Lemma 9.9. *Let Theorem 9.7 hold. Then* $\mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y})$*, the cross-validation score of the weighted least squares approximation defined in Section 2.2 with* $\omega_i = \beta(\boldsymbol{x}^i)$*, is* $\boldsymbol{c}$*-bounded for points* $\boldsymbol{X} \in \Xi$ *with*

$$\boldsymbol{c} = \frac{60K^2\|\beta\|_\infty^{5/2}}{n} N^{3/2}(V)\mathbb{1}\,.$$

Further, the L_2*-error* $\|f_q - T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1}\|^2_{L_2(D,\varrho_T)}$ *is* $\boldsymbol{c}$*-bounded for points* $\boldsymbol{X} \in \Xi$ *with*

$$\boldsymbol{c} = \frac{52K^2\|\beta\|_\infty^{3/2}}{n} N^{3/2}(V)\mathbb{1}\,.$$

Proof. **Step 1.** We first show the $\boldsymbol{c}$-boundedness of the cross-validation score in Ξ. We assume $\tilde{\boldsymbol{X}} \in \Xi$, $\tilde{\boldsymbol{y}} \in \mathbb{K}^n$, $|\tilde{y}_i| \le K$, be copies of $\boldsymbol{X}$, $\boldsymbol{y}$ differing in the first component without loss of generality. We have

$$\begin{aligned}
&\Big|\mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) - \mathrm{CV}_\beta(T_K S_V^{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{y}})\Big| \\
&\quad\le \frac{1}{n}\beta(\boldsymbol{x}^1)|(T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1})(\boldsymbol{x}^1) - y_1|^2 \\
&\qquad + \frac{1}{n}\beta(\tilde{\boldsymbol{x}}^1)|(T_K S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1})(\tilde{\boldsymbol{x}}^1) - \tilde{y}_1|^2 \\
&\qquad + \frac{1}{n}\sum_{i=2}^n \beta(\boldsymbol{x}^i)\Big(|(T_K S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i})(\boldsymbol{x}^i) - y_i|^2 \\
&\qquad - |(T_K S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}}_{-i})(\boldsymbol{x}^i) - y_i|^2\Big)\,.
\end{aligned}$$

Using the third binomial formula, we obtain

$$\begin{aligned}
&\Big|\mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) - \mathrm{CV}_\beta(T_K S_V^{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{y}})\Big| \\
&\quad\le \frac{8K\|\beta\|_\infty}{n} + \frac{1}{n}\sum_{i=2}^n \beta(\boldsymbol{x}^i)\Big|(T_K S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i})(\boldsymbol{x}^i) - (T_K S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}}_{-i})(\boldsymbol{x}^i)\Big| \cdot \\
&\qquad \Big|(T_K S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}}_{-i})(\boldsymbol{x}^i) + (T_K S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i})(\boldsymbol{x}^i) - 2y_i\Big| \\
&\quad\le \frac{8K\|\beta\|_\infty}{n} + \frac{4K\|\beta\|_\infty}{n}\sum_{i=2}^n \Big|(S_V^{\boldsymbol{X}_{-i}} \boldsymbol{y}_{-i})(\boldsymbol{x}^i) - (S_V^{\tilde{\boldsymbol{X}}_{-i}} \tilde{\boldsymbol{y}}_{-i})(\boldsymbol{x}^i)\Big|\,.
\end{aligned}$$

By Theorem 9.8 we obtain the $\boldsymbol{c}$-boundedness of the cross-validation score

$$\begin{aligned}
&\left|\mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) - \mathrm{CV}_\beta(T_K S_V^{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{y}})\right| \\
&\quad \le \frac{8K\|\beta\|_\infty}{n} + \frac{52K^2\|\beta\|_\infty^{5/2}}{n} N^{3/2}(V) \\
&\quad \le \frac{60K^2\|\beta\|_\infty^{5/2}}{n} N^{3/2}(V)\,.
\end{aligned}$$

Step 2. We now show the $\boldsymbol{c}$-boundedness of the L_2-error in Ξ. We have

$$\begin{aligned}
&\|f_q - T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1}\|_{L_2(D,\varrho_T)} - \|f_q - T_K S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1}\|_{L_2(D,\varrho_T)} \\
&= \int_D \left|f_q(\boldsymbol{x}) - (T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1})(\boldsymbol{x})\right|^2 \\
&\quad - \left|f_q(\boldsymbol{x}) - (T_K S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1})(\boldsymbol{x})\right|^2 \,\mathrm{d}\varrho_T(\boldsymbol{x})\,.
\end{aligned}$$

Using the third binomial formula, we obtain

$$\begin{aligned}
&\|f_q - T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1}\|_{L_2(D,\varrho_T)} - \|f_q - T_K S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1}\|_{L_2(D,\varrho_T)} \\
&= \int_D \left|2f_q(\boldsymbol{x}) - (T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1})(\boldsymbol{x}) - (T_K S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1})(\boldsymbol{x})\right| \cdot \\
&\quad \left|(T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1})(\boldsymbol{x}) - (T_K S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1})(\boldsymbol{x})\right| \mathrm{d}\varrho_T(\boldsymbol{x}) \\
&\le 4K \int_D \left|(S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1})(\boldsymbol{x}) - (S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1})(\boldsymbol{x})\right| \mathrm{d}\varrho_T(\boldsymbol{x})\,.
\end{aligned}$$

By Theorem 9.8 we obtain the $\boldsymbol{c}$-boundedness of the L_2-error

$$\begin{aligned}
&\|f_q - T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y}_{-1}\|_{L_2(D,\varrho_T)} - \|f_q - T_K S_V^{\tilde{\boldsymbol{X}}_{-1}} \tilde{\boldsymbol{y}}_{-1}\|_{L_2(D,\varrho_T)} \\
&\quad \le \frac{52K^2\|\beta\|_\infty^{3/2}}{n} N^{3/2}(V)\,. \qquad \blacksquare
\end{aligned}$$

Now we state the central theorem of this section showing the concentration of the cross-validation score for random points.

Theorem 9.10. *Let $n \in \mathbb{N}$, $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n$ be points drawn according to a probability measure $\mathrm{d}\varrho_S = 1/\beta\,\mathrm{d}\varrho_T$. Let further, V be an $m-1$-dimensional function space with $m \ge 2$ and an orthonormal basis $\eta_1, \ldots, \eta_{m-1}$ in L_2 with m satisfying*

$$20\|\beta\|_\infty N(V) \le \frac{n}{\log n}\,,$$

$t \geq 0$, and $\boldsymbol{y} \in \mathbb{K}^n$ with $|y_i| \leq K$.

Then, for $T_K S_V^{\boldsymbol{X}} \boldsymbol{y}$ the truncated weighted least squares approximation defined in Section 2.2 and (9.2) *with $\omega_i = \beta(\boldsymbol{x}^i)$, we have for the cross-validation score*

$$\Big| \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y})) - \mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big| \\ \leq 64 K^2 \|\beta\|_\infty^{5/2} \Big(\sqrt{\frac{t}{2}} + 1 \Big) \frac{N^{3/2}(V)}{\sqrt{n}}$$

with probability exceeding $1 - 2/\sqrt{n} - 2\exp(-t)$.

Further, we have for the L_2-error with probability exceeding $1 - 2/\sqrt{n} - 2\exp(-t)$

$$\Big| \mathbb{E}_{\boldsymbol{Z}}(\|T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y} - f_q\|_{L_2(D,\varrho_T)}^2) - \|T_K S_V^{\boldsymbol{X}_{-1}} \boldsymbol{y} - f_q\|_{L_2(D,\varrho_T)} \Big| \\ \leq 56 K^2 \|\beta\|_\infty^{3/2} \Big(\sqrt{\frac{t}{2}} + 1 \Big) \frac{N^{3/2}(V)}{\sqrt{n}} \,.$$

Proof. Using triangle inequality, we obtain

$$\Big| \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y})) - \mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big| \\ \leq \Big| \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y})) - \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) | \boldsymbol{Z} \in \Xi) \Big| \\ + \Big| \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) | \boldsymbol{Z} \in \Xi) - \mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big| \,.$$

By Theorem 6.4 with $t = \log(n)$, the with respect to $\mathrm{d}\varrho_S$ drawn points are in Ξ with probability exceeding $1 - \gamma = 1 - 1/n$. Thus, The first summand is bounded by the fail probability of Γ from Step 1 and $\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \leq 4K$:

$$\Big| \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y})) - \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) | \boldsymbol{Z} \in \Xi) \Big| \\ = \mathbb{P}_{\boldsymbol{X}}(\boldsymbol{X} \notin \Xi) \max_{\boldsymbol{X}, \boldsymbol{y}} \Big| \mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big| \leq \frac{4K}{n} \,.$$

Finally applying McDiarmid's concentration inequality together with the $\boldsymbol{c}$-boundedness from Theorem 9.9 in order to bound the second summand by

$$\Big| \mathbb{E}_{\boldsymbol{Z}}(\mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) | \boldsymbol{Z} \in \Xi) - \mathrm{CV}(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big| \\ \leq \Big(\sqrt{\frac{tn}{2}} + \sqrt{n} \Big) \frac{60 K^2 \|\beta\|_\infty^{5/2}}{n} N^{3/2}(V)$$

with probability exceeding $1 - 2/\sqrt{n} - 2\exp(-t)$ giving the assertion.

The concentration of the L_2-error follows analogous. ■

To get a feeling for Theorem 9.10, we will put it into perspective by formulating a guarantee for using cross-validation as an a posteriori parameter choice strategy.

Corollary 9.11. *Under the assumptions of Theorem 9.10, we have with probability exceeding* $1 - 4/\sqrt{n} - 4\exp(-t)$

$$
\begin{aligned}
\|f_q - T_K S_V^{\boldsymbol{X}\,-1}\|_{L_2(D,\varrho_T)} \leq{}& \mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) - \sigma_q^2 \\
&+ 120K^2 \|\beta\|_\infty^{5/2} \Big(\sqrt{\frac{t}{2}} + 1 \Big) \frac{N^{3/2}(V)}{\sqrt{n}} \,.
\end{aligned}
$$

Proof. By Theorem 9.2, we have

$$
\begin{aligned}
&\|f_q - T_K S_V^{\boldsymbol{X}\,-1}\|_{L_2(D,\varrho_T)} = \|f_q - T_K S_V^{\boldsymbol{X}\,-1}\|_{L_2(D,\varrho_T)} \\
&\quad - \mathbb{E}_{\boldsymbol{Z}_{-1}}\Big(\|f_q - T_K S_V^{\boldsymbol{X}\,-1} \boldsymbol{y}_{-1}\|_{L_2(D,\varrho_T)} \Big) + \mathbb{E}_{\boldsymbol{Z}}\Big(\mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big) - \sigma_q^2 \\
&\quad - \mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) + \mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \\
&\leq \Big| \|f_q - T_K S_V^{\boldsymbol{X}\,-1}\|_{L_2(D,\varrho_T)} - \mathbb{E}_{\boldsymbol{Z}_{-1}}\Big(\|f_q - T_K S_V^{\boldsymbol{X}\,-1} \boldsymbol{y}_{-1}\|_{L_2(D,\varrho_T)} \Big) \Big| \\
&\quad + \Big| \mathbb{E}_{\boldsymbol{Z}}\Big(\mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big) - \mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) \Big| \\
&\quad + \mathrm{CV}_\beta(T_K S_V^{\boldsymbol{X}} \boldsymbol{y}) - \sigma_q^2 \,.
\end{aligned}
$$

Applying Theorem 9.10, we obtain the assertion. ■

With Theorem 9.11, we have an estimator of the unknown L_2-error by computing the cross-validation score, which only relies on the given data. To discuss the involved rates in more detail we resume to Theorem 8.6.

Example 9.12. *We return to the one-dimensional learning problem on the torus* $\mathbb{T}$ *with the exponential functions from Theorem 8.6. There we established a theoretical guarantee of the* L_2*-error an deduced an optimal polynomial degree* $m^\star$*, which we do not know in practice.*

With Theorem 9.11 we have a bound for the a posteriori parameter estimator cross-validation. A bottleneck is the third summand with $N^{3/2}(V)/\sqrt{n}$ *determining its asymptotic behavior. In this setting we have* $N(V) = m$*. In the following table we compare the* L_2*-error for optimally chosen polynomial degree* $m^\star$ *and the bound for the cross-validation score for different* *smoothness of the regression function f_q. Note, we are only interested in the dependence in the number of points n and ignore constants.*

smoothness	$m^\star$	$\|f_q - S_V^{\boldsymbol{X}}\|_{L_2}^2$	*cross-validation bound*
$f_q \in H^s(\mathbb{T})$	$n^{\frac{1}{2s+1}}$	$n^{-\frac{2s}{2s+1}}$	$n^{\frac{1-s}{2s+1}}$
$f_q \in C^\infty(\mathbb{T})$	$\log(n)$	$\frac{\log n}{n}$	$\frac{\log^{3/2} n}{\sqrt{n}}$

Comparing the last two columns shows that the L_2-error with the optimal polynomial degree has approximately double the order of convergence compared to the cross-validation bound. Thus, we loose theoretically half the rate of convergence. In the numerical experiments in Section 9.3 the loss of half the rate is not observed indicating room for improvement.

Concluding this section, we have shown concentration inequalities for the cross-validation score in least squares approximation. We weakened the robustness concept from [BE02] to only hold with high probability and used an extension of McDiarmid's inequality suitable in this case. As many algorithms involve random components, this concept may be used in other scenarios as well. For our case of least squares approximation this allowed to obtain guarantees for the goodness of cross-validation as an a posteriori parameter choice strategy. We will validate the theory with numerical experiments in Section 9.3.

9.2 Fast cross-validation for least squares approximation

To compute the cross-validation score CV_β as defined in Theorem 9.1 we need as many least squares approximations as there are points in $\boldsymbol{X}$. In this section we show how this can be reduced to only the cost of computing one least squares approximation. Parts of the results are already published in [BHP20].

We start with a result, which shifts the computational heavy part from setting up the approximations to the points in question, cf. [GHW79].

Theorem 9.13. *The cross-validation score $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ of the least squares approximation $S_V^{\boldsymbol{X}}\boldsymbol{y}$ defined in Section 2.2 can be computed via*

$$\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n}\beta(\boldsymbol{x}^i)\Big|\frac{(S_V^{\boldsymbol{X}}\boldsymbol{y})(\boldsymbol{x}^i) - y_i}{1-h_{i,i}}\Big|^2,$$

where $h_{i,i}$ are the diagonal elements of the matrix of the least squares approximation $\boldsymbol{L}(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}$.

Proof. The statement follows by plugging Theorem 9.6 (i) in the definition of the cross-validation score. ■

With that we only need the residual of the least squares approximation using all points $\boldsymbol{X}$ and the diagonal entries $h_{i,i}$. For the particular example of trigonometric polynomials with a point- and frequency grid an analytic formula was shown in [TW96]. This is generalizable. Given a tight L_2-MZ inequality, the diagonal entries $h_{i,i}$ can be computed explicitly as the following theorem shows.

Theorem 9.14. *Let $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\}$ and $\boldsymbol{W} = \operatorname{diag}(\omega_1, \dots, \omega_n)$ form a tight L_2-MZ inequality for the approximation space $V = \operatorname{span}\{\eta_k\}_{k=1}^{m-1}$ with constant A, cf.* (6.1). *Then the diagonal entries from Theorem 9.13 evaluate to*

$$h_{i,i} = \frac{\omega_i}{A} N(V, \boldsymbol{x}^i)\,.$$

Proof. From Theorem 6.2 we know $\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L} = A\boldsymbol{I}$. Thus,

$$h_{i,i} = (\boldsymbol{L}(\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W})_{i,i} = \frac{\omega_i}{A}(\boldsymbol{L}\boldsymbol{L}^*)_{i,i}\,,$$

which evaluates to the assertion by the definition of $\boldsymbol{L}$ and the Christoffel function (3.14). ■

Therefore, the computation of the cross-validation score boils down to the computation the least squares approximation for all points $\boldsymbol{X}$ and the evaluation of the Christoffel function $N(V, \cdot)$ (3.14) at the given points.

Example 9.15. *In particular cases the Christoffel function $N(V, \cdot)$ is fast to compute:*

(i) On $\mathbb{T}^d$ the Fourier monomials $\eta_{\boldsymbol{k}} = \exp(2\pi \mathrm{i}\langle \boldsymbol{k}, \cdot\rangle)$ have absolute value one everywhere. Thus, for $V = \operatorname{span}\{\eta_{\boldsymbol{k}_1}, \dots, \eta_{\boldsymbol{k}_{m-1}}\}$ we obtain $N(V, \cdot) \equiv m - 1$. For the equispaced grid this was already observed in [TW96].

(ii) On the two-dimensional sphere $\mathbb{S}^2 = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\|_2 = 2\}$, there is a basis of spherical harmonics $\{Y_{k,l}\}_{k=0,\dots,m,l=-k,\dots,k}$, cf. [FGS98, AH12, Mic13, DX13]. Using the addition theorem from [Mic13, Theorem 5.11], we have $\sum_{l=-k}^{k} |Y_{k,l}|^2$ being constant for all $k \in \mathbb{N}$. Thus, the Christoffel function is constant in this case.

This works analogous for higher-dimensional spheres and the Wigner-D functions $D_k^{ll'}$, $l, l' = -k, \dots, k$ on the rotation group $SO(3)$, cf.

Definition [VMK88, Section 4.1, Equation (5)] using the corresponding addition theorem [VMK88, Section 4.7, Equation (4)].

(iii) *Even when the Christoffel function is not constant, the fast computation is possible. Taking the Chebyshev polynomials in one dimension from Section 3.5.3, we have for* $V = \operatorname{span}\{T_0, \dots, T_{m-1}\}$

$$N(V, x) = \frac{1}{\pi} + \frac{2}{\pi}\sum_{k=1}^{m-1} \cos^2(k \arccos(x)) \,.$$

Using the power reduction formula $\cos(2\alpha) = (\cos(2\alpha) + 1)/2$, *we obtain*

$$N(V, x) = \frac{m}{\pi} + \frac{1}{\pi}\sum_{k=1}^{m} \cos(2k \arccos(x)) \,,$$

which can be computed with a discrete cosine transform of type I (DCT-I) in $\mathcal{O}(m \log m)$ *for Chebyshev points* $x^i = \cos(\pi(2i-1)/(2m))$ *for* $i = 1, \dots, m$, *cf. [PPST18, Section 6.3] or using the nonequispaced cosine transform for arbitrary points [FP05].*

Analogously, this works for the half-period cosine from Theorem 3.26.

In Theorem 9.15 we have seen that there are cases where the Christoffel function $N(V, \cdot)$ simplifies yielding barely any computational cost. Even if we have to compute the Christoffel function naively, we end up with $\mathcal{O}(m \cdot n)$ arithmetic operations. This is the same as the naive matrix-vector product used in the least squares approximation and, thus, stems no bottleneck when fast algorithms are not available.

Next, we have a look what happens when there is no exact L_2-MZ inequality and Theorem 9.14 is not applicable. E.g. random points used in the statistical learning setting from Chapter 8 do not fall into this category. We have often seen, that the loss of the tightness of an L_2-MZ inequality is not problematic. Motivated by that, we propose an alternative for the exact cross-validation score in order to preserve the possibility of the fast computation.

Definition 9.16. *We introduce the* ***approximated cross-validation score*** *of the least squares approximation* $S_V^{\boldsymbol{X}}\boldsymbol{y}$ *via*

$$\operatorname{FCV}(S_V^{\boldsymbol{X}}\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} \beta(\boldsymbol{x}^i)\Big|\frac{(S_V^{\boldsymbol{X}}\boldsymbol{y}(\boldsymbol{x}^i) - y_i}{1 - \tilde{h}_{i,i}}\Big|^2$$

with $\tilde{h}_{i,i} = \min\{\frac{A+B}{2AB}\omega_i N(V, \boldsymbol{x}^i), 1\}$.

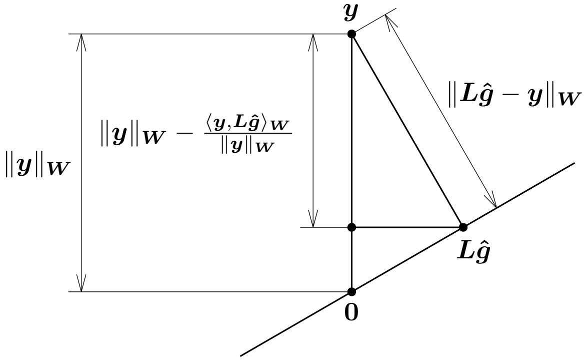

$\boldsymbol{y}$
$\|\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y}\|_{\boldsymbol{W}}$
$\|\boldsymbol{y}\|_{\boldsymbol{W}} - \frac{\langle \boldsymbol{y}, \boldsymbol{L}\hat{\boldsymbol{g}}\rangle_{\boldsymbol{W}}}{\|\boldsymbol{y}\|_{\boldsymbol{W}}}$
$\|\boldsymbol{y}\|_{\boldsymbol{W}}$
$\boldsymbol{L}\hat{\boldsymbol{g}}$
$\boldsymbol{0}$

Figure 9.1: Geometric proof of Theorem 9.17.

The boundedness by one is motivated by the behavior of the exact diagonal entries $h_{i,i}$ covered by the next lemma.

Lemma 9.17. *Let* $\hat{\boldsymbol{g}} = \arg\min_{\hat{\boldsymbol{a}} \in \mathbb{K}^m} \|\boldsymbol{L}\hat{\boldsymbol{a}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2$*. Then*

$$\langle \boldsymbol{y}, \boldsymbol{L}\hat{\boldsymbol{g}}\rangle_{\boldsymbol{W}} = \|\boldsymbol{y}\|_{\boldsymbol{W}}^2 - \|\boldsymbol{L}\hat{\boldsymbol{g}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 \in [0, \|\boldsymbol{y}\|_{\boldsymbol{W}}^2] \,.$$

In particular, for our case of interest $\hat{\boldsymbol{h}}_i = \arg\min_{\hat{\boldsymbol{a}} \in \mathbb{K}^m} \|\boldsymbol{L}\hat{\boldsymbol{a}} - \boldsymbol{e}_i\|_{\boldsymbol{W}}^2 = (\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{e}_i$*, we obtain*

$$h_{i,i} = 1 - \frac{\|\boldsymbol{L}\hat{\boldsymbol{h}}_i - \boldsymbol{e}_i\|_{\boldsymbol{W}}^2}{\omega_i} \in [0, 1] \,.$$

Proof. Since $\boldsymbol{L}\hat{\boldsymbol{g}}$ is the weighted orthogonal projection, the points $\boldsymbol{0}$, $\boldsymbol{y}$, and $\boldsymbol{L}\hat{\boldsymbol{g}}$ span a right triangle. Its hypotenuse from $\boldsymbol{L}\hat{\boldsymbol{g}}$ divides it in two triangles which are similar to the original one, cf. Figure 9.1. Thus,

$$\frac{\|\boldsymbol{y} - \boldsymbol{L}\hat{\boldsymbol{g}}\|_{\boldsymbol{W}}}{\|\boldsymbol{y}\|_{\boldsymbol{W}}} = \frac{\|\boldsymbol{y}\|_{\boldsymbol{W}} - \frac{\langle \boldsymbol{L}\hat{\boldsymbol{g}}, \boldsymbol{y}\rangle_{\boldsymbol{W}}}{\|\boldsymbol{y}\|_{\boldsymbol{W}}}}{\|\boldsymbol{y} - \boldsymbol{L}\hat{\boldsymbol{g}}\|_{\boldsymbol{W}}} \,,$$

which implies

$$\|\boldsymbol{y} - \boldsymbol{L}\hat{\boldsymbol{g}}\|_{\boldsymbol{W}}^2 = \|\boldsymbol{y}\|_{\boldsymbol{W}}^2 - \langle \boldsymbol{L}\hat{\boldsymbol{g}}, \boldsymbol{y}\rangle_{\boldsymbol{W}} \,. \qquad \blacksquare$$

Next, we want to estimate the error which comes from using the approximated cross-validation score. We start with an estimation of the approximated diagonal entries $\tilde{h}_{i,i}$.

Lemma 9.18. *Let points $\boldsymbol{X}$ and weights $\boldsymbol{W}$ form an L_2-MZ inequality for the function space $V = \operatorname{span}\{\eta_1, \dots, \eta_{m-1}\}$ with constants A and B. Further let $h_{i,i}$ be the diagonal entries from Theorem 9.13 and $\tilde{h}_{i,i}$ be the approximated ones from Theorem 9.16. Then*

$$\frac{2A}{A+B}\tilde{h}_{i,i} \leq h_{i,i} \leq \frac{2B}{A+B}\tilde{h}_{i,i} \,.$$

Proof. We show the assertion for $\tilde{g}_{i,i} = \frac{A+B}{2AB}\omega_i N(V, \boldsymbol{x}^i)$ instead of $\tilde{h}_{i,i}$, which differs in omitting the maximum. Because of the boundedness of $h_{i,i}$, the original assertion follows from Theorem 9.17.

Let $\boldsymbol{y}^i = (\eta_1(\boldsymbol{x}^i), \dots, \eta_{m-1}(\boldsymbol{x}^i))^\mathsf{T}$. Then

$$\begin{aligned} h_{i,i} &= \omega_i (\boldsymbol{y}^i)^* (\boldsymbol{L}^* \boldsymbol{W} \boldsymbol{L})^{-1} \boldsymbol{y}^i \\ &= \frac{A+B}{2AB}\omega_i (\boldsymbol{y}^i)^* \boldsymbol{y}^i \, \frac{2AB}{A+B} \, \frac{(\boldsymbol{y}^i)^* (\boldsymbol{L}^* \boldsymbol{W} \boldsymbol{L})^{-1} \boldsymbol{y}^i}{(\boldsymbol{y}^i)^* \boldsymbol{y}^i} \,. \end{aligned}$$

The first factor evaluates to $\frac{A+B}{2AB}\omega_i (\boldsymbol{y}^i)^* \boldsymbol{y}^i = \tilde{h}_{i,i}$. The latter is a Rayleigh quotient and is bounded by the reciprocals of the singular values of $\boldsymbol{L}^* \boldsymbol{W} \boldsymbol{L}$, which by Theorem 6.2 are $1/A$ and $1/B$, which yields the assertion. ■

By this lemma we obtain a bound on the difference of the original diagonal entry $h_{i,i}$ and the approximated one $\tilde{h}_{i,i}$.

Corollary 9.19. *Let points $\boldsymbol{X}$ and weights $\boldsymbol{W}$ form an L_2-MZ inequality for the function space V with bounds A and B. Further let $h_{i,i}$ the diagonal entries from Theorem 9.13 and $\tilde{h}_{i,i}$ the approximated ones from Theorem 9.16. Then*

$$|h_{i,i} - \tilde{h}_{i,i}| \leq \tilde{h}_{i,i} \frac{B-A}{A+B} \,.$$

With the bound on the approximated diagonal entries $\tilde{h}_{i,i}$ we now estimate the error of the approximated cross-validation score FCV_β.

Theorem 9.20. *Let $\boldsymbol{X}$ and $\boldsymbol{W}$ form an L_2-MZ inequality for V with constants A and B. Further, let $\mathrm{CV}_\beta(\boldsymbol{S}_V^{\boldsymbol{X}} \boldsymbol{y})$ be the cross-validation score from Theorem 9.1 and $\mathrm{FCV}_\beta(\boldsymbol{S}_V^{\boldsymbol{X}} \boldsymbol{y})$ be the approximated cross-validation score from Theorem 9.16 for the least squares approximation $S_V^{\boldsymbol{X}} \boldsymbol{y}$. Then*

$$\frac{|\,\mathrm{CV}_\beta(S_V^{\boldsymbol{X}} \boldsymbol{y}) - \mathrm{FCV}_\beta(S_V^{\boldsymbol{X}} \boldsymbol{y})|}{\mathrm{CV}_\beta(S_V^{\boldsymbol{X}} \boldsymbol{y})} \leq \frac{B-A}{B+A} \max_{i=1,\dots,n} \frac{2\tilde{h}_{i,i}}{(1-\tilde{h}_{i,i})^2}$$

for all data vectors $\boldsymbol{y} \in \mathbb{K}^n$.

In particular, if $\boldsymbol{X}$ *and* $\boldsymbol{W}$ *form a tight* L_2*-MZ inequality for* V*, we have* $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) = \mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ *for all data vectors* $\boldsymbol{y} \in \mathbb{K}^n$.

Proof. We have

$$\begin{aligned}
&|\,\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})| \\
&\quad\le \frac{1}{n}\sum_{i=1}^{n} \beta(\boldsymbol{x}^i)|(S_V^{\boldsymbol{X}}\boldsymbol{y})(\boldsymbol{x}^i) - y_i|^2 \Big|\frac{1}{(1-h_{i,i})^2} - \frac{1}{(1-\tilde{h}_{i,i})^2}\Big| \\
&\quad= \frac{1}{n}\sum_{i=1}^{n} \beta(\boldsymbol{x}^i)\frac{|(S_V^{\boldsymbol{X}}\boldsymbol{y})(\boldsymbol{x}^i) - y_i|^2}{(1-h_{i,i})^2}\Big|1 - \frac{(1-h_{i,i})^2}{(1-\tilde{h}_{i,i})^2}\Big| \\
&\quad\le \mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) \max_{i=1,\dots,n}\Big|1 - \frac{(1-h_{i,i})^2}{(1-\tilde{h}_{i,i})^2}\Big| .
\end{aligned}$$

We use $\tilde{h}_{i,i} \le 1$ and $h_{i,i} \le 1$, cf. Theorem 9.17, to estimate the second factor

$$\begin{aligned}
\Big|1 - \frac{(1-h_{i,i})^2}{(1-\tilde{h}_{i,i})^2}\Big| &\le \Big|1 + \frac{1-h_{i,i}}{1-\tilde{h}_{i,i}}\Big|\Big|1 - \frac{1-h_{i,i}}{1-\tilde{h}_{i,i}}\Big| \\
&\le \Big|\frac{2-h_{i,i}-\tilde{h}_{i,i}}{1-\tilde{h}_{i,i}}\Big|\Big|\frac{h_{i,i}-\tilde{h}_{i,i}}{1-\tilde{h}_{i,i}}\Big| \\
&\le \frac{2}{(1-\tilde{h}_{i,i})^2}|h_{i,i}-\tilde{h}_{i,i}|
\end{aligned}$$

By the Corollary 9.19 we obtain the assertion

$$\Big|1 - \frac{(1-h_{i,i})^2}{(1-\tilde{h}_{i,i})^2}\Big|^2 \le \frac{2\tilde{h}_{i,i}}{(1-\tilde{h}_{i,i})^2}\frac{B-A}{A+B} .$$

The in particular part follows from Theorem 6.2. ■

To discuss Theorem 9.11 we resume to Theorem 8.6.

Example 9.21. *We know from Theorem 9.15 that the Christoffel function for trigonometric polynomials on the torus* $\mathbb{T}$ *evaluates to* $N(V, \boldsymbol{x}) \equiv m$*. Let further* $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^n\} \subseteq \mathbb{T}^d$ *be points fulfilling an equal weighted, i.e.,* $\omega_i = 1$*,* L_2*-MZ inequality for* V *with constants* $A = 1/2n$ *and* $B = 3/2n$*. An example for these points would be an uniform random draw, cf. Theorem 6.4. Then* $\tilde{h}_{i,i} = \min\{4m/(3n), 1\}$.

The interesting range of polynomial degrees is around the optimal one. Consulting Theorem 8.6, we know that for $f \in H^s$ this is achieved by $m^\star \sim n^{\frac{1}{2s+1}}$. Up to constants Theorem 9.20 would than yield an relative error of

$$\frac{|\operatorname{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \operatorname{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})|}{\operatorname{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})} \lesssim \frac{m}{n} \sim n^{-\frac{2s}{2s+1}}\,.$$

The nearly linear decay from Theorem 9.21 for the relative difference of the cross-validation score $\operatorname{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ and approximated cross-validation score $\operatorname{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ is better than the expected error decay from Theorem 8.6 itself. Thus, using the approximated cross-validation score instead of the exact one does not affect the performance of the parameter choice strategy.

9.3 Numerical experiments

In this section we validate our theoretical findings from Sections 9.1 and 9.2. Namely, we will have a look at:

- the cross-validation score $\operatorname{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma^2$ being an estimator of the L_2-error $\|f_q - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2$ and
- the fast cross-validation score $\operatorname{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ being an estimator of the cross-validation score $\operatorname{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ whilst being fast to compute.

9.3.1 Numerical experiment on the torus $\mathbb{T}$

We start with a simple example on the one-dimensional torus $\mathbb{T}$ with a smaller number of points. As test function we use the B-spline of order two

$$B_2(x) = \begin{cases} \sqrt{12}x & \text{for } 0 \le x \le 1/2 \\ \sqrt{12}(1-x) & \text{for } 1/2 \le x \le 1\,, \end{cases}$$

which is L_2-normalized and has Sobolev smoothness $s = 3/2 - \varepsilon$ for any $\varepsilon > 0$. We know the exact Fourier coefficients $\hat{f}_k$, allowing us to compute the L_2-error of the least squares approximation $S_V^{\boldsymbol{X}}\boldsymbol{y}$ exactly using Parseval's equality

$$\|B_2 - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 = \|B_2\|_{L_2}^2 - \sum_{k\le m} |\hat{f}_k|^2 + \sum_{k\le m} |\hat{f}_k - \hat{g}_k|^2\,,$$

with $\hat{g}_k$ the Fourier coefficients of $S_V^{\boldsymbol{X}}\boldsymbol{y}$.

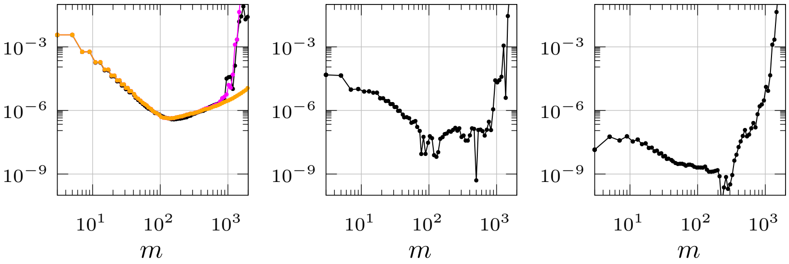


Figure 9.2: One-dimensional experiment on the torus $\mathbb{T}$. In the left figure shows the L_2-error in orange, the cross-validation score $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma_{\text{emp}}^2$ in black, and the fast cross-validation score $\mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma_{\text{emp}}^2$ in magenta. The middle shows the difference of L_2-error and cross-validation score $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma_{\text{emp}}^2$. The right figure shows the difference of cross-validation score $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma_{\text{emp}}^2$ and the fast cross-validation score $\mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma_{\text{emp}}^2$.

For our experiment we choose $n = 3\,000$ uniform random points in which we evaluate the test function B_2 and add Gaussian noise $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ with a variance $\sigma^2 = 0.05$. For different polynomial degrees m we compute the least squares approximation $S_V^{\boldsymbol{X}}\boldsymbol{y}$ for $V = \{\exp(2\pi \mathrm{i}(-1)^k \lfloor k/2 \rfloor \cdot)\}_{k=1}^{m-1}$. We measure the following quantities

- the L_2-error as shown above,

- the exact cross-validation score $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$, and

- the fast cross-validation score $\mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ with the assumption $A = B = 1$ for the approximated diagonal elements $\tilde{h}_{i,i}$ and the Christoffel function $N(V,\cdot) \equiv m$.

Further, we computed the empirical variance of the noise given by $\sigma_{\text{emp}}^2 = 1/n \sum_{i=1}^n |\varepsilon_i|^2$ being the real variance of the realization of the noise $\boldsymbol{\varepsilon}$.

The results are depicted in Figure 9.2. We see that the three quantities are close except for larger polynomial degrees, where the theory is limited anyways. In particular, the error of the cross-validation score against the L_2-error, $\|B_2 - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2 - \mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) + \sigma_{\text{emp}}^2$, is smaller than the actual L_2-error making cross-validation viable to use as its estimator. Furthermore, the difference between fast cross-validation score $\mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ and cross-validation score $\mathrm{CV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ is smaller than the L_2-error as well, making the fast cross-validation score a viable estimator of the cross-validation score.

In the experiment we see that the fast cross-validation score $\mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ is a good estimator for the L_2-error for small polynomial degrees whilst obtaining a speedup of a factor $n = 3\,000$, which is the number of points of the experiment.

The speedup and small error of the fast cross-validation score to the exact cross-validation score was expected from the theory. The small error of the cross-validation score to the L_2-error complies the theory and is even better suggesting room for improvement.

9.3.2 Numerical experiment on the unit cube $[0,1]^d$

As the theory is basis- and dimension-independent, in this section we resume to the more practical example from Section 8.4 being on the five-dimensional unit cube $[0,1]^5$ and we approximate with the $H^2([0,1])$ basis from Theorem 3.27.

The setup of the experiment is the same as in Section 8.4: We approximate the tensorized, shifted, and dilated B-spline of order two $f_q(\boldsymbol{x}) = \prod_{j=1}^d B_2^{\mathrm{cut}}(x_j)$ from (8.12), which we sample in $n = 1\,000\,000$ uniform random points. We use three different noise levels $\sigma^2 \in \{0.00, 0.03, 0.05\}$ and compute for different polynomial degrees

- the L_2-error $\|B_2^{\mathrm{cut}} - S_V^{\boldsymbol{X}}\boldsymbol{y}\|_{L_2}^2$ by using Parseval's identity as done in the previous section and

- the fast cross-validation score $\mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y})$ from Theorem 9.16.

Note, the Christoffel function $N(V,\cdot)$ is not constant as before and we compute it in a naive manner in $\mathcal{O}(n \cdot m)$. As we use matrix-vector multiplications for computing the least squares approximation, this is no bottleneck and does not deteriorate the overall computation complexity. We omit the computation of the exact cross-validation score as it would be too slow.

The results are depicted in Figure 9.3. With the high number of points $n = 1\,000\,000$, the theory suggests a smaller difference of L_2-error and fast cross-validation score. Indeed, in the plots these two quantities are indistinguishable and their difference is magnitudes smaller.

This experiments shows, that

- with a bigger number of points, the use of fast cross-validation is even more viable and

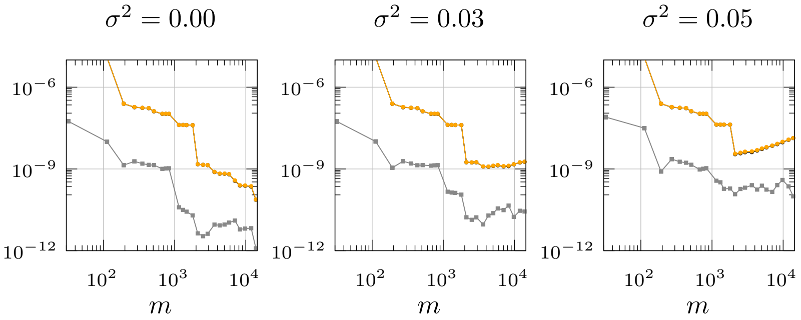


Figure 9.3: Five-dimensional experiment on the unit cube $[0,1]^5$ for different levels of noise $\sigma^2 \in \{0, 0.03, 0.05\}$. The figures shows the L_2-error in black and the fast cross-validation score $\mathrm{FCV}_\beta(S_V^{\boldsymbol{X}}\boldsymbol{y}) - \sigma^2_{\text{emp}}$ in orange. In gray we plotted their difference.

- even though we are on a different domain with a different basis, we obtain a small error, suggesting the generality of our approach.

Chapter 10

Outlook

We use this final chapter to comment on possible starting points for advancements including subjects which have not made it into the thesis but are already published. Overall, this thesis serves as a plea for the over 200 years old least squares approximation method showing its fast implementation and theoretical optimality. In Sections 7.4 and 8.5 we compared it to other algorithms in the worst-case and statistical learning setting, respectively. It shows that least squares approximation, when used right, is a universal tool applicable in all settings of function approximation. With the vast amount of topics covered in this thesis there remain several interesting research directions:

- In Theorem 3.27 we had a look at the $H^2([0,1])$-basis, which is now ready to use in practical applications as our numerical experiments in Sections 8.3.2, 8.4, and 9.3.2 show. Theoretically, higher order basis were also considered in [AIN12, Section 3] but not implemented. As approximation on the unit interval and cube has many practical applications, one can have a look at the numerical implementation of higher order bases.

- For the `BSS`-algorithm from Section 5.2 we observed cubic runtime in practice. We had to stop the numerical experiments for frames of dimension $\geq 1\,000$, as the algorithm was too slow. Many subsampling techniques were only introduced theoretically without big concern about the runtime. An interesting question is a possible speedup of the algorithm.

- In Theorem 7.8 the bounds are near-optimal up to a single logarithm. The existence of a optimal points set closing that gap is known by using the Kadison-Singer theorem in [DKU23], which is not constructive. It is open whether there is a polynomial time algorithm which gives points fulfilling the optimal error bound.

- On the same note, in Theorem 7.8 and in the tight error bound in [DKU23] the realizing algorithm was the **weighted** least squares approximation. In particular settings like on the torus $\mathbb{T}^d$ the weights turn out to all be equal. However, it is open if this is possible in general and the optimal rate is achieved via **plain** least squares approximation.

- In Section 7.2.3 we proposed using subsampled L_2-MZ inequalities for function approximation in the worst-case setting. This works also in the individual function approximation setting as we have already tried with the example of rank-1 lattices in [BT23]. However, this was not the focus of this paper and a thorough investigation of using general subsampled L_2-MZ inequalities in individual function approximation would be interesting.

- We have shown the concentration of the cross-validation score in Theorem 9.10. However, in Theorem 9.12 we see that the concentration bound deteriorates the theoretically optimal error. Further, the numerical experiments suggest a better performance of the cross-validation score. So, there is a gap one can investigate.

List of Symbols

Fields and domains

$\mathbb{N}$	set of positive integers
$\mathbb{N}_0$	set of nonnegative integers
$\mathbb{Z}$	set of integers
$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex numbers
$\mathbb{K}$	either the set of real or complex numbers
$\mathbb{T}$	one-dimensional torus
$\mathbb{T}^d$	d-dimensional torus

Vector and matrix related notions

$\|\boldsymbol{k}\|_p$	p-norm $(\sum_{j=1}^{d} \|k_j\|^p)^{1/p}$
$\|\boldsymbol{A}\|_F$	Frobenius norm
$\|\boldsymbol{A}\|_{2\to 2}$	spectral norm
$\sigma_{\min/\max}(\boldsymbol{A})$	minimal/maximal singular value
$\lambda_{\min/\max}(\boldsymbol{A})$	minimal/maximal eigenvalue

Important numbers and notions

$\delta_{k,l}$	Kronecker delta
$S_V^{\boldsymbol{X}}$	least squares approximation
$S^{\boldsymbol{X}}$	kernel method
g_n	sampling width
$g_{n,m}^{\mathrm{ls}}$	sampling width restricted to the least squares approximation
a_m	linear width
d_m	Kolmogorov width
$N(V,\cdot)$	Christoffel function
$N(V)$	$\sup_{x\in D} N(V,x)$
$\mathcal{O}(\cdot)$	Landau symbol
CV_β	cross-validation score

Function spaces and norms

$H(K)$	reproducing kernel Hilbert space (RKHS) with kernel K
$L_2(D, \varrho_T)$	space of square-integrable functions on D
$\lVert f \rVert_{L_2}$	$(\int_D \lvert f \rvert^2 \,\mathrm{d}\varrho_T)^{1/2}$
$L_\infty(D, \varrho_T)$	space of essentially bounded functions on D
$\lVert f \rVert_{L_\infty}$	$\operatorname{ess\,sup} \lvert f \rvert = \inf\{a \in \mathbb{R} : \varrho_T(\lvert f \rvert^{-1}((a, \infty))) = \emptyset\}$
$\ell_\infty(D)$	class of bounded functions on D
$\lVert f \rVert_\infty$	$\sup_{\boldsymbol{x} \in D} \lvert f(\boldsymbol{x}) \rvert$
$H^s(\mathbb{T})$	Sobolev space of smoothness s on the torus $\mathbb{T}$
$H^s([0,1])$	Sobolev space of smoothness s on the unit interval $[0,1]$
$\lVert f \rVert_{H^s}$	$(\lVert f \rVert_{L_2}^2 + \lVert f^{(s)} \rVert_{L_2}^2)^{1/2}$
$L_2^s([-1,1], v^{\alpha,\beta})$	Sobolev-type subpace
$p_k^{\alpha,\beta}$	Jacobi polynomial of degree k
$\lVert f \rVert_{L_2^s([-1,1], v^{\alpha,\beta})}$	$(\sum_{k=0}^\infty (k+1)^{2s} \lvert \langle f, p_k^{\alpha,\beta} \rangle_{\alpha,\beta} \rvert^2)^{1/2}$
$\langle f, g \rangle_{\alpha,\beta}$	$\int_{-1}^1 f(x) \overline{g(x)} v^{\alpha,\beta}(x) \,\mathrm{d}x$
$v^{\alpha,\beta(x)}$	$(1-x)^\alpha (1+x)^\beta$
$C^s(D)$	s-times continuously differentiable functions
$H^s(D)$	isotropic Sobolev space of smoothness s
$\lVert f \rVert_{H^s}$	$(\sum_{\lVert \boldsymbol{\alpha} \rVert_1 \leq s} \lVert D^{\boldsymbol{\alpha}} f \rVert_{L_2}^2)^{1/2}$
$H^s_{\text{mix}}(d)$	Sobolev space with dominating mixed smoothness
$\lVert f \rVert_{H^s_{\text{mix}}}$	$(\sum_{\boldsymbol{\alpha} \in \{0,s\}^d} \lVert D^{\boldsymbol{\alpha}} f \rVert_{L_2}^2)^{1/2}$

Bibliography

[AB22] B. Adcock and S. Brugiapaglia. Is Monte Carlo a bad sampling strategy for learning smooth functions in high dimensions? *arXiv:math/2208.09045v2*, 2022.

[ACDM22] B. Adcock, J. M. Cardenas, N. Dexter, and S. Moraga. *Towards Optimal Sampling for Learning Sparse Approximations in High Dimensions*, pages 9–77. Springer International Publishing, Cham, 2022.

[Adc10a] B. Adcock. *Modified Fourier expansions: theory, construction and applications*. PhD thesis, University of Cambridge, 2010.

[Adc10b] B. Adcock. Multivariate modified Fourier series and application to boundary value problems. *Numer. Math.*, 115(4):511–552, 2010.

[Adr08] R. Adrain. Research concerning the probabilities of the errors which happen in making observations. *The Analyst, or, Mathematical Museum*, 1(4):93–109, 1808.

[AF03] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[AH11] B. Adcock and D. Huybrechs. Multivariate modified Fourier expansions. In *Spectral and high order methods for partial differential equations*, volume 76 of *Lect. Notes Comput. Sci. Eng.*, pages 85–92. Springer, Heidelberg, 2011.

[AH12] K. Atkinson and W. Han. *Spherical harmonics and approximations on the unit sphere: an introduction*, volume 2044 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2012.

[AIN12] B. Adcock, A. Iserles, and S. P. Nørsett. From high oscillation to rapid approximation II: expansions in Birkhoff series. *IMA J. Numer. Anal.*, 32(1):105–140, 2012.

[APSV18] D. Alfke, D. Potts, M. Stoll, and T. Volkmer. NFFT meets krylov methods: Fast matrix-vector products for the graph laplacian of fully connected networks. *Frontiers in Applied Mathematics and Statistics*, 4, December 2018.

[Aro35] N. Aronszajn. Sur les décompositions des fonctions analytiques uniformes et sur leurs applications. *Acta Math.*, 65(1):1–156, 1935.

[Aro44] N. Aronszajn. La théorie générale des noyaux reprodusants et ses applications. *Proc. Cambridge Philos. Soc.*, 39:133, 1944.

[Aro50] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[Bab60] K. I. Babenko. Approximation of a certain class of periodical functions of many variables by trigonometric polynomials. *Dokl. Akad. Nauk SSSR*, 132:982–985, 1960.

[Bar02] Y. Baraud. Model selection for regression on a random design. *ESAIM Probab. Stat.*, 6:127–146, 2002.

[Bar23] F. Bartel. Stability and error guarantees for least squares approximation with noisy samples. *SMAI J. Comput. Math.*, 9:95–120, 2023.

[BDSU16] G. Byrenheid, D. Dũng, W. Sickel, and T. Ullrich. Sampling on energy-norm based sparse grids for the optimal recovery of Sobolev type functions in h^γ. *J. Approx. Theory*, 207:207–231, 2016.

[BE02] O. Bousquet and A. Elisseeff. Stability and generalization. *J. Mach. Learn. Res.*, 2:499–526, 2002.

[Bec11] S. M. A. Becker. Regularization of statistical inverse problems and the Bakushinskiĭ veto. *Inverse Problems*, 27(11):115010, 22, 2011.

[Bel19] P. C. Bellec. Concentration of quadratic forms under a Bernstein moment assumption. *arXiv:math/1901.08736*, 2019.

[Ber22] S. Bergmann. Über die Entwicklung der harmonischen Funktionen der Ebene und des Raumes nach Orthogonalfunktionen. *Math. Ann.*, 86(3-4):238–271, 1922.

[BH22] F. Bartel and R. Hielscher. Concentration inequalities for cross-validation in scattered data approximation. *J. Approx. Theory*, 277:Paper No. 105715, 17, 2022.

[BHP20] F. Bartel, R. Hielscher, and D. Potts. Fast cross-validation in harmonic approximation. *Appl. Comput. Harmon. Anal.*, 49(2):415–437, 2020.

[BHS92] D. Berthold, W. Hoppe, and B. Silbermann. A fast algorithm for solving the generalized airfoil equation. *J. Comput. Appl. Math.*, 43(1):185–219, 1992.

[BHT23] S. Bates, T. Hastie, and R. Tibshirani. Cross-validation: What does it estimate and how well does it do it? *J. Amer. Statist. Assoc.*, pages 1–12, 2023.

[Bjö96] Å. Björck. *Numerical methods for least squares problems*. SIAM, 1996.

[BKPU23] F. Bartel, L. Kämmerer, D. Potts, and T. Ullrich. Reconstructing functions from values at subsampled quadrature points. *Math. Comp.*, published electronically, 2023.

[BKUV17] G. Byrenheid, L. Kämmerer, T. Ullrich, and T. Volkmer. Tight error bounds for rank-1 lattice sampling in spaces of hybrid mixed smoothness. *Numer. Math.*, 136(4):993–1034, 2017.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013.

[BLV08] D. Bilyk, M. T. Lacey, and A. Vagharshakyan. On the small ball inequality in all dimensions. *Journal of Functional Analysis*, 254(9):2470–2502, May 2008.

[BN79] J. R. Bunch and C. P. Nielsen. Updating the singular value decomposition. *Numer. Math.*, 31(2):111–129, 1978/79.

[Boc22] S. Bochner. Über orthogonale Systeme analytischer Funktionen. *Math. Z.*, 14(1):180–207, 1922.

[BPS22] F. Bartel, D. Potts, and M. Schmischke. Grouped transformations and regularization in high-dimensional explainable ANOVA approximation. *SIAM J. Sci. Comput.*, 44(3):A1606–A1631, 2022.

[BR08] F. Bauer and M. Reiß. Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Problems*, 24(5):055009, 16, 2008.

[BS47a] S. Bergman and M. Schiffer. On Green's and Neumann's functions in the theory of partial differential equations. *Bull. Amer. Math. Soc.*, 53:1141–1151, 1947.

[BS47b] S. Bergman and M. Schiffer. A representation of Green's and Neumann's functions in the theory of partial differential equations of second order. *Duke math. J.*, 14:609–635, 1947.

[BS48] S. Bergman and M. Schiffer. Kernel functions in the theory of partial differential equations of elliptic type. *Duke math. J.*, 15:535–566, 1948.

[BS02] H. Blockeel and J. Struyf. Efficient algorithms for decision tree cross-validation. *J. Mach. Learn. Res.*, 3:621–650, 2002.

[BSS09] J. D. Batson, D. A. Spielman, and N. Srivastava. Twice-Ramanujan sparsifiers. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 255–262. ACM, New York, 2009.

[BSU23] F. Bartel, M. Schäfer, and T. Ullrich. Constructive subsampling of finite frames with applications in optimal function recovery. *Appl. Comput. Harmon. Anal.*, 65:209–248, 2023.

[BT23] F. Bartel and F. Taubert. Nonlinear approximation with subsampled rank-1 lattices. In *Fourteenth International Conference on Sampling Theory and Applications*, 2023.

[BTA04] A. Berlinet and C. Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Publishers, Boston, MA, 2004. With a preface by Persi Diaconis.

[CDL13] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13(5):819–834, 2013.

[CFM12] P. G. Casazza, M. Fickus, and D. G. Mixon. Auto-tuning unit norm frames. *Appl. Comput. Harmon. Anal.*, 32(1):1–15, 2012.

[Chr08] O. Christensen. *Frames and bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2008. An introductory course.

[CK03] P. G. Casazza and J. Kovačević. Equal-norm tight frames with erasures. *Adv. Comput. Math.*, 18(2/4):387–430, 2003.

[CK13] P. G. Casazza and G. Kutyniok, editors. *Finite Frames*. Birkhäuser Boston, 2013.

[CKNS16] R. Cools, F. Y. Kuo, D. Nuyens, and G. Suryanarayana. Tent-transformed lattice rules for integration and approximation of multivariate non-periodic functions. *J. Complexity*, 36:166–181, 2016.

[CM17] A. Cohen and G. Migliorati. Optimal weighted least-squares methods. *SMAI J. Comput. Math.*, 3:181–203, 2017.

[CMO97] R. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *J. Comput. Finance*, 1(1):27–46, 1997.

[CN07] R. Cools and D. Nuyens. An overview of fast component-by-component constructions of lattice rules and lattice sequences. *PAMM*, 7:1022609–1022610, 2007.

[Com15] R. Combes. An extension of McDiarmid's inequality. *arXiv:cs/1511.05240*, 2015.

[Con90] J. B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[CP11] R. Cools and K. Poppe. Chebyshev lattices, a unifying framework for cubature with Chebyshev weight function. *BIT*, 51(2):275–288, 2011.

[CS02] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin of the American mathematical society*, 39(1):1–49, 2002.

[CT65] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

[CZ07] F. Cucker and D. X. Zhou. *Learning Theory*. Cambridge University Press, 2007.

[DC22a] M. Dolbeault and A. Cohen. Optimal pointwise sampling for L^2 approximation. *J. Complexity*, 68:101602, 2022.

[DC22b] M. Dolbeault and A. Cohen. Optimal sampling and Christoffel functions on general domains. *Constr. Approx.*, 56(1):121–163, 2022.

[DG91] L. N. Deshpande and D. Girard. Fast computation of cross-validated robust splines and other non-linear smoothing splines. *Curves and Surfaces*, pages 143–148, 1991.

[DKP22] J. Dick, P. Kritzer, and F. Pillichshammer. *Lattice Rules*. Springer International Publishing, 2022.

[DKU23] M. Dolbeault, D. Krieg, and M. Ullrich. A sharp upper bound for sampling numbers in l2. *Applied and Computational Harmonic Analysis*, 63:113–134, 2023.

[DNP14] J. Dick, D. Nuyens, and F. Pillichshammer. Lattice rules for nonperiodic smooth integrands. *Numer. Math.*, 126(2):259–291, 2014.

[DNPV12] E. Di Nezza, G. Palatucci, and E. Valdinoci. Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.*, 136(5):521–573, 2012.

[DPS$^+$21] F. Dai, A. Prymak, A. Shadrin, V. Temlyakov, and S. Tikhonov. Entropy numbers and Marcinkiewicz-type discretization. *J. Funct. Anal.*, 281(6):109090, 2021.

[DR22] X. Dong and M. Rudelson. Approximately Hadamard matrices and Riesz bases in random frames. *arXiv:math/2207.07523*, 2022.

[DS52] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72:341–366, 1952.

[DS00] J. Dongarra and F. Sullivan. Guest editors introduction to the top 10 algorithms. *Computing in Science & Engineering*, 2(01):22–23, 2000.

[DuTU18] D. Dũng, V. N. Temlyakov, and T. Ullrich. *Hyperbolic cross approximation*. Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser/Springer, Cham, 2018. Edited and with a foreword by Sergey Tikhonov.

[DuTU19] D. Dũng, V. N. Temlyakov, and T. Ullrich. *Hyperbolic Cross Approximation*. Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser/Springer, 2019.

[DVPR10] E. De Vito, S. Pereverzyev, and L. Rosasco. Adaptive kernel methods using the balancing principle. *Found. Comput. Math.*, 10(4):455–479, 2010.

[DX13] F. Dai and Y. Xu. *Approximation theory and harmonic analysis on spheres and balls*. Springer Monographs in Mathematics. Springer, New York, 2013.

[FGS98] W. Freeden, T. Gervens, and M. Schreiner. *Constructive approximation on the sphere*. Numerical Mathematics and Scientific Computation. The Clarendon Press, Oxford University Press, New York, 1998. With applications to geomathematics.

[FM11] F. Filbir and H. N. Mhaskar. Marcinkiewicz-Zygmund measures on manifolds. *J. Complexity*, 27:568–598, 2011.

[FP05] M. Fenn and D. Potts. Fast summation based on fast trigonometric transforms at non-equispaced nodes. *Numer. Linear Algebra Appl.*, 12(2-3):161–169, 2005.

[FR13] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[Gau11] C. F. Gauss. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Cambridge Library Collection. Cambridge University Press, Cambridge, 2011. Reprint of the 1809 original.

[GHH11] M. Griebel, J. Hamaekers, and F. Heber. BOSSANOVA - A bond order dissection approach for efficient electronic structure calculations. *Oberwolfach Report*, 32:1804–1808, 2011.

[GHW79] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[GIKV21] C. Gross, M. A. Iwen, L. Kämmerer, and T. Volkmer. A deterministic algorithm for constructing multiple rank-1 lattices of near-optimal size. *Adv. Comput. Math.*, 47(6):Paper No. 86, 24, 2021.

[GKKW02] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York, 2002.

[GMM$^+$22] E. R. Gizewski, L. Mayer, B. A. Moser, D. H. Nguyen, S. Pereverzyev, S. V. Pereverzyev, N. Shepeleva, and W. Zellinger. On a regularization of unsupervised domain adaptation in RKHS. *Appl. Comput. Harmon. Anal.*, 57:201–227, 2022.

[God48] R. Godement. Les fonctions de type positif et la théorie des groupes. *Trans. Amer. Math. Soc.*, 63:1–84, 1948.

[Gre97] A. Greenbaum. *Iterative methods for solving linear systems*, volume 17 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[Grö20] K. Gröchenig. Sampling, Marcinkiewicz-Zygmund inequalities, approximation, and quadrature rules. *J. Approx. Theory*, 257:105455, 20, 2020.

[GSY19] T. Goda, K. Suzuki, and T. Yoshiki. Lattice rules in non-periodic subspaces of Sobolev spaces. *Numer. Math.*, 141(2):399–427, 2019.

[Gu13] C. Gu. *Smoothing Spline ANOVA Models*. Springer New York, 2013.

[Har97] D. A. Harville. *Matrix algebra from a statistician's perspective*. Springer-Verlag, New York, 1997.

[HD15] J. Hampton and A. Doostan. Coherence motivated sampling and convergence analysis of least squares polynomial Chaos regression. *Comput. Methods Appl. Mech. Engrg.*, 290:73–97, 2015.

[HGB$^+$06] J. Huang, A. Gretton, K. Borgwardt, B. Schölkopf, and A. Smola. Correcting sample selection bias by unlabeled data. *Advances in neural information processing systems*, 19, 2006.

[HNP22] C. Haberstich, A. Nouy, and G. Perrin. Boosted optimal weighted least-squares. *Math. Comp.*, 91(335):1281–1315, 2022.

[HNV08] A. Hinrichs, E. Novak, and J. Vybíral. Linear information versus function evaluations for L_2-approximation. *J. Approx. Theory*, 153(1):97–107, 2008.

[HS52] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436 (1953), 1952.

[HTF09] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. Springer Series in Statistics. Springer, New York, second edition, 2009. Data mining, inference, and prediction.

[IKP18] C. Irrgeher, P. Kritzer, and F. Pillichshammer. Integration and approximation in cosine spaces of smooth functions. *Math. Comput. Simulation*, 143:35–45, 2018.

[IN08] A. Iserles and S. P. Nørsett. From high oscillation to rapid approximation. I. Modified Fourier expansions. *IMA J. Numer. Anal.*, 28(4):862–887, 2008.

[IN09] A. Iserles and S. P. Nørsett. From high oscillation to rapid approximation. III. Multivariate expansions. *IMA J. Numer. Anal.*, 29(4):882–916, 2009.

[JMN21] P. Junghanns, G. Mastroianni, and I. Notarangelo. *Weighted polynomial approximation and numerical methods for integral equations*. Pathways in Mathematics. Birkhäuser/Springer, Cham, 2021.

[KK11] L. Kämmerer and S. Kunis. On the stability of the hyperbolic cross discrete Fourier transform. *Numer. Math.*, 117(3):581–600, 2011.

[KKK$^+$21a] V. Kaarnioja, Y. Kazashi, F. Y. Kuo, F. Nobile, and I. H. Sloan. Fast approximation by periodic kernel-based lattice-point interpolation with application in uncertainty quantification. *Numer. Math.*, 150(1):33–77, 2021.

[KKK$^+$21b] V. Kaarnioja, Y. Kazashi, F. Y. Kuo, F. Nobile, and I. H. Sloan. Fast approximation by periodic kernel-based lattice-point interpolation with application in uncertainty quantification. *Numerische Mathematik*, 150(1):33–77, November 2021.

[KKLT22] B. Kashin, E. Kosov, I. Limonova, and V. N. Temlyakov. Sampling discretization and related problems. *J. Complexity*, 71:Paper No. 101653, 55, 2022.

[KKP07] J. Keiner, S. Kunis, and D. Potts. Efficient reconstruction of functions on the sphere from scattered data. *J. Fourier Anal. Appl.*, 13(4):435–458, 2007.

[KKP09] J. Keiner, S. Kunis, and D. Potts. Using NFFT 3—a software library for various nonequispaced fast Fourier transforms. *ACM Trans. Math. Software*, 36(4):Art. 19, 30, 2009.

[KL93] J. Kuelbs and W. Li. Metric entropy and the small ball problem for Gaussian measures. *Journal of Functional Analysis*, 116(1):133–157, August 1993.

[KMNN21] F. Y. Kuo, G. Migliorati, F. Nobile, and D. Nuyens. Function integration, reconstruction and approximation using rank-1 lattices. *Math. Comp.*, 90(330):1861–1897, 2021.

[KMU16] T. Kühn, S. Mayer, and T. Ullrich. Counting via entropy: new preasymptotics for the approximation numbers of Sobolev embeddings. *SIAM J. Numer. Anal.*, 54(6):3625–3647, 2016.

[KN08] S. Kindermann and A. Neubauer. On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization. *Inverse Probl. Imaging*, 2(2):291–299, 2008.

[KOUU20] C. Kacwin, J. Oettershagen, M. Ullrich, and T. Ullrich. Numerical performance of optimized frolov lattices in tensor product reproducing kernel sobolev spaces. *Found. Comput. Math.*, 21(3):849–889, July 2020.

[KPP18] S. Kindermann, S. Pereverzyev, Jr., and A. Pilipenko. The quasi-optimality criterion in the linear functional strategy. *Inverse Problems*, 34(7):075001, 24, 2018.

[KPUU23] D. Krieg, K. Pozharska, M. Ullrich, and T. Ullrich. Sampling recovery in uniform and other norms. *arXiv:math/2305.07539*, 2023.

[KPV15] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate periodic functions by trigonometric polynomials based on rank-1 lattice sampling. *J. Complexity*, 31:543–576, 2015.

[KSU14] T. Kühn, W. Sickel, and T. Ullrich. Approximation numbers of Sobolev embeddings—sharp constants and tractability. *J. Complexity*, 30(2):95–116, 2014.

[KSU15] T. Kühn, W. Sickel, and T. Ullrich. Approximation of mixed order Sobolev functions on the d-torus: asymptotics, preasymptotics, and d-dependence. *Constr. Approx.*, 42(3):353–398, 2015.

[KSWW09] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski. On decompositions of multivariate functions. *Math. Comp.*, 79(270):953–966, 2009.

[KU21a] D. Krieg and M. Ullrich. Function values are enough for L_2-approximation. *Found. Comput. Math.*, 21(4):1141–1151, 2021.

[KU21b] D. Krieg and M. Ullrich. Function values are enough for L_2-approximation: Part II. *J. Complexity*, 66:Paper No. 101569, 14, 2021.

[KUV21] L. Kämmerer, T. Ullrich, and T. Volkmer. Worst-case recovery guarantees for least squares approximation using random samples. *Constr. Approx.*, 54(2):295–352, 2021.

[Kä14a] L. Kämmerer. *High Dimensional Fast Fourier Transform Based on Rank-1 Lattice Sampling*. Dissertation. Universitätsverlag Chemnitz, 2014.

[Kä14b] L. Kämmerer. Reconstructing multivariate trigonometric polynomials from samples along rank-1 lattices. In *Approximation theory XIV: San Antonio 2013*, volume 83 of *Springer Proc. Math. Stat.*, pages 255–271. Springer, Cham, 2014.

[Kä18] L. Kämmerer. Multiple rank-1 lattices as sampling schemes for multivariate trigonometric polynomials. *J. Fourier Anal. Appl.*, 24(1):17–44, 2018.

[Kä20] L. Kämmerer. A fast probabilistic component-by-component construction of exactly integrating rank-1 lattices and applications. *arXiv:math/2012.14263*, 2020.

[LdHA10] M. A. Lukas, F. R. de Hoog, and R. S. Anderssen. Efficient algorithms for robust generalized cross-validation spline smoothing. *J. Comput. Appl. Math.*, 235:102–107, 2010.

[Leg05] A. Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. Nineteenth Century Collections Online (NCCO): Science, Technology, and Medicine: 1780-1925. F. Didot, 1805.

[Li86] K.-C. Li. Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.*, 14(3):1101–1112, 1986.

[LMP20] S. Lu, P. Mathé, and S. V. Pereverzev. Balancing principle in supervised learning for a general regularization scheme. *Appl. Comput. Harmon. Anal.*, 48(1):123–148, 2020.

[LO06] R. Liu and A. B. Owen. Estimating mean dimensionality of analysis of variance decompositions. *J. Amer. Statist. Assoc.*, 101(474):712–721, 2006.

[LPU23] L. Lippert, D. Potts, and T. Ullrich. Fast hyperbolic wavelet regression meets ANOVA. *Numerische Mathematik*, 154(1-2):155–207, June 2023.

[LT22] I. Limonova and V. Temlyakov. On sampling discretization in L_2. *J. Math. Anal. Appl.*, 515(2):Paper No. 126457, 14, 2022.

[Luk06] M. A. Lukas. Robust generalized cross-validation for choosing the regularization parameter. *Inverse Problems*, 22(5):1883–1902, 2006.

[Mam15] I. G. Mamedov. On the well-posed solvability of the Dirichlet problem for a generalized Mangeron equation with nonsmooth coefficients. *Differ. Equ.*, 51(6):745–754, 2015. Translation of Differ. Uravn. **5**1 (2015), no. 6, 733–742.

[McD89] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[Mer77] M. Merriman. *A List of Writings Relating to the Method of Least Squares: With Historical and Critical Notes*. Transactions of the Connecticut Academy of Arts and Sciences. Academy, 1877.

[Mer09] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London Ser. A*, 209:415–446, 1909.

[Mer11] J. Mercer. Sturm-Liouville series of normal functions in the theory of integral equations. *Philos. Trans. Roy. Soc. London Ser. A*, 211:111–198, 1911.

[Mes62] H. Meschkowski. *Hilbertsche Räume mit Kernfunktion*. Die Grundlehren der mathematischen Wissenschaften, Band 113. Springer-Verlag, Berlin-Göttingen-Heidelberg, 1962.

[Mic13] V. Michel. *Lectures on constructive approximation*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013. Fourier, spline, and wavelet methods on the real line, the sphere, and the ball.

[MNvST14] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Analysis of discrete L^2 projection on polynomial spaces with random evaluations. *Found. Comput. Math.*, 14:419–456, 2014.

[MNW01] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward. Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature. *Math. Comput.*, 70:1113–1130, 2001. Corrigendum on the positivity of the quadrature weights in 71:453–454, 2002.

[Moo16] E. H. Moore. On properly positive Hermitian matrices. *Bull. Amer. Math. Soc.*, 23:59, 1916.

[Moo35] E. H. Moore. *General analysis Part I*. Memoirs of the American Philosophical Society, 1935.

[Moo39] E. H. Moore. *General analysis Part II*. Memoirs of the American Philosophical Society, 1939.

[MR77] C. A. Micchelli and T. J. Rivlin. *A Survey of Optimal Recovery*, pages 1–54. Springer US, Boston, MA, 1977.

[MS00] M. Mullin and R. Sukthankar. Complete cross-validation for nearest neighbor classifiers. In *17th International Conference on Machine Learning (ICML)*, 2000.

[MSS15] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. *Ann. of Math. (2)*, 182(1):327–350, 2015.

[MU21] M. Moeller and T. Ullrich. L_2-norm sampling discretization and recovery of functions from RKHS with finite trace. *Sampl. Theory Signal Process. Data Anal.*, 19(2):13, 2021.

[MVDV92] M. Moonen, P. Van Dooren, and J. Vandewalle. A singular value decomposition updating algorithm for subspace tracking. *SIAM J. Matrix Anal. Appl.*, 13(4):1015–1038, 1992.

[MX15] F. Marcellán and Y. Xu. On Sobolev orthogonal polynomials. *Expo. Math.*, 33(3):308–352, 2015.

[MZ37] J. Marcinkiewicz and A. Zygmund. Sur les fonctions indépendantes. *Fundamenta Mathematicae*, 29(1):60–90, 1937.

[NBUL23] N. Nagel, F. Bartel, T. Ullrich, and K. Lüttgen. Efficient recovery of non-periodic multivariate functions from few scattered samples. In *Fourteenth International Conference on Sampling Theory and Applications*, 2023.

[Nev79] P. G. Nevai. Orthogonal polynomials. *Mem. Amer. Math. Soc.*, 18(213):v+185, 1979.

[Nik63] S. M. Nikol'skii. Functions with dominant mixed derivative, satisfying a multiple Hölder condition. *Sibirsk. Mat. Zh.*, 4:1342–1364, 1963.

[NJZ17] A. Narayan, J. D. Jakeman, and T. Zhou. A Christoffel function weighted least squares algorithm for collocation approximations. *Math. Comp.*, 86(306):1913–1947, 2017.

[NP22] R. Nasdala and D. Potts. A note on transformed Fourier systems for the approximation of non-periodic signals. In *Monte Carlo* *and quasi-Monte Carlo methods*, volume 387 of *Springer Proc. Math. Stat.*, pages 253–271. Springer, Cham, [2022] ©2022.

[NS23] D. Nuyens and Y. Suzuki. Scaled lattice rules for integration on $\mathbb{R}^d$ achieving higher-order convergence with error analysis in terms of orthogonal projections onto periodic spaces. *Math. Comp.*, 92(339):307–347, 2023.

[NSU21] N. Nagel, M. Schäfer, and T. Ullrich. A new upper bound for sampling numbers. *Found. Comp. Math.*, 2021.

[Nuy07] D. Nuyens. *Fast construction of good lattice rules*. PhD thesis, KU Leuven, 2007.

[NW08] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Vol. 1: Linear information*, volume 6 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2008.

[PC12] K. Poppe and R. Cools. In search for good Chebyshev lattices. In *Monte Carlo and quasi-Monte Carlo methods 2010*, volume 23 of *Springer Proc. Math. Stat.*, pages 639–654. Springer, Heidelberg, 2012.

[Pie87] A. Pietsch. *Eigenvalues and s-numbers*, volume 13 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1987.

[PL13] S. V. Pereverzyev and S. Lu. *Regularization Theory for Ill-Posed Problems. Selected Topics*. DeGruyter, Berlin, Boston, 2013.

[Pla72] R. L. Plackett. Studies in the history of probability and statistics. XXIX. The discovery of the method of least squares. *Biometrika*, 59:239–251, 1972.

[PPST18] G. Plonka, D. Potts, G. Steidl, and M. Tasche. *Numerical Fourier analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, Cham, 2018.

[PS82] C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8(1):43–71, 1982.

[PS22] D. Potts and M. Schmischke. Learning multivariate functions with low-dimensional structures using polynomial bases. *J. Comput. Appl. Math.*, 403:113821, 2022.

[PU22] K. Pozharska and T. Ullrich. A note on sampling recovery of multivariate functions in the uniform norm. *SIAM J. Numer. Anal.*, 60(3):1363–1384, 2022.

[PV15] D. Potts and T. Volkmer. Fast and exact reconstruction of arbitrary multivariate algebraic polynomials in Chebyshev form. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 392–396, 2015.

[PV16] D. Potts and T. Volkmer. Sparse high-dimensional FFT based on rank-1 lattice sampling. *Appl. Comput. Harmon. Anal.*, 41(3):713–748, 2016.

[RFA99] H. Rabitz and O. F. Alis. General foundations of high dimensional model representations. *J. Math. Chem.*, 25:197–233, 1999.

[Ros09] S. Rosset. Bi-level path following for cross validated solution of kernel quantile regression. *J. Mach. Learn. Res.*, 10:2473–2505, 2009.

[RV07] M. Rudelson and R. Vershynin. Sampling from large matrices: an approach through geometric functional analysis. *J. ACM*, 54(4):Art. 21, 19, 2007.

[RV13] M. Rudelson and R. Vershynin. Hanson-wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18(none), 2013.

[RW12] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1-minimization. *J. Approx. Theory*, 164(5):517–533, 2012.

[SC08] I. Steinwart and A. Christmann. *Support vector machines*. Information Science and Statistics. Springer, New York, 2008.

[Sch64] L. Schwartz. Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *J. Analyse Math.*, 13:115–256, 1964.

[Sch22] M. Schmischke. *Interpretable approximation of high-dimensional data based on the ANOVA decomposition*. PhD thesis, Chemnitz University of Technology, 2022.

[Shi00] H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244, 2000.

[SJ94] I. H. Sloan and S. Joe. *Lattice methods for multiple integration*. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1994.

[SNC16] G. Suryanarayana, D. Nuyens, and R. Cools. Reconstruction and collocation of a class of non-periodic functions by sampling along tent-transformed rank-1 lattices. *J. Fourier Anal. Appl.*, 22(1):187–214, 2016.

[Sob50] S. L. Sobolev. Некоторые применения функционального анализа в математической физике. Изд-во Ленинградского гос. ун-та, 1950.

[SS11] D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011.

[SS12] I. Steinwart and C. Scovel. Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSs. *Constr. Approx.*, 35(3):363–417, 2012.

[Sti81] S. M. Stigler. Gauss and the invention of least squares. *Ann. Statist.*, 9(3):465–474, 1981.

[SU09] W. Sickel and T. Ullrich. Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross. *J. Approx. Theory*, 161(2):748–786, 2009.

[SWB08] R. B. Sidje, A. B. Williams, and K. Burrage. Fast generalized cross validation using Krylov subspace methods. *Numer. Algor.*, 47:109–131, 2008.

[Sze21] G. Szegő. Über orthogonale Polynome, die zu einer gegebenen Kurve der komplexen Ebene gehören. *Math. Z.*, 9(3-4):218–270, 1921.

[Sze75] G. Szegő. *Orthogonal polynomials*. American Mathematical Society Colloquium Publications, Vol. XXIII. American Mathematical Society, Providence, R.I., fourth edition, 1975.

[Tem93a] V. N. Temlyakov. *Approximation of periodic functions*. Computational Mathematics and Analysis Series. Nova Science Publishers Inc., Commack, NY, 1993.

[Tem93b] V. N. Temlyakov. On approximate recovery of functions with bounded mixed derivative. *J. Complexity*, 9(1):41–59, 1993. Festschrift for Joseph F. Traub, Part I.

[Tem18] V. N. Temlyakov. The Marcinkiewicz-type discretization theorems. *Constr. Approx.*, 48(2):337–369, 2018.

[Tem21] V. Temlyakov. On optimal recovery in L_2. *J. Complexity*, 65:Paper No. 101545, 11, 2021.

[Tre13] L. N. Trefethen. *Approximation theory and approximation practice*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013.

[Tri92] H. Triebel. *Theory of function spaces. II*, volume 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1992.

[Tri10] H. Triebel. *Theory of function spaces*. Modern Birkhäuser Classics. Birkhäuser Verlag, Basel, 2010.

[Tro12] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.

[TU21] V. N. Temlyakov and T. Ullrich. Bounds on Kolmogorov widths and sampling recovery for classes with small mixed smoothness. *J. Complexity*, 67:101575, 2021.

[TU22] V. N. Temlyakov and T. Ullrich. Approximation of functions with small mixed smoothness in the uniform norm. *J. Approx. Theory*, 277:105718, 2022.

[TW96] M. Tasche and N. Weyrich. Smoothing inversion of Fourier series using generalized cross-validation. *Results Math.*, 29(1-2):183–195, 1996.

[Vap00] V. N. Vapnik. *The nature of statistical learning theory*. Statistics for Engineering and Information Science. Springer-Verlag, New York, second edition, 2000.

[Ver11] R. Vershynin. A simple decoupling inequality in probability theory. *preprint*, 2011.

[Ver18] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer.

[VMK88] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ. *Quantum theory of angular momentum*. World Scientific Publishing Co., Inc., Teaneck, NJ, 1988. Irreducible tensors, spherical harmonics, vector coupling coefficients, $3nj$ symbols, Translated from the Russian.

[Wan23] H. Wang. New error bounds for legendre approximations of differentiable functions. *J. Fourier Anal. Appl.*, 29(4), July 2023.

[Wea04] N. Weaver. The Kadison-Singer problem in discrepancy theory. *Discrete Math.*, 278(1–3):227–239, 2004.

[Wei40] A. Weil. *L'intégration dans les groupes topologiques et ses applications*, volume 869 of *Actualités Scientifiques et Industrielles*. Hermann & Cie, Paris, 1940. [This book has been republished by the author at Princeton, N. J., 1941.].

[Wei07] H. L. Weinert. Efficient computation for Whittaker-Henderson smoothing. *Comp. Stat. & Data Analysis*, 52:959–974, 2007.

[Wen05a] H. Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.

[Wen05b] H. Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.

[Wer00] D. Werner. *Funktionalanalysis*. Springer-Verlag, Berlin, extended edition, 2000.

[WW09] A. G. Werschulz and H. Woźniakowski. Tractability of multivariate approximation over a weighted unanchored Sobolev space. *Constr. Approx.*, 30(3):395–421, 2009.

[Yse04] H. Yserentant. On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives. *Numer. Math.*, 98(4):731–759, 2004.

[Zar07] S. Zaremba. L'équation biharmonique et uni classe remarquable de fonctions fondamentales harmoniques. *Bulletin International de l'Académie des Sciences de Cracovie*, pages 147–196, 1907.

[Zar08] S. Zaremba. Sur le calcul numérique des fonctions demandées dans le problème de dirichlet et le problème hydrodynamoique. *Bulletin International de l'Académie des Sciences de Cracovie*, pages 125–195, 1908.