# SAMPLING INEQUALITIES AND APPLICATIONS 

Dissertation<br>zur Erlangung des Doktorgrades<br>der Mathematisch-Naturwissenschaftlichen Fakultäten<br>der Georg-August-Universität zu Göttingen<br>vorgelegt von<br>Christian Rieger<br>aus Göttingen

Göttingen 2008

D7
Referent: Professor Robert Schaback
Koreferent: Professor Gert Lube
Tag der mündlichen Prüfung: 28.03.2008

## Contents

1 Introduction ..... 1
1.1 Overview ..... 3
2 Sampling Inequalities ..... 7
2.1 Basic Properties ..... 7
2.1.1 General Framework ..... 7
2.1.2 Connection to Markov-Bernstein Inequalities ..... 8
2.1.3 Connection to General Stability Estimates ..... 9
2.1.4 Connection to Lebesgue Constants ..... 9
2.1.5 Connection to Norming Sets ..... 11
2.1.6 Strong Sampling Inequalities ..... 11
2.1.7 Weak Sampling Inequalities ..... 13
2.2 Reconstruction Problems ..... 14
2.2.1 Error Estimates for Reconstruction Problems ..... 15
2.3 Types of Reconstruction Methods ..... 16
2.3.1 Unsymmetric Methods ..... 16
2.3.2 Symmetric Methods ..... 18
2.4 Regularization ..... 19
3 Notation and Basic Facts ..... 23
3.1 Notation ..... 23
3.2 Kernels and Native Spaces ..... 25
3.2.1 Properties of the Native Space ..... 28
3.2.2 $\quad$ Native Spaces on $\mathbb{R}^{d}$ ..... 29
3.2.3 Sobolev Spaces on Bounded Domains ..... 30
3.3 Extension Maps ..... 31
3.3.1 Some Geometrical Terms ..... 32
3.3.2 Polynomial Approximation in Sobolev Spaces ..... 33
3.4 Norming Sets and Polynomial Reproduction ..... 34
4 Infinitely Smooth Functions ..... 37
4.1 Estimates on General Lipschitz Domains ..... 38
4.2 Estimates on Compact Cubes ..... 43
4.3 Oversampling Near the Boundary ..... 45
4.3.1 Polynomial Reproduction ..... 46
4.3.2 Sampling Inequality ..... 49
4.4 Kernels and Native Spaces ..... 50
4.5 Applications to Smoothed Interpolation ..... 53
5 Kernel Based Learning ..... 55
5.1 Regularized Problems in Native Hilbert Spaces ..... 55
5.2 Support Vector Regression ..... 57
5.3 A Sampling Inequality ..... 59
$5.4 \quad \nu$-SVR with Exact Data ..... 60
$5.5 \quad \nu$-SVR with Inexact Data ..... 62
$5.6 \epsilon$-SVR with Exact Data ..... 64
$5.7 \quad \epsilon$-SVR with Inexact Data ..... 65
5.8 Spectral Convergence Orders for Support Vector Machines ..... 66
5.9 Numerical Results ..... 68
5.9.1 Exact Training Data ..... 68
5.9.2 Inexact Data ..... 70
6 Bernstein Inequality ..... 73
6.1 Stationary Weak Data ..... 74
6.2 Norming Set Argument ..... 75
6.3 Extension Operator and Native Space ..... 78
6.4 Bernstein Inequality ..... 83
6.5 Interpolation ..... 86
7 Stationary Weak Data ..... 89
7.1 Local Estimates ..... 91
7.1.1 Norming Set Argument ..... 91
7.1.2 Local Polynomial Approximation ..... 93
7.1.3 Local Sampling Inequality ..... 96
7.2 Global Estimates ..... 98
7.3 Error Estimates for Optimal Recovery ..... 99
7.4 Unsymmetric Weak Recovery ..... 101
8 Galerkin Methods ..... 105
8.1 Model Problem: Elliptic Partial Differential Equations ..... 105
8.2 A Norming Set Approach for Variational Problems ..... 106
8.3 Sampling Inequality for Galerkin Data ..... 109
8.4 Error Estimates for Best Approximation ..... 113
8.5 Sampling Inequality with Polynomials and Galerkin Data ..... 114
9 Discussion and Outlook ..... 119

## Chapter 1

## Introduction

In many practical applications it is necessary to approximate or reconstruct a function as a formula from given strong or weak scattered data. Important examples are domain modeling, surface reconstruction, kernel-based learning or the numerical solution of partial differential equations (for a detailed overview see the recent review [49] and the references therein). There are mainly two ways in which the reconstruction can be done, namely interpolation and approximation. An interpolation process tries to match the given data exactly, which makes sense only if the data is not disturbed by any error. Interpolation processes often have numerical disadvantages because they are usually expensive, in storage and computationally. Therefore, one often asks for a reconstruction that reproduces the given data not exactly, but only approximately. Sometimes it is even useful to allow the accuracy to be a free parameter. One possibility to deal with this kind of reconstruction is provided by sampling inequalities.
Sampling inequalities give a precise formulation of the observation that a differentiable function cannot attain large values anywhere if its derivatives are bounded, and if it produces small data on a sufficiently dense discrete set. Inequalities of this kind can be used to derive a priori error estimates for various regularized approximation problems [49] as they occur for instance in many machine learning algorithms [51] or PDE solvers [46].
In the univariate setting such inequalities are quite easy to obtain [44]. We assume a sufficiently smooth function $f$ on an interval $[a, b]$ and a discrete ordered set of centers $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset[a, b]$ with $a=x_{1}<x_{2}<\cdots<x_{N-1}<x_{N}=b$. In this easy univariate setting we define the fill distance

$$
h:=h_{X,[a, b]}:=\frac{1}{2} \max _{2 \leq j \leq N}\left|x_{j}-x_{j-1}\right|
$$

to be the largest possible distance any point $x \in[a, b]$ has from the set $X$. With this notation we can easily compute for any point $x \in[a, b]$ and the closest point $x_{j} \in X$

$$
\begin{aligned}
f(x) & =f\left(x_{j}\right)+\int_{x_{j}}^{x} f^{\prime}(t) d t, \text { i.e., } \\
|f(x)| & \leq\left|f\left(x_{j}\right)\right|+\sqrt{\left|x-x_{j}\right|} \sqrt{\int_{x_{j}}^{x}\left|f^{\prime}(t)\right|^{2} d t}
\end{aligned}
$$

which yields a typical sampling inequality

$$
\|f\|_{L_{\infty}([a, b])} \leq \sqrt{h}|f|_{W_{2}^{1}[a, b]}+\left\|\left.f\right|_{X}\right\|_{\ell_{\infty}(X)} .
$$

This easy example already reveals the basic phenomenon, i.e., it bounds a weak continuous norm in terms of a strong continuous norm weighted by the fill distance, and some discrete values sampled on the finite set $X$.
Recently several such sampling inequalities for very general functions $u \in W_{p}^{k}(\Omega)$ from certain Sobolev spaces $W_{p}^{k}(\Omega)$ with $1<p<\infty$ and $k>d / p$, or with $p=1$ and $k \geq d$ on a domain $\Omega \subset \mathbb{R}^{d}$ were obtained. As a first step in this direction, Narcowich, Ward and Wendland considered the case of functions with scattered zeros [38]. They found the inequality

$$
|u|_{W_{q}^{m}} \leq C h^{k-m-d\left(\frac{1}{p}-\frac{1}{q}\right)}+|u|_{W_{p}^{k}}
$$

for functions $u \in W_{p}^{k}$ with $k-m>d / p$ and $u(X)=0$, where the discrete set $X$ has sufficiently small fill distance

$$
\begin{equation*}
h:=h_{X, \Omega}:=\sup _{x \in \Omega} \min _{x_{j} \in X}\left\|x-x_{j}\right\|_{2} . \tag{1.0.1}
\end{equation*}
$$

The fill distance may be interpreted geometrically as the radius of the largest ball that is completely contained in $\Omega$, and that does not contain any of the points from the discrete set $X$. It is a useful quantity for the deterministic error analysis in an isotropic setting. The case $h=0$ implies that $X$ is dense in $\Omega$, and therefore convergence is studied for the limit $h \rightarrow 0$, which means that the domain $\Omega$ is nearly uniformly filled with points from $X$. Here and throughout the thesis, we denote by $C$ a generic positive constant.
In the author's Diploma thesis [41] this result was generalized to functions with arbitrary values on scattered locations. A typical sampling inequality then takes the form [67]

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq C\left(h^{k-|\alpha|-d\left(\frac{1}{p}-\frac{1}{q}\right)}+|u|_{W_{p}^{k}(\Omega)}+h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}\right),
$$

where $q \in[1, \infty]$, and $h$ denotes the fill distance of the discrete set $X$ in $\Omega$. A similar result was established by Madych [31], namely

$$
\|u\|_{L_{p}(\Omega)} \leq C\left(h^{k}|u|_{W_{p}^{k}(\Omega)}+h^{d / p}\left\|\left.u\right|_{X}\right\|_{\ell_{p}(X)}\right),
$$

for all $u \in W_{p}^{k}(\Omega)$. These bounds were for instance used [67] to derive optimal algebraic convergence orders for kernel based smoothed interpolation methods.
In this thesis we shall generalize these results in various ways. On the one hand we will derive several new sampling inequalities, e.g., for functions with special smoothness properties, or for more general discrete data. On the other hand we illustrate various applications of sampling inequalities, in particular in the error analysis of manifold reconstruction processes.

### 1.1 Overview

## Background and Notation

In Chapter 2 we provide the general context of sampling inequalities and explain the underlying background. For that, we define more precisely the notion of sampling inequalities and explain their relations to various kinds of reconstruction processes.
Afterwards, in Chapter 3 we set the notation and recall some basic facts, which we need throughout the thesis. For convenience, we frequently use the recent monograph [65] as a reference although most of the cited results have been published earlier, partly by different authors. We mainly stick to the standard notations as used in [65].

## Strong Sampling Inequalities and Applications

In Chapters 4 and 5, we address recovery from strong discrete data. The content of these chapters is partly based on joint work with Barbara Zwicknagl, see [42, 43].
We first derive sampling inequalities for infinitely smooth functions where the convergence orders turn out to vary exponentially with the fill distance $h$. We are handling infinitely smooth functions by normed linear function spaces $\mathcal{H}(\Omega)$ on a domain $\Omega \subset \mathbb{R}^{d}$ that can for a fixed $1 \leq p<\infty$ be uniformly continuously embedded into every classical Sobolev space $W_{p}^{k}(\Omega)$. The embedding constants are allowed to depend on the smoothness $k$. Details can be found in equation (2.1.6). There are various examples of spaces with this property, e.g., Sobolev spaces of infinite order as they occur for instance in the study of partial differential equations of infinite orders [1], or reproducing kernel Hilbert spaces of Gaussians and inverse multiquadrics.
As a typical result we obtain inequalities of the form

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq e^{C \log (C h) / \sqrt{h}}\|u\|_{\mathcal{H}(\Omega)}+C h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)} .
$$

As a corollary we obtain that the technique presented here reproduces the well-known error estimates for the standard interpolation problem in the native Hilbert space of the inverse multiquadric and Gaussian kernels.
In Chapter 5 we provide a theoretical framework to derive deterministic error bounds for some popular support vector machines. Support vector (SV) machines and related kernelbased algorithms are modern learning systems motivated by results of statistical learning theory [57]. The concept of SV machines is to provide a prediction function that is accurate on the given training data, and that is sparse in the sense that it can be written in terms of a typically small subset [50] of all examples, called the support vectors. Therefore, SV regression and classification algorithms are closely related to regularized problems from classical approximation theory [23], and techniques from functional analysis were applied to derive probabilistic error bounds for SV regression [17].
We show exemplarily how sampling inequalities can be used to bound the worst-case generalization error for the $\nu$ - and the $\epsilon$-regression, as introduced by Schölkopf, Williamson and Bartlett [52], for exact as well as for inexact given training-data. Here we do not make any statistical assumptions on the inaccuracy of the training data. In contrast to the literature, our error bounds explicitly depend on the pointwise noise in the data. Thus they can be
used for any subsequent probabilistic analysis modelling certain assumptions on the noise distribution.
In Section 6 we prove Bernstein inequalities. For that we consider for appropriate radial basis functions $\Phi$ the linear spaces $V_{\Phi, X}:=\left\{\Phi\left(\cdot-x_{j}\right): x_{j} \in X\right\}$. Since $V_{\Phi, X}$ is finite dimensional, all norms on it are equivalent. The Bernstein inequalities quantify the equivalence constants between the $W_{2}^{\tau}$ - and the $L_{2}$-norm in terms of the geometry of $X$. They take the form

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C q_{X}^{-\tau}\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}(\Omega)},
$$

where

$$
q_{X}:=\frac{1}{2} \min _{\substack{x_{i}, x_{j} \in X \\ x_{i} \neq x_{j}}}\left\|x_{i}-x_{j}\right\|_{2}
$$

denotes the separation distance of the discrete set $X$. Therefore, our results generalize estimates from [39] to bounded domains $\Omega$. These technical estimates are very useful to prove stability estimates for unsymmetric methods, as presented in the subsequent chapters.

## Weak Sampling Inequalities and Applications

To work towards an analysis of meshless local Petrov-Galerkin (MLPG) methods [4, 5, 6] we shall treat in Chapter 7 sampling inequalities for weak data in the sense of [47]. The main drawback of the sampling inequalities considered so far is that one assumes functions that allow continuous point evaluations. Especially for many practical applications in the field of partial differential equations (PDE's) this assumption is too restrictive. Following [47] we shall use convolution-type data of the form

$$
\begin{equation*}
\lambda_{j}(u)=\int_{\Omega} K\left(x-x_{j}\right) u(x) d x \tag{1.1.1}
\end{equation*}
$$

where $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ is a discrete set of points and $K: \mathbb{R}^{d} \rightarrow \mathbb{R}$ is called test kernel. In contrast to [47] we shall use stationary data, i.e., the support of the test kernel shall be scaled with the fill distance or mesh-norm of $X$ in $\Omega$ as defined in equation (1.0.1). The term stationary is a usual notation in the kernel based meshless literature. This approach generalizes in a way the finite volume method from [64]. Our main result is a sampling inequality involving weak convolution-type data. We shall fit our work in the general framework of recent research as can, e.g., be found in [47, 45, 46].
In the last part we shall analyze another kind of weak data, which arises naturally from the numerical study of partial differential equations. We start with a partial differential equation in its weak formulation

$$
\begin{equation*}
\text { find } u \in W: a(u, v)=F(v) \quad \text { for all } v \in W, \tag{1.1.2}
\end{equation*}
$$

where $W=W_{2}^{m}(\Omega)$ is typically a Sobolev space, $a(\cdot, \cdot)$ is a bilinear form and $F$ is a linear form. We discretize the problem in a Petrov-Galerkin style [11] and consider the finite
dimensional problem

$$
\text { find } u \in V_{\Phi, X}: a(u, v)=F(v) \text { for all } v \in V_{\Phi, X},
$$

where again for a kernel $\Phi$ we use the notation

$$
V_{\Phi, X}:=\operatorname{span}\left\{\Phi\left(\cdot-x_{j}\right): x_{j} \in X\right\}
$$

with an orthonormal basis $\left\{\phi_{j}\right\}_{j=1, \ldots, N}$. Under weak assumptions on the bilinear form $a(\cdot, \cdot)$ we derive a sampling inequality of the form

$$
\|u\|_{L_{2}(\Omega)} \leq C h^{m}\|u\|_{W_{2}^{m}(\Omega)}+C_{X} \sum_{j=1}^{N} a\left(u, \phi_{j}\right)^{2}
$$

for all $u \in W_{2}^{m}(\Omega)$ and a specific $m>0$. This bound is used to get an error estimate for the best approximation $s_{f} \in V_{\Phi, X}$ to a function $f \in W_{2}^{1}(\Omega)$. Unfortunately this best approximation is numerically unavailable. Therefore we present an approximation strategy by testing for which we prove a convergence rate. For the finite dimensional approximation we use two approaches, kernel-based and polynomial spaces.
Finally we give a short discussion and outlook.

## Chapter 2

## Sampling Inequalities

In this chapter we shall explain some general features of sampling inequalities and give an overview of some applications. To avoid technical details, we always assume $\Omega \subset \mathbb{R}^{d}$ to be a nicely bounded domain. By $X$ we mean a discrete set $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$. The more precise technical details are given in the following chapters.

### 2.1 Basic Properties

Sampling inequalities were developed to quantify the following observation:
If a sufficiently smooth function is small on scattered points, and if its derivatives are bounded, it must be small in the whole domain.

As a first step in this direction, Narcowich, Ward and Wendland considered the case of functions with scattered zeros [38]. In the author's Diploma thesis [41] this result was generalized to functions with arbitrary values on scattered locations. A similar result has been found by Madych [31]. Another recent extension of these results and their applications in spline smoothing can be found in [2]. In this section this concept is described in a generalized form, its connections to other techniques are explained, and various applications are presented. We start from a rather abstract viewpoint and show the connections between sampling inequalities and certain other techniques. Then we look at special classes of sampling inequalities, and finally we give a general overview of how sampling inequalities can be applied. Our special results on sampling inequalities and applications start from Chapter 4 on.

### 2.1.1 General Framework

We assume a normed linear function space $\mathcal{F}$ with norms $\|\cdot\|_{S}$ and $\|\cdot\|_{W}$, where $\|\cdot\|_{S}$ is stronger than $\|\cdot\|_{W}$, i.e.,

$$
\begin{equation*}
\|f\|_{W} \leq C\|f\|_{S} \quad \text { for all } f \in \mathcal{F} . \tag{2.1.1}
\end{equation*}
$$

Here and in the following, $C$ denotes a generic positive constant. Furthermore, we consider a set of $N$ linear independent functionals $\Lambda=\left\{\lambda_{1}, \ldots, \lambda_{N}\right\}$ from the dual space $\mathcal{F}^{*}$ with
respect to $\|\cdot\|_{S}$. These functionals are used to sample a function from $\mathcal{F}$ via the continuous and linear sampling operator

$$
\begin{aligned}
S_{\Lambda} & : \mathcal{F} \rightarrow \mathbb{R}^{N}, \\
& f
\end{aligned} \mapsto\left(\lambda_{1}(f), \ldots, \lambda_{N}(f)\right)^{T} .
$$

Then a rather general form of a sampling inequality is

$$
\begin{equation*}
\|f\|_{W} \leq C\left(h_{\Lambda}^{\sigma}\|f\|_{S}+C\left(h_{\Lambda}\right)\left\|S_{\Lambda}(f)\right\|_{\mathbb{R}^{N}}\right) \quad \text { for all } f \in \mathcal{F} . \tag{2.1.2}
\end{equation*}
$$

Or, in case that $\|\cdot\|_{S}$ is only a semi-norm, which is denoted by $|\cdot|_{S}$, we get the analogue result

$$
\begin{equation*}
\|f\|_{W} \leq C\left(h_{\Lambda}^{\sigma}|f|_{S}+C\left(h_{\Lambda}\right)\left\|S_{\Lambda}(f)\right\|_{\mathbb{R}^{N}}\right) \quad \text { for all } f \in \mathcal{F} . \tag{2.1.3}
\end{equation*}
$$

The term $h_{\Lambda}$ denotes some discretization parameter which should be small, i.e., $h_{\Lambda} \xrightarrow{N \rightarrow \infty} 0$. The exponent $\sigma>0$ will be called sampling order. This means that there is a small factor in front of the term with the strong continuous norm and a possibly large term in front of the term with the discrete norm. If the sampling operator contains only point evaluations, i.e., evaluations of $f$ and its derivatives on certain finite point sets, we speak of strong sampling inequalities. If some other functionals are involved, which may be well defined even if point evaluation is not continuous, we speak of weak sampling inequalities.

### 2.1.2 Connection to Markov-Bernstein Inequalities

By $\mathcal{P}_{S} \subset \mathcal{F}$, we will denote the kernel of the semi-norm $|\cdot|_{S}$. If we insert an element $p \in \mathcal{P}_{S}$ from this kernel into the sampling inequality, we obtain

$$
\|p\|_{W} \leq C\left(h_{\Lambda}^{\sigma}|p|_{S}+C\left(h_{\Lambda}\right)\left\|S_{\Lambda}(p)\right\|_{\mathbb{R}^{N}}\right)=C\left(h_{\Lambda}\right)\left\|S_{\Lambda}(p)\right\|_{\mathbb{R}^{N}} .
$$

This means that we can bound a continuous norm by a discrete norm on the data. Inequalities of this kind will be called stability bounds. Such bounds can hold only if $\mathcal{P}_{S}$ is finite-dimensional, but this is just a necessary condition. If $\mathcal{P}_{S}$ is a space of polynomials, these estimates imply Markov-Bernstein inequalities [9]. Let us explain this in some more detail. Let us assume for instance that the sampling operator consists only of Lagrange data. Furthermore, we assume that $\|\cdot\|_{W}=|\cdot|_{W_{\infty}^{1}(\Omega)}$, and that $\|\cdot\|_{S}=|\cdot|_{W_{\infty}^{k}(\Omega)}$ are classical Sobolev-seminorms. This yields for all $1 \leq \ell \leq d$

$$
\begin{align*}
\left\|\partial_{\ell} p\right\|_{L_{\infty}(\Omega)} & \leq C\left(h_{\Lambda}\right) \max _{x_{j} \in X}\left|p\left(x_{j}\right)\right| \\
& \leq C\left(h_{\Lambda}\right)\|p\|_{L_{\infty}(\Omega)}, \text { for all } p \in \pi_{k-1}(\Omega), \tag{2.1.4}
\end{align*}
$$

where $\partial_{\ell}$ denotes the partial derivative in direction of the $\ell$-th coordinate. This is a special case of classical Markov-Bernstein-inequalities [9], and we see that the proofs for sampling inequalities should contain those classical estimates.

### 2.1.3 Connection to General Stability Estimates

In general, stability estimates do not scale with the discretization parameter $h_{\Lambda}$, but with an other discretization parameter $q_{\Lambda} \leq h_{\Lambda}$. In the meshless literature $q_{X}$ expresses how irregularly distributed the data is. Roughly speaking, $q_{\Lambda}$ measures the shortest distance between two data points. In the finite element literature, it is the radius of the largest ball inside an element. It is intuitively clear that both quantities are responsible for stability. For reasons that will be apparent later we will introduce two different sets of functionals $\Lambda$ and $\tilde{\Lambda}$. We assume that $\mathcal{R} \subset \mathcal{F}$ is a finite dimensional trial space with an inverse inequality of the form

$$
\begin{equation*}
|r|_{S} \leq C q_{\tilde{\Lambda}}^{-\tau}\|r\|_{W} \quad \text { for all } r \in \mathcal{R} . \tag{2.1.5}
\end{equation*}
$$

Inequalities of this kind will be called inverse inequalities since they bound a stronger norm by a weaker one. In many cases, we are able to couple these discretization parameters via

$$
C h_{\Lambda}^{\sigma} q_{\tilde{\Lambda}}^{-\tau} \leq \frac{1}{2}
$$

In the special case $\sigma=\tau$ this coupling allows for an interpretation in the language of finite elements. In this case the coupling just means that one considers isotropic elements. In the meshless context, there are two possibilities to achieve such a coupling, either by coupling $\sigma$ and $\tau$, or by coupling $\Lambda$ and $\tilde{\Lambda}$. We may combine the estimates (2.1.3) and (2.1.5) to get a stability estimate of the form

$$
\|r\|_{W} \leq 2 C\left(h_{\Lambda}, q_{\Lambda}\right)\left\|S_{\Lambda}(r)\right\|_{\mathbb{R}^{N}} \quad \text { for all } r \in \mathcal{R}
$$

In this general form, stability estimates are crucial for all discretization processes defined via trial spaces $\mathcal{R}$ and a sampling operator $S_{\Lambda}$. They imply invertibility of the sampling operator on $\mathcal{R}$ and guarantee that small errors in the sampled data imply small deviations in the norm on the trial space. The upshot of the above argument is that stability bounds follow from a sampling inequality and an inverse inequality.

Guideline: Sampling inequalities and inverse inequalities imply stability bounds.

### 2.1.4 Connection to Lebesgue Constants

Furthermore, sampling inequalities are closely related to Lebesgue constants. See [15] for more details in the setting of radial basis functions. Nevertheless, we shall briefly point out how Lebesgue constants are related to the notion of stability. Let us assume some interpolation space $\mathcal{R}$, i.e., a space which allows unique interpolation, which we call trial space for reasons that become apparent later. Since $\mathcal{R}$ is an interpolation space, we can build the generalized cardinal interpolants $u_{\lambda_{i}}$ from $\mathcal{R}$, which fulfill $\lambda_{j}\left(u_{\lambda_{i}}\right)=\delta_{i, j}$, where $\delta_{i, j}$ denotes the Kronecker symbol. Now we can directly form an interpolant to a function $f$, namely $I_{\Lambda}(f)(\cdot)=\sum_{j=1}^{N} \lambda_{j}(f) u_{\lambda_{j}}(\cdot)$. Then we can derive a stability estimate of the
form

$$
\begin{aligned}
\left\|I_{\Lambda}(f)\right\|_{W} & =\left\|\sum_{j=1}^{N} \lambda_{j}(f) u_{\lambda_{j}}\right\|_{W} \leq \sum_{j=1}^{N}\left|\lambda_{j}(f)\right|\left\|u_{\lambda_{j}}\right\|_{W} \\
& \leq \max _{1 \leq j \leq N}\left|\lambda_{j}(f)\right| \sum_{j=1}^{N}\left\|u_{\lambda_{j}}\right\|_{W}=\|\lambda(f)\|_{\ell_{\infty}(X)} L_{\mathcal{R}, \Lambda},
\end{aligned}
$$

where $L_{\mathcal{R}, \Lambda}$ is called Lebesgue constant. Evaluation stability is another concept of stability that is closely related to Lebesgue constants. Suppose the data $S_{\Lambda}(f)$ is corrupted by some relative errors $E=\left(\epsilon_{1}, \ldots, \epsilon_{N}\right)^{T}$ with $\max _{1 \leq j \leq N}\left|\epsilon_{j}\right|=\epsilon$. Then,

$$
\begin{aligned}
& \left\|\sum_{j=1}^{N} \lambda_{j}(f) u_{\lambda_{j}}-\sum_{j=1}^{N} \lambda_{j}(f)\left(1+\epsilon_{j}\right) u_{\lambda_{j}}\right\|_{L_{\infty}(\Omega)}=\left\|\sum_{j=1}^{N} \epsilon_{j} \lambda_{j}(f) u_{\lambda_{j}}\right\|_{L_{\infty}(\Omega)} \\
\leq & \epsilon \cdot \max _{1 \leq j \leq N}\left|\lambda_{j}(f)\right| \sup _{x \in \Omega} \sum_{j=1}^{N}\left|u_{\lambda_{j}}(x)\right| \leq \epsilon \cdot \max _{1 \leq j \leq N}\left|\lambda_{j}(f)\right| \sum_{j=1}^{N}\left\|u_{\lambda_{j}}(x)\right\|_{L_{\infty}(\Omega)} .
\end{aligned}
$$

This shows that the evaluation of the interpolation process is very stable in the basis of generalized cardinal functions. In addition, this short detour indicates that many concepts of univariate polynomial interpolation can be carried over to kernel based methods [35].
Now we explain how sampling inequalities lead to bounds on the Lebesgue constants. If we measure the discrete term in the $\ell_{\infty}\left(\mathbb{R}^{N}\right)$ norm, i.e.,

$$
\left\|S_{\Lambda}(u)\right\|_{\mathbb{R}^{N}}=\max _{1 \leq i \leq N}\left|\lambda_{i}(u)\right|
$$

we immediately get $\left\|S_{\Lambda}\left(u_{\lambda_{i}}\right)\right\|_{\mathbb{R}^{N}}=1$ for all $1 \leq i \leq N$. Applying the sampling inequality (2.1.3) hence yields

$$
\left\|u_{\lambda_{i}}\right\|_{W} \leq C\left(h_{\Lambda}^{\sigma}\left|u_{\lambda_{i}}\right|_{S}+C\left(h_{\Lambda}\right)\left\|S_{\Lambda}\left(u_{\lambda_{i}}\right)\right\|_{\mathbb{R}^{N}}\right) \leq C\left(h_{\Lambda}^{\sigma}\left|u_{\lambda_{i}}\right|_{S}+C\left(h_{\Lambda}\right)\right)
$$

If we now assume also a stability inequality of the form $\left|u_{\lambda_{i}}\right|_{S} \leq\left\|f_{i}\right\|_{S}$ for all $f_{i} \in S$ fulfilling the interpolation condition $\lambda_{j}\left(f_{i}\right)=\delta_{i, j}$ as well, we get

$$
\left\|u_{\lambda_{i}}\right\|_{W} \leq C h_{\Lambda}^{\sigma}\left\|f_{i}\right\|_{S}+C\left(h_{\Lambda}\right) .
$$

This is well known in the kernel based interpolation case, see an example below. If we now assume that there is an $f_{i}$ fulfilling the interpolation condition and satisfying the estimate $\left\|f_{i}\right\|_{S} \leq \tilde{C}\left(h_{\Lambda}\right)$, we can deduce $\left\|u_{\lambda_{i}}\right\|_{W} \leq \tilde{C}\left(h_{\Lambda}\right) C\left(h_{\Lambda}\right)$. Thus, the Lebesgue constant is bounded by

$$
\sum_{i=1}^{N}\left\|u_{\lambda_{i}}\right\|_{W} \leq N \tilde{C}\left(h_{\Lambda}\right) C\left(h_{\Lambda}\right)
$$

We spell out all these constants in the special situation of [15]: The authors consider as trial space $\mathcal{R}=\operatorname{span}\left\{\Phi\left(\cdot-x_{j}\right): x_{j} \in X \subset \Omega\right\}$ the translates of a single radial basis function with Fourier transform $\hat{\Phi}(\omega) \approx\left(1+\|\omega\|_{2}^{2}\right)^{-\tau}$. The functionals $\lambda_{j}=\delta_{x_{j}}$ are simply point evaluations at $X$, where $X$ is a grid with $N$ points. In this special case, they obtain: $\tilde{C}\left(h_{\Lambda}\right)=N^{\frac{\tau}{d}}$ and $C\left(h_{\Lambda}\right)=N^{-\frac{T}{d}}$. This, together with the Cauchy-Schwarz inequality, finally yields $\sum_{i=1}^{N}\left\|u_{\lambda_{i}}\right\|_{W} \leq C \sqrt{N}$.

### 2.1.5 Connection to Norming Sets

As these examples indicate, that sampling inequalities imply that the data $S_{\Lambda}(p)$ contains already enough information about $p \in \mathcal{P}_{S}$. This leads to the general concept of norming sets [26]. We shall briefly explain this concept, but we will be more precise in Chapter 3. $\Lambda$ is called a norming set for $\mathcal{P}_{S}$ if the sampling operator

$$
\begin{aligned}
\left.S_{\Lambda}\right|_{\mathcal{P}_{S}}: \quad \mathcal{P}_{S} & \left.\rightarrow S_{\Lambda}\right|_{\mathcal{P}_{S}}\left(\mathcal{P}_{S}\right) \subset \mathbb{R}^{N} \\
v & \left.\mapsto S_{\Lambda}\right|_{\mathcal{P}_{S}}(v)=(\lambda(v))_{\lambda \in \Lambda}
\end{aligned}
$$

is injective. Then it is well-known (see Chapter 3) that we can introduce another norm on $\mathcal{P}_{S}$ by $\left\|\left.S_{\Lambda}\right|_{\mathcal{P}_{S}}(\cdot)\right\|_{\mathbb{R}^{N}}$. Furthermore, by application of Theorem3.4.2, we get a reproduction formula on $\mathcal{P}_{S}$ of the form

$$
\Psi(p)=\sum_{j=1}^{N} a_{j}(\Psi) \lambda_{j}(p) \quad \text { for all } p \in \mathcal{P}_{S} \text { and all } \Psi \in \mathcal{P}_{S}^{*}
$$

where we can bound the term

$$
\sum_{j=1}^{N}\left|a_{j}(\lambda)\right| \leq C
$$

i.e., an inequality of the form (2.1.4). This explains the terminology norming set. This detour indicates that norming sets are crucial in the proofs of sampling inequalities.

### 2.1.6 Strong Sampling Inequalities

If we now set $\|\cdot\|_{S}=\|\cdot\|_{W_{p}^{k}(\Omega)}$ and $\|\cdot\|_{W}=\|\cdot\|_{W_{q}^{m}(\Omega)}$, the condition (2.1.1) simply becomes

$$
W_{q}^{k}(\Omega) \hookrightarrow W_{p}^{m}(\Omega) \quad \text { i.e., } k \geq m
$$

If the functions are regular enough to allow for point evaluation, i.e., $k>d / p$, or $k \geq d$ for $d=1$, we can state a basic form of a sampling inequality, which can be found already in [67]. To do so, we define a sampling operator

$$
\begin{array}{rll}
S_{X} & : W_{p}^{k}(\Omega) \rightarrow \mathbb{R}^{N} \\
& u & \mapsto\left(u\left(x_{1}\right), \ldots, u\left(x_{N}\right)\right)^{T}
\end{array}
$$

for some points $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$, i.e., $\lambda_{i}=\delta_{x_{i}}$. The discretization will be measured in terms of the fill distance as given in equation (1.0.1). Because of the geometric interpretation we are interested in the limit $h_{X, \Omega} \rightarrow 0$. We get the bound

$$
\|u\|_{L_{2}(\Omega)} \leq C h_{X, \Omega}^{\tau}\|u\|_{W_{2}^{\tau}(\Omega)}+C\left\|S_{X}\right\|_{\ell_{\infty}(X)}
$$

for all $u \in W_{p}^{k}(\Omega)$. We already see that the sampling rate depends only on the "smoothness difference" of the two continuous norms involved.

## Univariate Sampling Inequalities

As an easy example we consider the univariate case [44]. We assume a sufficiently smooth function $f$ on an interval $[a, b]$ and a discrete ordered set $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset[a, b]$. We can easily compute for any point $x \in[a, b]$ and the closest point $x_{j} \in X$

$$
\begin{aligned}
f(x) & =f\left(x_{j}\right)+\int_{x_{j}}^{x} f^{\prime}(t) d t \\
\Rightarrow \quad|f(x)| & \leq\left|f\left(x_{j}\right)\right|+\sqrt{\left|x-x_{j}\right|} \sqrt{\int_{x_{j}}^{x}\left|f^{\prime}(t)\right|^{2} d t}
\end{aligned}
$$

which yields a typical form of a sampling inequality,

$$
\|f\|_{L_{\infty}[a, b]} \leq \sqrt{h}|f|_{W_{2}^{1}[a, b]}+\left\|\left.f\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

If $f$ is the error-function of some discretized numerical process, we can conclude that the $L_{\infty}$-error is small, provided that the discrete errors are small, and that the $|\cdot|_{W_{2}^{1}}$-seminorm of the error is bounded independent of the discretization. Now we will turn to the multivariate case.

## Multivariate Sampling Inequalities

Recently several sampling inequalities for general functions $u \in W_{p}^{k}(\Omega)$ from certain Sobolev spaces with $1<p<\infty$ and $k>d / p$, or $p=1$ and $k \geq d$ on a domain $\Omega \subset \mathbb{R}^{d}$ were obtained. They usually take the following form [67].

Theorem 2.1.1 Let $k$ be a positive real number and $\alpha \in \mathbb{N}_{0}^{d}$ a multi-index with $\lfloor k-|\alpha|\rfloor>$ $\frac{d}{2}$. Then there exists a positive constant $C>0$ such that for all discrete sets $X \subset \Omega$ with sufficiently small fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor k-1\rfloor^{-2}$ the inequality

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq C\left(h^{k-|\alpha|-\max \left\{d\left(\frac{1}{2}-\frac{1}{q}\right), 0\right\}}|u|_{W_{2}^{k}(\Omega)}+h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}\right)
$$

holds for all $u \in W_{2}^{k}(\Omega)$, with $1 \leq q \leq \infty$, where the explicit expression for $C_{\Omega}$ can be found in equation (5.3.1).

In [31] there is an analogous result. In this case the estimate takes the form

$$
\|u\|_{L_{p}(\Omega)} \leq C\left(h^{k}|u|_{W_{p}^{k}(\Omega)}+h^{d / p}\left\|\left.u\right|_{X}\right\|_{\ell_{p}(X)}\right)
$$

for all $u \in W_{p}^{k}(\Omega)$. Other extensions both concerning the range of the smoothness parameter and unbounded domains are done in [2, 3]. Here, the fill distance $h$ and the smoothness $k$ have to be coupled appropriately. As we saw above, these kinds of inequalities are closely related to Markov-Bernstein inequalities [9]. If we for instance insert a polynomial $p \in \pi_{k-1}(\Omega)$ in the first inequality with $\alpha=0$, we directly see

$$
\|p\|_{L_{\infty}(\Omega)} \leq C\left\|\left.p\right|_{X}\right\|_{\ell_{\infty}(X)} .
$$

In this case the inherent coupling of $h$ and $k$ implies that $\pi_{k-1}(\Omega)$ is a Haar space over $X$. Hence sampling inequalities in Sobolev spaces always imply stable polynomial reproductions.
In this thesis (Chapter 4) we shall derive sampling inequalities with exponential sampling order. These inequalities take for instance the form

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq e^{C \log (C h) / \sqrt{h}}\|u\|_{\mathcal{H}(\Omega)}+C h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)},
$$

where the space $\mathcal{H}(\Omega)$ consists of infinitely smooth functions. This means that for a fixed $p \in[1, \infty)$ and all $k \in \mathbb{N}$ we assume that there are embedding operators $I_{k}^{(p)}$ and constants $E(k)$ such that

$$
\begin{align*}
& I_{k}^{(p)}: \mathcal{H}(\Omega) \rightarrow W_{p}^{k}(\Omega) \quad \text { with } \\
& \left\|I_{k}^{(p)}\right\|_{\left\{\mathcal{H}(\Omega) \rightarrow W_{p}^{k}(\Omega)\right\}} \leq E(k) \quad \text { for all } k \in \mathbb{N}_{0} \tag{2.1.6}
\end{align*}
$$

The native Hilbert spaces of Gaussian kernels are important examples of such spaces.

### 2.1.7 Weak Sampling Inequalities

Now we will focus on weak sampling operators. We consider a set of functionals

$$
\Lambda_{S}=\left\{\lambda_{1}, \ldots, \lambda_{N_{s}}\right\} \subset\left(W_{2}^{\tau}(\Omega)\right)^{*} .
$$

These functionals define a weak sampling operator

$$
S_{\Lambda}:=\left(\begin{array}{c}
\lambda_{1} \\
\vdots \\
\lambda_{N_{S}}
\end{array}\right)
$$

We deal with the natural question whether a sampling inequality of the form

$$
\|u\|_{L_{2}(\Omega)} \leq C h_{X, \Omega}^{\tau}|u|_{W_{2}^{\tau}(\Omega)}+C\left\|S_{\Lambda} u\right\|_{\ell_{\infty}\left(\mathbb{R}^{N}\right)}
$$

holds. This can be true only if the functionals $\lambda_{j}$ contain enough information on the function $u$. To be precise we can again insert a polynomial of degree less than $\tau$ to derive

$$
\|p\|_{L_{2}(\Omega)} \leq C\left\|S_{\Lambda} p\right\|_{\ell_{\infty}\left(\mathbb{R}^{N}\right)} .
$$

This shows that the functionals $\Lambda$ have to form a norming set (for a definition, see Section 3.4) for the polynomials of degree less than $\tau$. We will present two examples of such functionals, which are of current research interest.

## Weak Convolution-Type Data

Following [47] we consider weak convolution-type data of the form

$$
\begin{equation*}
\lambda_{j}(u)=\int_{\Omega} K\left(x-x_{j}\right) u(x) d x \tag{2.1.7}
\end{equation*}
$$

where $X=\left\{x_{1}, \ldots, x_{n}\right\} \subset \Omega$ is a discrete set of points, and $K: \mathbb{R}^{d} \rightarrow \mathbb{R}$ is a test kernel. Schaback investigates in [47] the case of non-stationary data, i.e., a fixed test kernel. In contrast, we shall use stationary data, i.e., the support of the test kernel shall be scaled with the fill distance $h$ of $X$ in $\Omega$. This generalizes the recovery of functions under construction of local volume integrals, as needed in finite volume methods, introduced by Sonar (partly jointly with Iske) in a series of papers [25,53, 54, 55], where the analytical background was provided by Wendland [64]. They consider the case where the kernel $K$ is replaced by a characteristic function.
Our main result is a sampling inequality for the described convolution-type sampling. This is a crucial ingredient for the analysis of the unsymmetric meshless local Petrov-Galerkin schemes (MLPG) by Atluri and Zhu [6, 7, 4, 5], and it fits into the general framework of recent research on the theoretical background of these methods, as can be found in [47, 45, 46].

## Galerkin Methods

In Chapter 8 we shall analyze another kind of weak data, which arises naturally in the study of partial differential equations. Namely, we consider a partial differential equation in its weak formulation

$$
\begin{equation*}
\text { find } u \in W: a(u, v)=F(v) \quad \text { for all } v \in X \tag{2.1.8}
\end{equation*}
$$

where $W=W_{2}^{m}(\Omega)$ is typically a Sobolev space, $a: W \times W \rightarrow \mathbb{R}$ is a bilinear form, and $F \in W^{*}$ is a linear functional. To solve the problem (2.1.8) approximately we use a Ritz-Galerkin approach [11, (2.5.7)] and consider the finite dimensional problem

$$
\text { find } u \in V_{\Phi, X}: a(u, v)=F(v) \text { for all } v \in V_{\Phi, X}
$$

where

$$
V_{\Phi, X}:=\operatorname{span}\left\{\Phi\left(\cdot-x_{j}\right): x_{j} \in X\right\}
$$

with an orthonormal basis $\left\{\phi_{j}\right\}_{j=1, \ldots, N}$. Under weak assumptions on the bilinear form $a(\cdot, \cdot)$, which are standard in the theory of finite element methods [11], we derive a sampling inequality of the form

$$
\|u\|_{L_{2}(\Omega)} \leq C h^{m}\|u\|_{W_{2}^{m}(\Omega)}+C_{X} \sum_{j=1}^{N} a\left(u, \phi_{j}\right)^{2}
$$

for all $u \in W_{2}^{m}(\Omega)$ and a specific $m>0$.

### 2.2 Reconstruction Problems

In this section we turn our attention to very general reconstruction problems, which provide one of the most important applications of sampling inequalities. We explain how sampling
inequalities can be used to measure the quality of almost every numerical recovery process where we try to recover an unknown function $f \in \mathcal{F}$ from its given data. Let

$$
\Lambda_{S}=\left\{\lambda_{1}, \ldots, \lambda_{N_{S}}\right\} \subset \mathcal{F}^{*},
$$

be a set of functionals, which will be used for testing. These functionals define a sampling operator via

$$
S_{\Lambda_{S}}:=\left(\begin{array}{c}
\lambda_{1} \\
\vdots \\
\lambda_{N_{S}}
\end{array}\right)
$$

which we need to define sampling inequalities. We consider data given by the sampling operator, namely $S_{\Lambda_{S}}(f)$.
Now we have to choose a finite dimensional space $\mathcal{R}$, from which we would like to pick a reconstruction. This space is called trial space. It is at the moment completely independent of the given data. We shall therefore carefully distinguish between the test and the trial side. The test side consists of the given functionals $\Lambda_{S}$. The trial side consists of a finite dimensional trial space $\mathcal{R}$, which is used to generate an approximate solution to the fitting problem. The approximation properties of the reconstruction process depend only on the trial side. In this thesis, we will mainly restrict ourselves to kernel-based methods. This means that the trial functions are generated by a single kernel $K(\cdot, \cdot)$, i.e.,

$$
\mathcal{R}:=\operatorname{span}\left\{K\left(x_{j}, \cdot\right): x_{j} \in X_{R}\right\},
$$

where the set $X_{R}:=\left\{x_{1}, \ldots, x_{N_{R}}\right\} \subset \Omega$ is called the set of trial points. Sampling inequalities can be used to quantify the approximation quality of $\mathcal{R}$. Note that we use $\mathcal{R}$ for the $\mathbf{t}$ Rial side and $S$ for the the $\mathbf{S t}$ side.
The main application of the sampling inequalities are, however, on the test side. Sampling inequalities are used in this framework to show to what extent small discrete residuals $\left\|S_{\Lambda_{S}}\left(R_{\mathcal{R}} f-f\right)\right\|_{\ell_{p}(X)}$, where $R_{\mathcal{R}} f \in \mathcal{R}$ is a reconstruction of $f$, imply global small errors. This leads to conditions on the quantity of given data.
Up to now we have treated test and trial side completely independent of each other. To prove convergence results for recovery methods, both sides have to be coupled. However, there is some general guideline for error estimates, which we explain in the next sections.

### 2.2.1 Error Estimates for Reconstruction Problems

Sampling inequalities can be applied in various ways. We turn our main attention on deterministic a priori error bounds for numerical processes $\mathcal{I}$. Suppose we want to reconstruct an unknown function $f \in \mathcal{F}$ from its data $S_{\Lambda}(f)$. Then we need to find a function $\mathcal{I}_{\Lambda}(f) \in \mathcal{R}$ (called an approximant to $f$ ) from a large trial space $\mathcal{R} \subset \mathcal{F}$, which fulfills the following two properties

$$
\begin{align*}
\left\|\mathcal{I}_{\Lambda}(f)\right\|_{S} & \leq C\|f\|_{S} \\
\left\|S_{\Lambda}\left(\mathcal{I}_{\Lambda}(f)-f\right)\right\|_{\mathbb{R}^{N}} & \leq g(f, N) \xrightarrow{N \rightarrow \infty} 0 . \tag{2.2.1}
\end{align*}
$$

These conditions are fulfilled in many kernel based methods as will be seen later. They can be seen as stability and consistency properties which are the main ingredient of nearly any convergence proof in numerical analysis. If we have a sampling inequality of the form (2.1.2), and if the assumptions (2.2.1) are satisfied, we get an error estimate of the form

$$
\left\|f-\mathcal{I}_{\Lambda}(f)\right\|_{W} \leq C h_{\Lambda}^{\sigma}\|f\|_{S}+C\left(h_{\Lambda}\right) g(f, N) .
$$

This means that the error is controlled easily, and convergence orders are obtained. Especially important is the situation, where the residual $C\left(h_{\Lambda}\right) g(f, N)$ is bounded by $C\left(h_{\Lambda}\right) g(f, N) \leq h_{\Lambda}^{\tilde{\sigma}}\|f\|_{S}$. This yields

$$
\left\|f-\mathcal{I}_{\Lambda}(f)\right\|_{W} \leq C\left(h_{\Lambda}^{\sigma}+h_{\Lambda}^{\tilde{\sigma}}\right)\|f\|_{S} .
$$

### 2.3 Types of Reconstruction Methods

There are two major approaches to numerical reconstruction processes, namely symmetric and unsymmetric methods, which will be treated separately in the following. The main focus will be on the coupling of trial and test side which is the key difference between both methods. From a certain viewpoint one can say that the unsymmetric methods generalize the symmetric methods.
We point out how sampling inequalities can be used in the analysis of both methods. For the sake of simplicity we will consider from now on sampling inequalities in Sobolev spaces on a bounded domain $\Omega \subset \mathbb{R}^{d}$ instead of sampling inequalities on general normed spaces. For more details on Sobolev spaces we refer to Chapter 3.

### 2.3.1 Unsymmetric Methods

Suppose we try to recover an unknown function $f \in W_{2}^{\tau}(\Omega)$ from its given data

$$
S_{\Lambda_{S}}(f):=\left(\begin{array}{c}
\lambda_{1}(f) \\
\vdots \\
\lambda_{N_{S}}(f)
\end{array}\right)
$$

The reconstruction is taken from the finite dimensional trial space

$$
\mathcal{R}:=\operatorname{span}\left\{K\left(x_{j}, \cdot\right): x_{j} \in X_{R}\right\} .
$$

There is a good but numerically unavailable candidate $\Pi_{\mathcal{R}}(f)$, namely the best approximation from $\mathcal{R}$ to $f \in W_{2}^{\tau}(\Omega)$ (see [13] for details). Suppose we measure the reconstruction error in the $L_{2}(\Omega)$-norm. Then we assume an approximation property of the form

$$
\left\|g-\Pi_{\mathcal{R}}(g)\right\|_{L_{2}(\Omega)} \leq \epsilon\left(h_{R}\right)\|g\|_{W_{2}^{\tau}(\Omega)} \quad \text { for any } g \in W_{2}^{\tau}(\Omega),
$$

where $h_{R}:=h_{X_{R}, \Omega}$ denotes the fill distance or mesh norm of the set of nodes $X_{R}$ associated to the trial space $\mathcal{R}$ in the domain $\Omega$. The approximation rate $\epsilon\left(h_{R}\right)$ may be estimated by means of appropriate sampling inequalities. Provided that all assumptions are satisfied, we may apply a sampling inequality to the above situation and get

$$
\left\|g-\Pi_{\mathcal{R}}(g)\right\|_{L_{2}(\Omega)} \leq C h_{R}^{\tau}\|g\|_{W_{2}^{\tau}(\Omega)}+C\left\|g-\Pi_{\mathcal{R}}(g)\right\|_{\ell_{p}\left(X_{R}\right)} .
$$

If $\Pi_{\mathcal{R}}$ is for instance an interpolation operator, we immediately get $\epsilon\left(h_{R}\right)=C h_{R}^{\tau}$. Unfortunately, the function $\Pi_{\mathcal{R}}(f)$ is numerically unavailable because its computation might use infinitely many data. We therefore have to construct a reasonable approximation from the trial space using the given data. We shall employ the technique of residual minimization. This means that we try to solve the system

$$
\begin{equation*}
S_{\Lambda_{S}}\left(u_{r}\right)=S_{\Lambda_{S}}(f) \tag{2.3.1}
\end{equation*}
$$

for some $u_{r} \in \mathcal{R}$ to some accuracy. Let us first spell out the last equation,

$$
S_{\Lambda_{S}}\left(u_{r}\right)=\left(\begin{array}{ccc}
\lambda_{1}^{x} K\left(x, x_{1}\right) & \ldots & \lambda_{1}^{x} K\left(x, x_{N_{R}}\right) \\
\vdots & \vdots & \vdots \\
\lambda_{N_{S}}^{x} K\left(x, x_{1}\right) & \ldots & \lambda_{N_{S}}^{x} K\left(x, x_{N_{R}}\right)
\end{array}\right)\left(\begin{array}{c}
\alpha_{1} \\
\vdots \\
\alpha_{N_{R}}
\end{array}\right)=: A_{\Lambda_{S}, \mathcal{R}} \alpha
$$

This is an unsymmetric system, and it is not clear whether it is solvable at all. If it was solvable, there would be a good approximate solution because of

$$
\left\|S_{\Lambda_{S}}\left(f-\Pi_{\mathcal{R}}(f)\right)\right\| \leq\left\|S_{\Lambda_{S}}\right\| \cdot\left\|f-\Pi_{\mathcal{R}}(f)\right\| \leq \epsilon\left(h_{R}\right)\left\|S_{\Lambda_{S}}\right\| \cdot\|f\|
$$

To show full rank of $A_{\Lambda_{S}, \mathcal{R}}$, we need to couple the trial and the test discretization. To be precise, we have to assume an inverse estimate of the form

$$
\begin{equation*}
\left\|u_{r}\right\|_{W_{2}^{\tau}(\Omega)} \leq \gamma\left(h_{R}\right)\left\|u_{r}\right\|_{L_{2}(\Omega)} \quad \text { for all } u_{r} \in \mathcal{R} \tag{2.3.2}
\end{equation*}
$$

In Chapter 6 we shall derive estimates of this kind under certain technical assumptions. Now we may apply a sampling inequality based on the functionals $\lambda_{i}$ with sampling order $\sigma>0$ and define $\beta(s):=h_{s}^{\sigma}$ and a factor $\tilde{\beta}(s)$ in front of the discrete term to the functions $u_{r} \in \mathcal{R}$. This yields an estimate of the form

$$
\begin{equation*}
\left\|u_{r}\right\|_{L_{2}(\Omega)} \leq C \beta(s)\left\|u_{r}\right\|_{W_{2}^{\tau}(\Omega)}+C \tilde{\beta}(s)\left\|S_{\Lambda_{s}}\left(u_{r}\right)\right\|_{\ell_{p}\left(\mathbb{R}^{N_{s}}\right)} \tag{2.3.3}
\end{equation*}
$$

Combining the inequalities (2.3.2) and (2.3.3) yields

$$
\left\|u_{r}\right\|_{L_{2}(\Omega)} \leq C_{1} \beta(s) \gamma\left(h_{R}\right)\|u\|_{L_{2}(\Omega)}+C_{2} \tilde{\beta}(s)\left\|S_{\Lambda_{s}}\left(u_{r}\right)\right\|_{\ell_{p}\left(\mathbb{R}^{N_{s}}\right)}
$$

Now we have to couple the discretizations in test and trial side via the additional condition

$$
\begin{equation*}
C_{1} \beta(s) \gamma\left(h_{R}\right) \leq \frac{1}{2} \tag{2.3.4}
\end{equation*}
$$

This is always possible by making the test discretization $S$ fine enough, i.e., $N_{s}$ large enough. For practical applications we therefore need an explicit lower bound for $N_{s}$. Nevertheless, we end up with

$$
\left\|u_{r}\right\|_{L_{2}(\Omega)} \leq 2 C_{2} \tilde{\beta}(s)\left\|S_{\Lambda_{s}}\left(u_{r}\right)\right\|_{\ell_{p}\left(\mathbb{R}^{N_{s}}\right)}
$$

This inequality bounds a continuous norm via a discrete norm under the condition (2.3.4). This assures the full rank of the system (2.3.1). This procedure indicates that sampling inequalities can be used to detect parameter choices, which guarantee good generalization properties of recovery processes.

### 2.3.2 Symmetric Methods

In symmetric methods the test and trial side are essentially the same up to the Riesz isomorphism between a linear space and its dual. In this case we also have $R=S$, since they are finite dimensional spaces and dual to each other, so we may as well skip these parameters here. If we start with a set of $N$ test functionals $\lambda_{1}, \ldots, \lambda_{N}$, the trial space is constructed as

$$
\mathcal{R}:=\operatorname{span}\left\{\lambda_{j}^{x} K(x, \cdot): 1 \leq j \leq N\right\},
$$

where the notation $\lambda_{j}^{x} K(x, \cdot)$ indicates that the functional $\lambda_{j}$ acts on $K(x, \cdot)$ with respect to the variable $x$. For the special case of point evaluation functionals

$$
\lambda_{j}=\delta_{x_{j}} \text { with } x_{j} \in X=\left\{x_{1}, \ldots, x_{N}\right\}
$$

this yields again the trial space

$$
\mathcal{R}:=\operatorname{span}\left\{K\left(x_{j}, \cdot\right): x_{j} \in X\right\} .
$$

In the case that trial and test side discretization are equal, the error analysis is much simpler since the system (2.3.1) takes the form

$$
\begin{equation*}
S_{\Lambda_{S}}\left(u_{r}\right)=\left(\lambda_{i}^{x} \lambda_{j}^{y} K(x, y)\right)=: A_{\Lambda, \mathcal{R}} \alpha, \tag{2.3.5}
\end{equation*}
$$

where the matrix $A_{\Lambda, \mathcal{R}}$ is symmetric and positive definite if the functionals are linear independent and if the kernel $K$ is positive definite. Therefore we can skip the stability analysis. Hence we can choose the operator $\Pi_{\mathcal{R}}$ as a possibly approximate (generalized) interpolation operator, i.e., $\lambda_{i}(g) \approx \lambda_{i}\left(\Pi_{\mathcal{R}}(g)\right)$. The error analysis is then provided by a sampling inequality of the form

$$
\left\|g-\Pi_{\mathcal{R}}(g)\right\|_{L_{2}(\Omega)} \leq C h_{r}^{\tau}\|g\|_{W_{2}^{\tau}(\Omega)}+C\left\|g-\Pi_{\mathcal{R}}(g)\right\|_{\ell_{p}\left(X_{R}\right)} .
$$

Although this approach is obviously less flexible than the unsymmetric case, it is nevertheless quite popular in the context of generalized interpolation [36, 69], and it has been used successfully in various applications, in particular for elliptic partial differential equations (see, e.g., [18, 19, 20, 21, 66, 27, 28]). One advantage of these meshless collocation methods is for instance, that they provide more flexibility for adaptive methods, because no re-meshing is needed. In most applications, there is a boundary value problem of the form

$$
\begin{align*}
L u & =f \\
& \text { in } \Omega  \tag{2.3.6}\\
B u & =g
\end{align*} \text { on } \partial \Omega .
$$

Here $L$ is a linear partial differential operator of order $m$ with possibly variable coefficients, i.e.,

$$
L u(x)=\sum_{|\alpha| \leq m} c_{\alpha}(x) D^{\alpha} u(x),
$$

and $B$ is a typical boundary operator, e.g., $B=\mathrm{Id}$ in the case of Dirichlet boundary conditions. For the sake of simplicity we restrict ourselves here to the case of strong collocation.

Hence the system of equations (2.3.6) is discretized by point evaluations in certain collocation nodes.
In our example we could choose two sets of points $X_{\Omega}:=\left\{x_{1}, \ldots, x_{N_{1}}: x_{j} \in \Omega\right\}$ and $X_{\partial \Omega}:=\left\{x_{N_{1}+1}, \ldots, x_{N}: x_{j} \in \partial \Omega\right\}$ and build up a set of $N$ test functionals

$$
\begin{aligned}
\Lambda & :=\left\{\lambda_{1}, \ldots, \lambda_{N}\right\}, \text { where } \\
\lambda_{j} & =\delta_{x_{j}} \circ L \quad \text { for } 1 \leq j \leq N_{1} \text { and } \\
\lambda_{j} & =\delta_{x_{j}} \circ B \quad \text { for } N_{1}+1 \leq j \leq N
\end{aligned}
$$

As introduced above, the set $X_{s}:=\left\{x_{1}, \ldots, x_{N}\right\}$ is called the set of test points. This test part is the same for both, the symmetric and the unsymmetric methods. From this point on, we may continue as in the previous sections with symmetric or unsymmetric methods. An example of how sampling inequalities can be used in the analysis of these methods can be found in [66]. We shall not repeat all arguments, but simply refer to the literature for details.

### 2.4 Regularization

General recovery processes can be split up in two categories, namely interpolation and approximation. Interpolation intrinsically means that the error function $f-R f$ between the function $f$ to be recovered and its approximation $R f$ vanishes at the given data, whereas an approximation process may produce small errors at the given data. While in some applications interpolation is required, others, in particular those involving errors or noise in the given data, prefer approximation methods. Sometimes it is even useful to allow the accuracy to be a free parameter.
An important example of non-interpolatory recovery processes are regularized or approximate interpolation processes [67]. In this section we briefly outline how sampling inequalities can be used to derive worst-case convergence rates for regularized reconstruction processes. We shall concentrate on regularization methods that avoid exact solving of the system (2.3.5). For the sake of simplicity we shall consider only the symmetric case. This is not a big restriction, since we are mostly interested in the approximation quality of the trial space. This concerns unsymmetric recovery methods in the same way.
Besides improving condition numbers, most regularization methods have several advantages, as will be pointed out in detail in Chapter 5, e.g., regularization is closely related to sparse approximation [23]. The crucial point for all regularized reconstruction processes $\Pi^{\nu}$, where $\nu$ is a regularization parameter, is to show the following two properties.

$$
\begin{aligned}
\left\|\Pi^{\nu}(f)\right\|_{W_{2}^{\tau}(\Omega)} & \leq\|f\|_{W_{2}^{\tau}(\Omega)} \text { and } \\
\max _{1 \leq j \leq N}\left|\lambda_{j}\left(f-\Pi^{\nu} f\right)\right| & \leq g(\nu, f)\|f\|_{W_{2}^{\tau}(\Omega)}
\end{aligned}
$$

The function $g(\nu, f)$ determines the approximation quality of $\Pi^{\nu}$. These properties can again be seen as stability and consistency properties.

## Spline Smoothing

For the sake of simplicity we shall in the following concentrate on the simple case that is known as spline smoothing or $\ell_{2}$-spline-regression. A more detailed overview can be found in [58] and [65].
For a given $f \in W_{2}^{\tau}(\Omega)$ and the functionals $\lambda_{j}$ from the previous sections we can formulate the smoothed optimal recovery problem

$$
\begin{equation*}
\min _{s \in W_{2}^{\tau}(\Omega)} \sum_{j=1}^{N}\left|\lambda_{j}(s-f)\right|^{2}+\nu\|s\|_{W_{2}^{\tau}(\Omega)}^{2} \tag{2.4.1}
\end{equation*}
$$

where $\nu \geq 0$ is called the smoothing parameter. For a more detailed discussion of the smoothing parameter see [58]. We simply note that the special case $\nu=0$ corresponds to finding a generalized interpolant, i.e., a function $s_{f} \in W_{2}^{\tau}(\Omega)$ that satisfies the generalized interpolation conditions

$$
\lambda_{j}(f)=\lambda_{j}\left(s_{f}\right) \text { for all } j=1, \ldots, N
$$

It is well known [65] that there always exists a solution to this relaxed interpolation problem (2.4.1) in the linear space $s_{f}^{(\nu)} \in \operatorname{span}\left\{\lambda_{j}^{x} K(\cdot, x) \mid 1 \leq j \leq N\right\}$. The coefficients $a \in \mathbb{R}^{N}$ with respect to the basis $\left\{\lambda_{j}^{x} K(\cdot, x)\right\}$ can be found by solving the linear system

$$
\left(A_{\Lambda, K}+\nu \mathrm{Id}\right) a=f_{\Lambda},
$$

where

$$
A_{\Lambda, K}:=\left(\lambda_{j} \lambda_{k} K(\cdot, \cdot)\right)_{j, k=1, \ldots N}, \text { and } f_{\Lambda}=\left(\lambda_{1}(f), \ldots, \lambda_{N}(f)\right)^{T} .
$$

As elaborated in [67], we have the following two inequalities,

$$
\begin{aligned}
\left\|s_{f}^{(\nu)}\right\|_{W_{2}^{\tau}(\Omega)} & \leq\|f\|_{W_{2}^{\tau}(\Omega)}, \\
\max _{1 \leq j \leq N}\left|\lambda_{j}\left(f-s_{f}^{(\nu)}\right)\right| & \leq \sqrt{\nu}\|f\|_{W_{2}^{\tau}(\Omega)} .
\end{aligned}
$$

Applying a sampling inequality yields the bound

$$
\left\|f-s_{f}^{(\nu)}\right\|_{L_{2}(\Omega)} \leq C\left(h^{\tau}+\sqrt{\nu}\right)\|f\|_{W_{2}^{\tau}(\Omega)}
$$

This inequality suggests an a priori choice of the smoothing parameter as $\nu \leq h^{2 \tau}$, which leads to the optimal approximation order [48].

## Kernel-Based Learning

There is a close link between the theory of kernel-based approximation and machine learning. Although there is a broad collection of literature on this topic (to name just two: [49] and [51]), we shall briefly introduce the different viewpoint of learning theory and some of the notation. From now on, we deal only with strong recovery or approximation problems.

In approximation theory one typically wants to recover an unknown continuous function $f \in H$ from scattered function values $F=\left(f\left(x_{1}\right), \ldots, f\left(x_{N}\right)\right)^{T}$, where $f$ is evaluated at certain evaluation points $X:=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$. The aim of approximation theory is to develop effective reconstruction methods and to derive error bounds for them. They are usually given in terms of the fill distance, as defined in (1.0.1). The error is normally measured in some continuous norm, for instance

$$
\left\|f-s_{f}\right\|_{L_{2}(\Omega)} \leq C h_{X, \Omega}^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)}
$$

provided that $H \subset W_{2}^{\tau}(\Omega)$, and $s_{f}$ denotes the reconstruction. The most common choice for $s_{f}$ in approximation theory is an interpolant, i.e., $s_{f}\left(x_{i}\right)=f\left(x_{i}\right)$ for $1 \leq i \leq N$. This method obviously makes the best use of the data $F$. But there are also some drawbacks of classical interpolation, as will be indicated in the beginning of Section5.1. On the one hand, the reconstruction is very unstable if we consider $F$ to be spoiled by noise. On the other hand, there are also numerical disadvantages, namely the computation of the interpolant may be ill-conditioned. Furthermore, if $s_{f}=\sum_{j=1}^{N} \alpha_{j} K\left(\cdot, x_{j}\right)$ denotes the interpolant, there will be many non-zero coefficients, i.e., this method is non-sparse.
One way out of these problems is provided by learning theory. In learning theory, one would call the reconstruction problem an example of supervised regression, because the real values $F$ are generated by an unknown, but fixed function $f$. We shall consider instead of $F$ a vector of possibly disturbed values $y=\left(y_{1}, \ldots, y_{N}\right)^{T}$ with $y_{i} \approx f\left(x_{i}\right)$. One typically relaxes the interpolation condition by using a more general similarity measure, e.g., by using a loss function. A loss function $c: \Omega \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is a function of the form $c(x, f(x), y)$ satisfying

$$
\begin{aligned}
& c(x, f(x), y)=0 \quad \text { for } f(x)=y \\
& c(x, f(x), y) \geq 0 \quad \text { otherwise. }
\end{aligned}
$$

A typical example is Vapnik's $\epsilon$-intensive loss function [57]

$$
|f(x)-y|_{\epsilon}=\left\{\begin{array}{cl}
0 & i f|f(x)-y| \leq \epsilon \\
|f(x)-y|-\epsilon & i f|f(x)-y|>\epsilon
\end{array}\right.
$$

which allows deviations up to a positive parameter $\epsilon>0$. A reconstruction is then for instance obtained as solution to the optimization problem [52]

$$
\begin{equation*}
\min _{f \in H} \frac{1}{N} \sum_{j=1}^{N}\left|f\left(x_{j}\right)-y_{j}\right|_{\epsilon}+\lambda\|f\|_{H}^{2} \tag{2.4.2}
\end{equation*}
$$

with $\epsilon$ and $\lambda$ being fixed problem parameters. Under certain assumptions, this problem possesses a solution $w=\sum_{j=1}^{N} w_{j} K\left(\cdot, x_{j}\right)$, which can be computed by the following finite dimensional problem (see the representer theorem)

$$
\begin{align*}
& \min _{\substack{\mathbf{w} \in \mathbb{R}^{\mathbf{N}} \\
\xi^{*}, \xi \in \mathbb{R}^{\mathbf{N}}}} \frac{1}{2} \mathbf{w}^{T} \mathbf{K} \mathbf{w}+C \cdot\left(\frac{1}{N} \sum_{j=1}^{N}\left(\xi_{j}+\xi_{j}^{*}\right)\right) \\
& \text { subject to } \quad(\mathbf{K w})_{j}-y_{j} \leq \epsilon+\xi_{j} \\
&(-\mathbf{K w})_{j}+y_{j} \leq \epsilon+\xi_{j}^{*} \\
& \xi_{j}^{*}, \xi_{j} \geq 0, \epsilon \geq 0 \tag{2.4.3}
\end{align*}
$$

where $\mathbf{K}$ denotes the Gram matrix of the kernel $K$. The solution to this problem is known to be more sparse than the solution to the interpolation problem. Sampling inequalities can now be used to provide a worst-case error analysis for those problems, even in the case of noisy data. Such bounds require no knowledge about the statistical noise model. We refer to Chapter 5 for more details.

## Chapter 3

## Notation and Basic Facts

In this chapter we collect some basic facts, which we need throughout the whole thesis. We will not give any proofs, but refer to the literature instead. Especially important is the recent monograph [65], which contains most of the facts collected here and gives also some broader background. We point out that most of the results have been published earlier by various authors. We mainly stick to the notations from [65].

### 3.1 Notation

We frequently use the multi-index notation $\alpha:=\left(\alpha_{1}, \ldots, \alpha_{d}\right)^{T} \in \mathbb{N}_{0}^{d}$, in particular for monomials $x \mapsto x^{\alpha}=x_{1}^{\alpha_{1}} \cdots \cdots x_{d}^{\alpha_{d}}$ for $x \in \mathbb{R}^{d}$ and $\alpha \in \mathbb{N}_{0}^{d}$. The degree is then given by $|\alpha|:=|\alpha|_{d}:=\sum_{j=1}^{d}\left|\alpha_{j}\right|$. Similarly we define derivatives with a multi-index by $D^{\alpha} u(x)=\partial_{x_{1}}^{\alpha_{1}} \ldots \partial_{x_{d}}^{\alpha_{d}} u(x)$ for sufficiently smooth functions $u$.
The space $\pi_{k}\left(\mathbb{R}^{d}\right)$ denotes the set of all $d$-variate polynomials with degree at most $k$, i.e., $\pi_{k}\left(\mathbb{R}^{d}\right):=\operatorname{span}\left\{x^{\alpha}:|\alpha| \leq k\right\}=\operatorname{span}\left\{x_{1}^{\alpha_{1}} \cdots x_{d}^{\alpha_{d}}: \sum_{j=1}^{d}\left|\alpha_{j}\right| \leq k\right\}$. The dimension of $\pi_{k}\left(\mathbb{R}^{d}\right)$ is given by $\operatorname{dim} \pi_{k}\left(\mathbb{R}^{d}\right)=\binom{k+d}{d}$ (see [65]). For $\Omega \subset \mathbb{R}^{d}$ we denote the restriction of $\pi_{k}\left(\mathbb{R}^{d}\right)$ to $\Omega$ by $\pi_{k}^{d}(\Omega)$ or by $\pi_{k}(\Omega)$.
We also use the notation $(t)_{+}:=\max \{0, t\}$.
The Gamma function is defined by

$$
\Gamma(z):=\lim _{n \rightarrow \infty} \frac{n!n^{z}}{z(z+1) \ldots(z+n)} \quad \text { for } z \in \mathbb{C}
$$

There is an important bound on the factorials, namely Stirling's formula.

$$
\sqrt{2 \pi n}\left(\frac{n}{e}\right)^{n} \leq \Gamma(n+1)=n!\leq \sqrt{2 \pi n}\left(\frac{n}{e}\right)^{n} \cdot e^{\frac{1}{12 n}}
$$

which can be found in many textbooks, e.g., in [22].
We further use two important Bessel functions.
Definition 3.1.1 The Bessel function of the first kind of order $\nu \in \mathbb{C}$ is defined by

$$
J_{\nu}(z):=\sum_{m=0}^{\infty} \frac{(-1)^{m}(z / 2)^{2 m+\nu}}{m!\Gamma(\nu+m+1)} \quad \text { for } z \in \mathbb{C} \backslash\{0\}
$$

and the Bessel function of the third kind of order $\nu \in \mathbb{C}$ is

$$
K_{\nu}(z):=\int_{0}^{\infty} e^{-\cosh (z)} \cosh (\nu z) d t
$$

for all $z \in \mathbb{C}$ with $|\arg (z)| \leq \pi / 2$.
Now we briefly introduce the Fourier transform.
Definition 3.1.2 For $f \in L_{1}\left(\mathbb{R}^{d}\right)$ we define its Fourier transform and its inverse Fourier transform by

$$
\begin{aligned}
\hat{f}(\omega) & :=(2 \pi)^{-\frac{d}{2}} \int_{\mathbb{R}^{d}} f(x) e^{-i x^{T} \omega} d x, \omega \in \mathbb{R}^{d} \\
\breve{f}(x) & :=(2 \pi)^{-\frac{d}{2}} \int_{\mathbb{R}^{d}} f(\omega) e^{i x^{T} \omega} d \omega, x \in \mathbb{R}^{d}
\end{aligned}
$$

Function values may be recovered from the Fourier transform:
Lemma 3.1.3 If $f \in L_{1}\left(\mathbb{R}^{d}\right)$ is continuous and has a Fourier transform $\hat{f} \in L_{1}\left(\mathbb{R}^{d}\right)$, then $f$ can be recovered by its Fourier transform via

$$
f(x)=(2 \pi)^{-\frac{d}{2}} \int_{\mathbb{R}^{d}} \hat{f}(\omega) e^{i x^{T} \omega} d \omega, x \in \mathbb{R}^{d}
$$

We define Schwartz space, on which the Fourier transform is an automorphism.
Definition 3.1.4 The Schwartz space $\mathcal{S}$ consists of all functions $\gamma \in C^{\infty}\left(\mathbb{R}^{d}\right)$ that satisfy

$$
\left|x^{\alpha} D^{\beta} \gamma(x)\right| \leq C(\alpha, \beta, \gamma) \quad \text { for all } x \in \mathbb{R}^{d}
$$

for all multi-indices $\alpha, \beta \in \mathbb{N}_{0}^{d}$ with a constant $C(\alpha, \beta, \gamma)$ independent of $x \in \mathbb{R}^{d}$.
Now we can formulate the main theorem on the Fourier transform on $L_{2}$.

Theorem 3.1.5 The Fourier transform defines an automorphism on $\mathcal{S}$. The inverse mapping is given by the inverse Fourier transform. Furthermore the mapping is an isometry with respect to the $L_{2}$-norm, i.e., $\|f\|_{L_{2}\left(\mathbb{R}^{d}\right)}=\|\hat{f}\|_{L_{2}\left(\mathbb{R}^{d}\right)}$.

Proof: See [65, Theorem 5.23].
Since the Schwartz space is dense in $L_{2}\left(\mathbb{R}^{d}\right)$ (see [65, Corollary 5.24]) we can define the Fourier transform on $L_{2}\left(\mathbb{R}^{d}\right)$ as the unique continuous extension of the Fourier transform on $\mathcal{S}$.

### 3.2 Kernels and Native Spaces

In this section we shall give a short introduction to the theory of positive (semi-) definite kernels. For the treatment of conditionally positive (semi-) definite kernels we refer to the book [65].
At the beginning, we need some basic definitions. We shall deal only with real-valued kernels $K$, i.e., $K: \Omega \times \Omega \rightarrow \mathbb{R}$ for $\Omega \subset \mathbb{R}^{d}$. The first notion we need is that of positive semi-definiteness.

Definition 3.2.1 A continuous kernel $K: \Omega \times \Omega \rightarrow \mathbb{R}$ is called positive semi-definite on $\Omega \subset \mathbb{R}^{d}$ if for all $N \in \mathbb{N}$, all sets of pairwise distinct points $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ and all $\alpha \in \mathbb{R}^{d}$ we have

$$
\sum_{j=1}^{N} \sum_{k=1}^{N} \alpha_{j} \alpha_{k} K\left(x_{j}, x_{k}\right) \geq 0
$$

It is called positive definite if equality holds only for $\alpha=0$.
If $\Omega$ is a finite set, i.e., $\Omega$ consists of only $N_{\max }$ distinct points, the inequality is required only for $N \leq N_{\text {max }}$.

The notion of a kernel is a rather general concept. In the later applications we shall mostly consider kernels with some additional structure.

Definition 3.2.2 A kernel is called

- translation-invariant if it depends only on the difference of the two arguments, i.e., if there is a function $\tilde{K}: \mathbb{R}^{d} \rightarrow \mathbb{R}$ such that

$$
K(x, y)=\tilde{K}(x-y) \quad \text { for all } x, y \in \Omega .
$$

- radial if it depends only on the Euclidian norm of the difference of the two arguments, i.e., if there is a function $\tilde{K}:[0, \infty) \rightarrow \mathbb{R}$ such that

$$
K(x, y)=\tilde{K}\left(\|x-y\|_{2}\right) \quad \text { for all } x, y \in \Omega .
$$

- zonal if it depends only on the inner of product of the two arguments, i.e., if there is a function $\tilde{K}: \mathbb{R} \rightarrow \mathbb{R}$ such that

$$
K(x, y)=\tilde{K}(<x, y>) \quad \text { for all } x, y \in \Omega .
$$

The concept of radial kernels can be easily carried over to functions.
Definition 3.2.3 A function $F: \mathbb{R}^{d} \rightarrow \mathbb{R}$ is called radial if it depends only on the norm of the argument, i.e., if there is a function $f:[0, \infty) \rightarrow \mathbb{R}$ such that $F(x)=f\left(\|x\|_{2}\right)$ for all $x \in \Omega$.

The Fourier transform of a radial function is given by the following lemma.

Lemma 3.2.4 Suppose that $F \in L_{1}\left(\mathbb{R}^{d}\right) \cap C\left(\mathbb{R}^{d}\right)$ is radial, i.e., $F(x)=f\left(\|x\|_{2}\right)$, then the Fourier transform $\hat{F}$ is also radial, and

$$
\hat{F}(\omega)=\|\omega\|_{2}^{-(d-2) / 2} \int_{0}^{\infty} f(t) J_{(d-2) / 2}\left(\|\omega\|_{2} t\right) d t
$$

where $J_{(d-2) / 2}$ is a Bessel function of the first kind.
There is a close relation between positive definite kernels and reproducing kernel Hilbert spaces. Let us explain the concept of a reproducing kernel.

Definition 3.2.5 Let $\mathcal{H}(\Omega)$ be a (pre-) Hilbert space of real-valued functions $f: \Omega \rightarrow \mathbb{R}$. A function $K: \Omega \times \Omega \rightarrow \mathbb{R}$ is called reproducing kernel of $\mathcal{H}$ if

- $K(\cdot, y) \in \mathcal{H}(\Omega)$ for all $y \in \Omega$
- $f(y)=(f, K(\cdot, y))_{\mathcal{H}(\Omega)}$ for all $f \in \mathcal{H}(\Omega)$ and all $y \in \Omega$.

The following theorem shows that reproducing kernel Hilbert spaces are very natural.
Theorem 3.2.6 Suppose that $\mathcal{H}(\Omega)$ is a Hilbert space of functions $f: \Omega \rightarrow \mathbb{R}$.
The following statements are equivalent:

- The point evaluation functionals are continuous, i.e., $\delta_{x} \in \mathcal{H}(\Omega)^{*}$ for all $x \in \Omega$.
- $\mathcal{H}(\Omega)$ has a reproducing kernel.

Proof: See [65, Theorem 10.2]
It is possible to show that there is a one-to-one correspondence between reproducing kernel Hilbert spaces and positive semi-definite kernels. We shall explain this connection in more detail because it is of great importance for all applications.

Theorem 3.2.7 Suppose $\mathcal{H}(\Omega)$ is a reproducing kernel Hilbert space with kernel $K(\cdot, \cdot)$. Then $K$ is positive semi-definite. Furthermore $K$ is positive definite if and only if the point evaluation functionals are linearly independent over $\mathcal{H}(\Omega)$.

Proof: See [65, Theorem 10.4].
To show the other direction we directly construct a reproducing kernel Hilbert space that belongs to the positive semi-definite kernel $K$. We start with defining

$$
F_{K}(\Omega):=\operatorname{span}\{K(\cdot, y): y \in \Omega\}
$$

equipped with the bilinear form

$$
\left(\sum_{j=1}^{N} \alpha_{j} K\left(\cdot, x_{j}\right), \sum_{k=1}^{M} \beta_{k} K\left(\cdot, y_{k}\right)\right)_{K}:=\sum_{j=1}^{N} \sum_{k=1}^{M} \alpha_{j} \beta_{k} K\left(x_{j}, y_{k}\right) .
$$

By this construction $\left(F_{K}(\Omega),(\cdot, \cdot)_{K}\right)$ forms a pre-Hilbert space with reproducing kernel $K(\cdot, \cdot)$. We shall denote the Hilbert space completion of $\left(F_{K}(\Omega),(\cdot, \cdot)_{K}\right)$ by $\mathcal{F}_{K}(\Omega)$. In
the positive definite case it is clear that the completion consists of functions, i.e., that the point evaluation functionals are continuous. To be precise, it can be shown [65, Lemma 10.8] that the map

$$
\mathcal{R}: \mathcal{F}_{K}(\Omega) \rightarrow C(\Omega), \mathcal{R}(f)(x):=(f, K(\cdot-x))_{K}
$$

is well defined and injective. This allows us finally to define the native space of $K$.
Definition 3.2.8 The native space $\mathcal{N}_{K}(\Omega)$ of $K$ is given by

$$
\mathcal{N}_{K}(\Omega)=\mathcal{R}\left(\mathcal{F}_{K}(\Omega)\right)
$$

equipped with the inner product

$$
(f, g)_{\mathcal{N}_{K}(\Omega)}:=\left(\mathcal{R}^{-1} f, \mathcal{R}^{-1} g\right)_{K} .
$$

It is well known [65, Theorem 10.12] that $\left(\mathcal{N}_{\Phi}(\Omega),(\cdot, \cdot)_{\mathcal{N}_{\Phi}(\Omega)}\right)$ is a reproducing kernel Hilbert space with reproducing kernel $K$. Let us summarize these results in a theorem.

Theorem 3.2.9 Suppose that $K: \Omega \times \Omega \rightarrow \mathbb{R}$ is a symmetric positive definite kernel. Then its associated native space is a Hilbert space of functions $f: \Omega \rightarrow \mathbb{R}$ with reproducing kernel $K$.

The remaining question is whether this space is unique. This question is answered affirmatively by the following theorem.

Theorem 3.2.10 We have the two results:

- The reproducing kernel in a reproducing kernel Hilbert space is uniquely defined.
- Suppose that $K$ is a symmetric positive definite kernel. Suppose further that $\mathcal{G}$ is a reproducing kernel Hilbert space of functions on $\Omega$ with reproducing kernel $K$. Then $\mathcal{G}$ is the native space $\mathcal{N}_{K}(\Omega)$ for $K$, and the inner products are the same.


## Proof:

- This is a direct consequence of the reproduction property. Assume there are two reproducing kernels $K_{1}, K_{2}$. The reproducing property yields

$$
\left(f, K_{1}(\cdot, y)-K_{2}(\cdot, y)\right)_{\mathcal{N}_{K}(\Omega)}=0
$$

for all $f \in \mathcal{H}(\Omega)$ and all $y \in \Omega$. Inserting $f=K_{1}(\cdot, y)-K_{2}(\cdot, y)$ shows the first claim.

- See [65, Theorem 10.11].

This finishes our short detour through the abstract theory of reproducing kernels.

### 3.2.1 Properties of the Native Space

Here we collect some general features of native spaces.
Theorem 3.2.11 Suppose we are given data pairs $\left\{\left(x_{1}, y_{1}\right), \ldots,\left(x_{N}, y_{N}\right)\right\} \subset \Omega \times \mathbb{R}$. Then there is a function $s_{X, y, K} \in F_{K}(X)$ that fulfills the interpolation condition $s_{X, y, K}\left(x_{i}\right)=y_{i}$ for all $1 \leq i \leq N$.

Proof: If we spell out the interpolation condition for a specific interpolant $s_{X, y, K}(x)=$ $\sum_{j=1}^{N} \alpha_{j} K\left(\cdot, x_{j}\right)$, we find

$$
A_{X, K} \alpha:=\left(\begin{array}{ccc}
K\left(x_{1}, x_{1}\right) & \cdots & K\left(x_{1}, x_{N}\right) \\
\vdots & \ddots & \vdots \\
K\left(x_{N}, x_{1}\right) & \cdots & K\left(x_{N}, x_{N}\right)
\end{array}\right)\left(\begin{array}{c}
\alpha_{1} \\
\vdots \\
\alpha_{N}
\end{array}\right)=\left(\begin{array}{c}
y_{1} \\
\vdots \\
y_{N}
\end{array}\right) .
$$

Since $K$ is positive definite this system is uniquely solvable.
If we fix now the set $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ and assume that the Lagrange data $y=$ $\left(y_{1}, \ldots, y_{N}\right)^{T}$ are generated by a function $f \in \mathcal{N}_{K}(\Omega)$, we can characterize the interpolant by variational principles. First of all we know that $s_{X, f(X), K}$ is the best approximation from $F_{K}(\Omega)$ to $f$, i.e., we have

Theorem 3.2.12 With the notation from above we get

$$
\left(f-s_{X, f(X), K}, s\right)_{\mathcal{N}_{K}(\Omega)}=0 \text { for all } s \in \mathcal{N}_{K}(\Omega)
$$

Proof: See [65, Lemma 10.24].
This yields as an important corollary a Pythagoras-law in the native space.
Corollary 3.2.13 Under the assumption from above we get

$$
\left\|f-s_{X, f(X), K}\right\|_{\mathcal{N}_{K}(\Omega)}^{2}+\left\|s_{X, f(X), K}\right\|_{\mathcal{N}_{K}(\Omega)}^{2}=\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}
$$

for all $f \in \mathcal{N}_{K}$. This yields immediately

$$
\left\|f-s_{X, f(X), K}\right\|_{\mathcal{N}_{K}(\Omega)} \leq\|f\|_{\mathcal{N}_{K}(\Omega)} \text {, and }\left\|s_{X, f(X), K}\right\|_{\mathcal{N}_{K}(\Omega)} \leq\|f\|_{\mathcal{N}_{K}(\Omega)}
$$

Proof: See [65, Corollary 10.25] .
There are two more minimality properties of the interpolant.
Theorem 3.2.14 With the notations from above we get:

- $s_{X, f(X), K}=\sum_{j=1}^{N} \alpha_{j} K\left(\cdot, x_{j}\right)$ is the solution to the finite-dimensional optimization problem

$$
\min _{s \in F_{K}(\Omega)}\|f-s\|_{\mathcal{N}_{K}(\Omega)}=\min _{\alpha \in \mathbb{R}^{N}}\left\|f-\sum_{j=1}^{N} \alpha_{j} K\left(\cdot, x_{j}\right)\right\|_{\mathcal{N}_{K}(\Omega)} .
$$

- Further, $s_{X, f(X), K}$ is given by the solution to the infinite-dimensional variational problem

$$
\left\|s_{X, f(X), K}\right\|_{\mathcal{N}_{K}(\Omega)}=\min _{\substack{\left.s \in \mathcal{N}_{K}(\Omega) \\ s\right|_{X}=\left.f\right|_{X}}}\|s\|_{\mathcal{N}_{K}(\Omega)}
$$

Proof: See [44, Theorem $3.1 \&$ Theorem 3.3].
Now we shall give some examples of native spaces. It is remarkable that all the theory from above holds true in these special function spaces such as Sobolev spaces.

### 3.2.2 Native Spaces on $\mathbb{R}^{d}$

At the beginning of this section we briefly explain why radial kernels arise naturally in the context of native spaces on the whole $\mathbb{R}^{d}$.

Definition 3.2.15 Let $\mathcal{T}$ be a group of transformations $\mathcal{T} \ni T: \Omega \rightarrow \Omega$. We say that a Hilbert space $\mathcal{F}$ of functions $\Omega \rightarrow \mathbb{R}$ is invariant under the group $\mathcal{T}$ if

- $f \circ T \in \mathcal{F}$ for all $T \in \mathcal{T}$ and all $f \in \mathcal{F}$, and
- $(f \circ T, g \circ T)_{\mathcal{F}}=(f, g)_{\mathcal{F}}$ for all $T \in \mathcal{T}$ and all $f, g \in \mathcal{F}$.

Now it can be shown that the kernel inherits the invariance of its reproducing kernel Hilbert space.

Theorem 3.2.16 Suppose that the reproducing kernel Hilbert space $\mathcal{H}(\Omega)$ is invariant un$\operatorname{der} \mathcal{T}$. Then the reproducing kernel $K$ satisfies

$$
K(T(x), T(y))=K(x, y)
$$

for all $x, y \in \Omega$ and all $T \in \mathcal{T}$.
Proof: See [65, Theorem 10.16]
Let us now suppose $\Omega=\mathbb{R}^{d}$, and let $\mathcal{T}$ be group of all translations and orthogonal transformations on $\mathbb{R}^{d}$. Translations can be parametrized by $\tau_{x}: \mathbb{R}^{d} \rightarrow \mathbb{R}^{d}, x \in \mathbb{R}^{d}$, such that $\tau_{x}(y)=x-y$, rotations are transformations $\rho_{x}: \mathbb{R}^{d} \rightarrow \mathbb{R}^{d}, x \in \mathbb{R}^{d}$ with $\rho_{x}(y)=A_{x} y$ for $A_{x} \in S O\left(\mathbb{R}^{d}\right)$, such that $A_{x}(x)=\|x\| e_{1}$. Then we get for arbitrary $x, y \in \mathbb{R}^{d}$

$$
\begin{aligned}
K(x, y) & =K\left(\tau_{x}(x), \tau_{x}(y)\right)=K(0, x-y)=K\left(\rho_{x-y}(0), \rho_{x-y}(x-y)\right) \\
& =K\left(0,\|x-y\| e_{1}\right)=: \tilde{K}(\|x-y\|) .
\end{aligned}
$$

Hence, radial kernels are quite natural on $\mathbb{R}^{d}$. In this situation we shall frequently denote the kernel by $\Phi$ instead of $K$ and call it a function. In the case $\Omega=\mathbb{R}^{d}$ we can further characterize the native space in terms of Fourier transforms. A more detailed elaboration can be found in [65, Theorem 10.12].

Proposition 3.2.17 Let $\Phi \in C\left(\mathbb{R}^{d}\right) \cap L_{1}\left(\mathbb{R}^{d}\right)$ be a real-valued positive definite function. Then the native space has a representation

$$
\mathcal{N}_{\Phi}\left(\mathbb{R}^{d}\right):=\left\{f \in L_{2}\left(\mathbb{R}^{d}\right) \cap C\left(\mathbb{R}^{d}\right): \hat{f} / \sqrt{\hat{\Phi}} \in L_{2}\left(\mathbb{R}^{d}\right)\right\}
$$

equipped with the inner product

$$
(f, g)_{\mathcal{N}_{\Phi}\left(\mathbb{R}^{d}\right)}:=(2 \pi)^{-\frac{d}{2}}(\hat{f} / \sqrt{\hat{\Phi}}, \hat{g} / \sqrt{\hat{\Phi}})_{L_{2}\left(\mathbb{R}^{d}\right)}=(2 \pi)^{-\frac{d}{2}} \int_{\mathbb{R}^{d}} \frac{\hat{f}(\omega) \overline{\hat{g}}(\omega)}{\hat{\Phi}(\omega)} d \omega .
$$

The inner product is equivalent to the native space inner product.
A direct consequence of this result is that there is a close relation between native spaces and Sobolev spaces. If we assume a radial basis function $\Phi \in L_{1}\left(\mathbb{R}^{d}\right)$ satisfying

$$
\begin{equation*}
c_{1}\left(1+\|\omega\|_{2}^{2}\right)^{-\tau} \leq \hat{\Phi}(\omega) \leq c_{2}\left(1+\|\omega\|_{2}^{2}\right)^{-\tau}, \quad \omega \in \mathbb{R}^{d} \tag{3.2.1}
\end{equation*}
$$

for some $c_{1}, c_{2}>0$ and $\tau>d / 2$, then we have $\mathcal{N}_{\Phi}\left(\mathbb{R}^{d}\right)=W_{2}^{\tau}\left(\mathbb{R}^{d}\right)$. This connection holds true also on subsets $\Omega \subset \mathbb{R}^{d}$. To see this we first have to specify Sobolev spaces on bounded domains.

### 3.2.3 Sobolev Spaces on Bounded Domains

We recall that a function $f: U \subset \mathbb{R}^{d} \rightarrow \mathbb{R}$ is called uniformly Lipschitz if there is a constant $L>0$ such that $|f(x)-f(y)| \leq L|x-y|$ holds for all $x, y \in U$.

Definition 3.2.18 Let $\Omega \subset \mathbb{R}^{d}$ be bounded. Then $\Omega$ is said to be of class $C^{k}$ for $k \in \mathbb{N}$ if and only if for every point $x_{0} \in \partial \Omega$ there is a linear orthogonal coordinate-tansformation $y=A x$, an $r>0$ and a $k$-times differentiable function $\gamma$ such that

$$
\begin{aligned}
B\left(x_{0}, r\right) \cap \Omega & =\left\{x=A^{-1} y \in B\left(x_{0}, r\right): y_{d}>\gamma\left(y_{1}, \ldots, y_{d-1}\right)\right\} \\
B\left(x_{0}, r\right) \backslash \bar{\Omega} & =\left\{x=A^{-1} y \in B\left(x_{0}, r\right): y_{d}<\gamma\left(y_{1}, \ldots, y_{d-1}\right)\right\} \\
B\left(x_{0}, r\right) \cap \partial \Omega & =\left\{x=A^{-1} y \in B\left(x_{0}, r\right): y_{d}=\gamma\left(y_{1}, \ldots, y_{d-1}\right)\right\} .
\end{aligned}
$$

If the function $\gamma$ is uniformly Lipschitz, we call the domain a Lipschitz domain.
From now on we shall always assume that $\Omega$ is at least a Lipschitz domain, unless we explicitly say something else. We shall establish our error estimates using a variety of Sobolev spaces, which we want to introduce now.

Definition 3.2.19 Let $\Omega \subseteq \mathbb{R}^{d}$ be a domain. For $k \in \mathbb{N}_{0}$ and $1 \leq p<\infty$ we define the Sobolev space $W_{p}^{k}(\Omega)$ to consist of all functions $u$ with distributional derivatives $D^{\alpha} u \in$ $L_{p}(\Omega),|\alpha| \leq k$. Associated with these spaces are the (semi-) norms

$$
|u|_{W_{p}^{k}(\Omega)}=\left(\sum_{|\alpha|=k}\left\|D^{\alpha} u\right\|_{L_{p}(\Omega)}^{p}\right)^{1 / p} \text { and }\|u\|_{W_{p}^{k}(\Omega)}=\left(\sum_{|\alpha| \leq k}\left\|D^{\alpha} u\right\|_{L_{p}(\Omega)}^{p}\right)^{1 / p} .
$$

The case $p=\infty$ is defined in the standard way:

$$
|u|_{W_{\infty}^{k}(\Omega)}=\mathrm{ess} \sup _{|\alpha|=k}\left\|D^{\alpha} u\right\|_{L_{\infty}(\Omega)} \text { and } \quad\|u\|_{W_{\infty}^{k}(\Omega)}=\operatorname{ess} \sup _{|\alpha| \leq k}\left\|D^{\alpha} u\right\|_{L_{\infty}(\Omega)} .
$$

We will also be dealing with Sobolev spaces of fractional order. Let $1 \leq p<\infty, k \in \mathbb{N}_{0}$ and $0<s<1$. We define the fractional order Sobolev spaces $W_{p}^{k+s}(\Omega)$ to consist of all $u$, for which the following (semi-)norms are finite:

$$
\begin{aligned}
|u|_{W_{p}^{k+s}(\Omega)} & :=\left(\sum_{|\alpha|=k} \int_{\Omega} \int_{\Omega} \frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|^{p}}{\|x-y\|_{2}^{d+p s}} d x d y\right)^{1 / p}, \\
\|u\|_{W_{p}^{k+s}(\Omega)} & :=\left(\|u\|_{W_{p}^{k}(\Omega)}^{p}+|u|_{W_{p}^{k+s}(\Omega)}^{p}\right)^{1 / p}
\end{aligned}
$$

### 3.3 Extension Maps

We recall two essential definitions from [56].
Definition 3.3.1 Let $\phi: \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ be a function that satisfies the Lipschitz condition

$$
|\phi(x)-\phi(y)| \leq M|x-y|, \quad \forall x, y \in \mathbb{R}^{d-1} .
$$

The smallest constant, for which the Lipschitz condition is true, is denoted by $C_{\phi}$. In terms of this function the special Lipschitz domain is the set of point lying above the hypersurface $y=\phi(x)$ in $\mathbb{R}^{d}$, i.e.,

$$
\Omega=\left\{x \in \mathbb{R}^{d}: x_{d}>\phi\left(x_{1}, x_{2}, \ldots, x_{d-1}\right)\right\} .
$$

The Lipschitz constant of $\Omega$ is defined by $C_{\Omega}:=C_{\phi}$.
Definition 3.3.2 Let $\Omega$ be a domain in $\mathbb{R}^{d}$. The boundary $\partial \Omega$ is called minimally smooth if there are constants $\delta>0, N \in \mathbb{N}$ and $M>0$, and a sequence $\left\{U_{i}\right\}_{i \in \mathbb{N}}$ of open sets, such that

1. No point of $\mathbb{R}^{d}$ is contained in more than $N$ of the $U_{i}$ 's.
2. If $x \in \partial \Omega$, then $B(x, \delta) \subset U_{i}$ for some $i$.
3. For each i there exists a special Lipschitz domain $D_{i}$ whose Lipschitz constant does not exceed $M$ such that

$$
U_{i} \cap \Omega=U_{i} \cap D_{i}
$$

With these notions we cite the following theorem.
Theorem 3.3.3 [56] Let $\Omega \subset \mathbb{R}^{d}$ be a domain with minimally smooth boundary. There is a bounded linear extension operator $\mathcal{E}$ such that for any $k \in \mathbb{N}$ and $1 \leq p \leq \infty$

$$
\mathcal{E}: W_{p}^{k}(\Omega) \rightarrow W_{p}^{k}\left(\mathbb{R}^{d}\right)
$$

with a norm $\|\mathcal{E}\|$ which is bounded by a constant, depending on $k, p, d$ and the constants of Definition 3.3.2.

Such extension theorems allow us to show that native spaces on bounded domains $\Omega \subset \mathbb{R}^{d}$ are Sobolev spaces.

Proposition 3.3.4 Suppose that $\Phi$ satisfies (3.2.1), and that $\Omega \subset \mathbb{R}^{d}$ has a Lipschitz boundary. Then $\mathcal{N}_{\Phi}(\Omega)=W_{2}^{\tau}(\Omega)$, and the norms are equivalent.

Proof: See [65, Corollary 10.48].
To continue we need some geometrical notations.

### 3.3.1 Some Geometrical Terms

We shall review some technical notions concerning the geometry of $\Omega$.
Definition 3.3.5 A set $\Omega \subset \mathbb{R}^{d}$ is said to satisfy an interior cone condition if there exists an angle $\theta \in(0, \pi / 2)$ and a radius $r>0$ such that for every $x \in \Omega$ a unit vector $\xi(x)$ exists such that the cone $C:=C(x, \xi(x), \theta, r)$

$$
\begin{equation*}
C:=\left\{x+\lambda y: y \in \mathbb{R}^{d},\|y\|_{2}=1, y^{T} \xi(x) \geq \cos (\theta), \lambda \in[0, r]\right\} \tag{3.3.1}
\end{equation*}
$$

is contained in $\Omega$.
In particular, a domain satisfying an interior cone condition cannot have any outward pointing cusps. The notion of an interior cone condition generalizes the notion of convexity in the following sense.

Lemma 3.3.6 Suppose that the cone $C(x, \xi(x), \theta, r)$ is a cone defined as in Definition 3.3.5. Then for every $h<r /(1+\sin (\theta))$ the closed ball
$B=B(x+h \xi(x), h \sin (\theta))$ is contained in $C(x, \xi(x), \theta, r)$. In particular, if z is a point from this ball then the whole line segment $x+t(z-x) /\|z-x\|_{2}, t \in[0, r]$ is contained in the cone.

Proof: See [65, Lemma 3.7].
We further make use of the fact that a ball and a cone itself obey an interior cone condition.

Lemma 3.3.7 Every ball with radius $\delta>0$ satisfies an interior cone condition with radius $\delta$ and angle $\theta=\pi / 3$.
Let $C\left(x_{0}, \xi, \theta, r\right)$ be a cone with angle $\left.\theta \in 0, \pi / 5\right]$ and radius $r>0$. Then $C\left(x_{0}, \xi, \theta, r\right)$ satisfies an interior cone condition with angle $\tilde{\theta}=\theta$ and radius

$$
\tilde{r}=\frac{3 \sin (\theta)}{4(1+\sin (\theta))} r .
$$

Proof: See [65, Lemma 3.10 \& 3.12].
Furthermore, we need a generalization of star-shaped domains.

Definition 3.3.8 A domain $\Omega \subset \mathbb{R}^{d}$ is said to be star-shaped with respect to a ball $B\left(x_{c}, \rho\right):=\left\{x \in \mathbb{R}^{d}:\left\|x-x_{c}\right\|_{2} \leq \rho\right\}$ if for every $x \in \Omega$ the closed convex hull of $\{x\} \cup B\left(x_{c}, \rho\right)$ is contained in $\Omega$. If $\Omega$ is bounded, i.e., $\Omega \subset B\left(x_{c}, R\right)$ for some $R>0$, then the chunkiness parameter $\gamma$ is defined to be the ratio of the diameter diam ${ }_{\Omega}$ to the radius $\rho_{\max }$ of the largest ball relative to which $\Omega$ is star-shaped, i.e., $\gamma=\frac{\operatorname{diam}_{\Omega}}{\rho_{\max }}$.

From now on we shall assume $\frac{\rho_{\max }}{2} \leq \rho \leq \rho_{\max }$. Then the chunkiness parameter is bounded (see [38]) by

$$
\frac{1}{2} \leq \frac{\operatorname{diam}_{\Omega}}{2 \rho} \leq \gamma=\frac{\operatorname{diam}_{\Omega}}{\rho_{\max }} \leq \frac{2 R}{\rho}
$$

Furthermore, such domains obey an interior cone condition.
Proposition 3.3.9 If $\Omega \subset B\left(x_{c}, R\right)$ is star-shaped with respect to $B\left(x_{x}, \rho\right)$, then $\Omega$ satisfies an interior cone condition with radius $r=\rho$ and angle $\theta=2 \arcsin (\rho /(2 R))$.

Proof: See [65, Proposition 11.16].

We shall make use of the following covering property, which was established in [38].
Theorem 3.3.10 [38] Suppose $\Omega \subset \mathbb{R}^{d}$ is bounded, has a Lipschitz boundary and satisfies an interior cone condition with maximum radius $R$ and angle $\phi \in(0, \pi / 2)$. Then there is $a$ constant $Q_{\theta, R}$ such that for $h \leq \frac{Q_{\theta, R}}{k^{2}}$ there is a covering $\Omega=\bigcup_{t \in T_{h}} \mathcal{D}_{t}$ with the following properties.

1. Each set $\mathcal{D}_{t}$ is star-shaped with respect to a ball $B_{t} \subset B_{R}(t) \cap \Omega$.
2. Each set $D_{t}$ satisfies an interior cone condition with radius $r$ and angle $\theta$, where $r$ and $\theta$ can be expressed explicitly by $R$ and $\phi$.
3. There are constants $D_{\phi}, \tilde{D}_{\phi}$ such that $\tilde{D}_{\phi} \cdot h k^{2} \leq \delta_{\mathcal{D}_{t}} \leq D_{\phi} \cdot h k^{2}$.
4. There is a constant $M_{1}=M_{1}(\theta, d)>0$ such that $\sum_{t \in T_{r}} \chi_{\mathcal{D}_{t}} \leq M_{1}$.
5. There is a constant $M_{2}(\theta, d)>0$ such that $\# T_{h}<M_{2}\left(h k^{2}\right)^{-d}$.

In the following we state some results from polynomial approximation theory.

### 3.3.2 Polynomial Approximation in Sobolev Spaces

Several proofs of error estimates in radial basis function approximation theory are based on good polynomial approximations. For our purposes, the most important polynomials are the so called averaged Taylor polynomials. They are defined by

$$
\begin{equation*}
Q^{k} u(x):=\sum_{|\alpha|<k} \frac{1}{\alpha!} \int_{B} D^{\alpha} u(y)(x-y)^{\alpha} \phi(y) d y \tag{3.3.2}
\end{equation*}
$$

where $B$ is a ball with radius $\geq 1 / 2 \rho_{\max }$ relative to which $\mathcal{D}$ is star-shaped, and $\phi \in C^{\infty}$ is a bump function supported on $\bar{B}$ satisfying both $\int_{B} \phi(y) d y=1$ and $\max \phi \leq$ $C_{d} \operatorname{diam}(B)^{-d}$.
$Q^{k} u(x)$ is indeed a polynomial of degree $k$ [11]. In the following we shall denote the diameter of $B$ by $\rho:=\operatorname{diam}(B)$. For the remainder

$$
R^{k}:=u-Q^{k} u
$$

there is the following bound from [11], while the explicit constants can be found in [38]. The Bramble-Hilbert Lemma [11, Lemma 4.3.8] is crucial for the following computations, so we state it here with constants, where we made the dependence on the domain explicit.
Lemma 3.3.11 For $u \in W_{p}^{k}(\mathcal{D})$ and $p \geq 1$ we have

$$
\begin{equation*}
\left|R^{k} u\right|_{W_{p}^{j}(\mathcal{D})}=\left|u-Q^{k} u\right|_{W_{p}^{j}(\mathcal{D})} \leq C_{\gamma, k, d} \delta_{\mathcal{D}}^{k-j}|u|_{W_{p}^{k}(\mathcal{D})} \quad j=0,1, \ldots, k, \tag{3.3.3}
\end{equation*}
$$

where $C_{\gamma, k, d}=C_{k, d}(1+2 \gamma)^{d}$. Here, $\gamma$ denotes again the chunkiness parameter from Definition 3.3.8.

### 3.4 Norming Sets and Polynomial Reproduction

This section deals with general Banach spaces. In the following let $V$ be a finite dimensional linear space endowed with the norm $\|\cdot\|_{V}$. We denote by $V^{*}$ the dual space of $V$, i.e., the space of all bounded linear functionals on $V$, and by $Z \subset V^{*}$ a finite subset of $V^{*}$ with $|Z|=N$.

Definition 3.4.1 $Z$ is called $a$ norming set for $V$ if the map

$$
\begin{aligned}
T: \quad & \rightarrow T(V) \subset \mathbb{R}^{N} \\
v & \mapsto T(v)=(z(v))_{z \in Z}
\end{aligned}
$$

is injective. $T$ will be called the sampling operator.
To explain why $Z$ is called a norming set we have to make a detour. First we have to introduce a norm $\|\cdot\|_{\mathbb{R}^{N}}$ on the $\mathbb{R}^{N}$ and a corresponding dual norm $\|\cdot\|_{\left(\mathbb{R}^{N}\right)^{*}}$ on $\left(\mathbb{R}^{N}\right)^{*} \cong$ $\mathbb{R}^{N}$. Note that $Z$ being a norming set means that $T$ is injective, which implies that $T$ is invertible on this image, i.e., there is an inverse map $T^{-1}: T(V) \rightarrow V$. The quantity

$$
\begin{equation*}
\left\|T^{-1}\right\|:=\sup _{\substack{x \in T(V) \\ x \neq 0}} \frac{\left\|T^{-1} x\right\|_{V}}{\|x\|_{\mathbb{R}^{N}}}=\sup _{\substack{v \in V \\ v \neq 0}} \frac{\|v\|_{V}}{\|T v\|_{\mathbb{R}^{N}}} \tag{3.4.1}
\end{equation*}
$$

is called norming constant. Now we can explain the name norming set, which is due to the fact that $Z$ allows us to introduce an equivalent norm on $V$ by $\|T(\cdot)\|_{\mathbb{R}^{N}}$. The equivalence constants are given by

$$
\frac{1}{\|T\|}\|T v\|_{\mathbb{R}^{N}} \leq\|v\|_{V} \leq\left\|T^{-1}\right\|\|T v\|_{\mathbb{R}^{N}} \text { for all } v \in V
$$

The norming constant plays an essential rôle as can be seen in the next theorem.

Theorem 3.4.2 Suppose $V$ is a finite dimensional normed linear space, and the set of functionals $Z=\left\{z_{1}, \ldots, z_{N}\right\}$ is a norming set for $V$ with sampling operator $T$. Then there is for every $\psi \in V^{*}$ a vector $u=u(\psi) \in \mathbb{R}^{N}$, such that

$$
\begin{aligned}
\psi(v) & =\sum_{j=1}^{N} u_{j} z_{j}(v), \text { and } \\
\|u\|_{\left(\mathbb{R}^{N}\right)^{*}} & \leq\|\psi\|_{V^{*}}\left\|T^{-1}\right\|
\end{aligned}
$$

for every $v \in V$.
Proof: [65, Theorem 3.4].
Now we introduce the central idea of norming sets, namely that we can control the norms of $T$ and its inverse by enlarging the size of $Z$. This will be called oversampling. We shall exemplify this idea in a special case, see [63]. Let $V$ be the restriction of $\pi_{m}\left(\mathbb{R}^{d}\right)$ to $\Omega$ where $\pi_{m}\left(\mathbb{R}^{d}\right)$ again denotes the space of all algebraic polynomials on $\mathbb{R}^{d}$ with degree at most $m$, i.e., $V=\pi_{m}(\Omega)$. Let $Z=\left\{\delta_{x_{1}}, \ldots, \delta_{x_{N}}\right\}$ consist of point evaluation functionals. Then we have the following result, where we choose the $\|\cdot\|_{\ell_{\infty}}$-norm on the $\mathbb{R}^{N}$ with dual norm $\|\cdot\|_{\ell_{1}}$. In the following we shall denote by $\alpha \in \mathbb{N}_{0}^{d}$ a multi-index.

Theorem 3.4.3 Suppose $\Omega$ is compact and satisfies an interior cone condition with radius $r>0$ and angle $\theta \in(0, \pi / 2)$. Let $m \in \mathbb{N}$ be fixed, and suppose $h>0$. If $X=$ $\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ satisfies

1. $h \leq \frac{r \sin \theta}{4(1+\sin \theta) m^{2}}$
2. for every $B(x, h) \subset \Omega$ there is a center $x_{j} \in X \cap B(x, h)$,
then $Z=\left\{D^{\alpha} \circ \delta_{x_{1}}, \ldots, D^{\alpha} \circ \delta_{x_{N}}\right\}$ is for every multi-index $|\alpha| \leq m$ a norming set for $V=\pi_{m}(\Omega)$. The norming constant is bounded by $2\left(\frac{2 m^{2}}{r \sin \theta}\right)^{|\alpha|}$.

Proof: See [65, Proposition 11.7].
If we in particular choose $\psi=D^{\alpha} \circ \delta_{x}$ for some $x \in \Omega$, this implies that for every $p \in \pi_{m}(\Omega)$ and any $|\alpha| \leq m$ there are real numbers $a_{j}^{(\alpha)}(x)$ such that

$$
D^{\alpha} p(x)=\sum_{j=1}^{N} a_{j}^{(\alpha)}(x) p\left(x_{j}\right),
$$

where

$$
\sum_{j=1}^{N}\left|a_{j}^{(\alpha)}(x)\right| \leq 2\left(\frac{2 m^{2}}{r \sin \theta}\right)^{|\alpha|} .
$$

We point out that the constant arises from a Bernstein inequality for multivariate polynomials. Furthermore, all of this implies that recovery of functions by polynomials from
function values at scattered locations can be made with bounded Lebesgue constants if moderate oversampling is allowed. Therefore, the functions $a_{j}^{(\alpha)}(\cdot)$ are called a polynomial reproduction formula of degree $m$. Finally this yields a finite version of a Markov-Bernstein inequality.

Theorem 3.4.4 Suppose that the domain $\Omega \subset \mathbb{R}^{d}$ is compact and satisfies an interior cone condition with radius $r>0$ and angle $\theta$. If $p \in \pi_{m}(\Omega)$ and $\alpha \in \mathbb{N}_{0}^{d}$ with $|\alpha| \leq m$, we have

$$
\left\|D^{\alpha} p\right\|_{L_{\infty}(\Omega)} \leq\left(\frac{2 m^{2}}{r \sin \theta}\right)^{|\alpha|}\|p\|_{L_{\infty}(\Omega)}
$$

Proof: See [65, Proposition 11.6].
This is only a brief detour through the theoretical basics we need throughout the thesis. For more details we refer to the literature.

## Chapter 4

## Infinitely Smooth Functions

We derive in this chapter sampling inequalities for infinitely smooth functions where the sampling orders turn out to depend exponentially on the fill distance $h$.
We are handling infinitely smooth functions by normed linear function spaces $\mathcal{H}(\Omega)$ on a domain $\Omega \subset \mathbb{R}^{d}$ that can for a fixed $1 \leq p<\infty$ be continuously embedded into every classical Sobolev space $W_{p}^{k}(\Omega)$. More precisely, for a fixed $p \in[1, \infty)$ and all $k \in \mathbb{N}$ we assume that there are embedding operators $I_{k}^{(p)}$ and constants $E(k)$ such that

$$
\begin{align*}
& I_{k}^{(p)}: \mathcal{H}(\Omega) \rightarrow W_{p}^{k}(\Omega) \quad \text { with } \\
& \left\|I_{k}^{(p)}\right\|_{\left\{\mathcal{H}(\Omega) \rightarrow W_{p}^{k}(\Omega)\right\}} \leq E(k) \quad \text { for all } k \in \mathbb{N}_{0} \tag{4.0.1}
\end{align*}
$$

There are various examples of spaces with this property, e.g., Sobolev spaces of infinite order as they occur for instance in the study of partial differential equations of infinite order [1], or reproducing kernel Hilbert spaces of Gaussians and inverse multiquadrics (see Section 4.4).
In the case of infinitely smooth functions, the shape of the domain $\Omega$ crucially influences our sampling inequalities. For general Lipschitz domains $\Omega$, which satisfy interior cone conditions, we use a polynomial reproduction [65], which accepts slight oversampling to bound the Lebesgue-constants. This results in a good behavior of the term with the discrete norm. A typical result in this case is that for sufficiently small fill distance $h$ there are generic constants $c>0$ such that with $\tilde{q} \in\{q, \infty\}$ the inequality

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq c e^{c \log (c h) / \sqrt{h}}\|u\|_{\mathcal{H}(\Omega)}+c h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\bar{q}}(X)}
$$

holds for all $u \in \mathcal{H}(\Omega)$. The best approximation orders for the first term can be obtained on compact cubes because we then may use a polynomial reproduction from [32]. Unfortunately this approach is limited to cubes and cannot cope with derivatives on the left hand-side of our sampling inequalities. Nevertheless we obtain as a typical result, which applies for instance to the functions from the native space of Gaussian kernels, that there are generic constants $c>0$, such that the inequality

$$
\|u\|_{L_{q}(\Omega)} \leq e^{c \log (c h) / h}\|u\|_{\mathcal{H}(\Omega)}+c^{1 / h}\left\|\left.u\right|_{X}\right\|_{\ell_{\tilde{q}}(X)}
$$

holds for all $u \in \mathcal{H}(\Omega)$ with $\tilde{q} \in\{q, \infty\}$ if the fill distance $h$ is sufficiently small.
It is an open research problem to improve the polynomial reconstruction results in [26] and [65]. There is some further discussion on this point in the outlook (see Chapter 9 ). Our main examples deal with reconstruction problems in Hilbert spaces. Therefore, in the second part we will focus on the native Hilbert spaces of Gaussian and inverse multiquadric kernels. In this case, we suppose $u$ to be an error function $u=f-R f$, where $f$ denotes the function we would like to reconstruct, and $R f$ is the reconstruction. To obtain optimal order error bounds one needs again stability and consistency of the reconstruction, namely

$$
\|R f\|_{\mathcal{H}} \leq C\|f\|_{\mathcal{H}}, \quad \text { and } \quad\left\|\left.(R f-f)\right|_{X}\right\|_{\ell_{p}(X)} \leq g(f, h),
$$

where $g$ determines the expected approximation order. This can be used to show that the theory presented here reproduces the well-known exponential error estimates for the standard interpolation problem in the native Hilbert space of the inverse multiquadrics and Gaussian kernels.

### 4.1 Estimates on General Lipschitz Domains

Following [38], we first obtain estimates on local domains $\mathcal{D} \subset \mathbb{R}^{d}$ and use a covering argument to get global results. We assume a domain $\mathcal{D}$ that is is star-shaped with respect to a ball $B_{r}\left(x_{c}\right)$, and that is contained in a ball $B_{R}\left(x_{c}\right)$. In this case we know from [38] that $\mathcal{D}$ satisfies an interior cone condition as well. We denote the associated chunkiness parameter with

$$
\gamma=\frac{\delta_{\mathcal{D}}}{\rho_{\max }},
$$

where $\rho_{\max }=\sup \{\rho: \mathcal{D}$ is star-shaped with respect to a ball of radius $\rho\}$, and $\delta_{\mathcal{D}}$ denotes the diameter of $\mathcal{D}$.
Let $\left\{a_{j}^{(\alpha)}: j=1, \ldots, N\right\}$ be a polynomial reproduction of degree $k$ with respect to a discrete set $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \mathcal{D}$, i.e.,

$$
D^{\alpha} q(x)=\sum_{j=1}^{N} a_{j}^{(\alpha)}(x) q\left(x_{j}\right)
$$

holds for every $\alpha \in \mathbb{N}_{0}^{d}$ with $|\alpha| \leq k$, all $x \in \mathcal{D}$ and all $q \in \pi_{k}^{d}(\mathcal{D})$ where $\pi_{k}^{d}$ denotes the space of all $d$-variate polynomials of degree not exceeding $k$. Then we have

$$
\begin{align*}
& \left|D^{\alpha} u(x)\right| \leq\left|D^{\alpha} u(x)-D^{\alpha} p(x)\right|+\left|D^{\alpha} p(x)\right| \\
& \leq\left\|D^{\alpha} u-D^{\alpha} p\right\|_{L_{\infty}(\mathcal{D})}+\sum_{j=1}^{N}\left|a_{j}^{(\alpha)}(x)\right|\left|p\left(x_{j}\right)\right| \\
& \leq\left\|D^{\alpha} u-D^{\alpha} p\right\|_{L_{\infty}(\mathcal{D})}+\sum_{j=1}^{N}\left|a_{j}^{(\alpha)}(x)\right|\left\|\left.p\right|_{X}\right\|_{\ell_{\infty}(X)} \\
& \leq\left\|D^{\alpha} u-D^{\alpha} p\right\|_{L_{\infty}(\mathcal{D})} \\
& +\sum_{j=1}^{N}\left|a_{j}^{(\alpha)}(x)\right|\left(\|u-p\|_{L_{\infty}(\mathcal{D})}+\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}\right) \tag{4.1.1}
\end{align*}
$$

for arbitrary $u \in W_{p}^{k}(\mathcal{D})$ and any polynomial $p \in \pi_{k}^{d}(\mathcal{D})$. As a polynomial approximation we use again averaged Taylor polynomials. We recall the definition (3.3.2)

$$
Q^{k} u(x):=\sum_{|\alpha|<k} \frac{1}{\alpha!} \int_{B} D^{\alpha} u(y)(x-y)^{\alpha} \phi(y) d y
$$

where $B$ is a ball with radius $\geq 1 / 2 \rho_{\max }$, relative to which $\mathcal{D}$ is star-shaped, and $\phi \in$ $C^{\infty}$ is a bump function supported on $\bar{B}$ satisfying both $\int_{B} \phi(y) d y=1$ and $\max \phi \leq$ $C_{d} \operatorname{diam}(B)^{-d}$. For the remainder $R^{k}:=u-Q^{k} u$ there is the following bound from [11], where the explicit constants can be found in [38]. This bound differs slightly from the bound in Lemma 3.3.11. We state it here, because we need to calculate the explicit dependence of the constants on $k$.
Lemma 4.1.1 For $u \in W_{p}^{k}(\mathcal{D})$ with $1<p<\infty$ and $k>|\alpha|+d / p$ or in the case $p=1$ and $k \geq|\alpha|+d$ we get

$$
\left\|D^{\alpha} u-D^{\alpha} Q^{k} u\right\|_{L_{\infty}(\mathcal{D})} \leq C_{d, \theta} \frac{d^{k-|\alpha|}}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-|\alpha|-d / p}|u|_{W_{p}^{k}(\mathcal{D})}
$$

where the constant $C_{d, \theta}$ depends only on the space dimension $d$ and the angle $\theta$.
Proof: We use the identity [11] $D^{\beta} Q^{k} u=Q^{k-|\beta|} D^{\beta} u$, for all $|\beta| \leq k$. This leads to

$$
\begin{aligned}
\left\|D^{\alpha} u-D^{\alpha} Q^{k} u\right\|_{L_{\infty}(\mathcal{D})} & =\left\|D^{\alpha} u-Q^{k-|\alpha|} D^{\alpha} u\right\|_{L_{\infty}(\mathcal{D})} \\
& \leq C_{d}(1+\gamma)^{d} \frac{d^{k-|\alpha|}}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-|\alpha|-d / p}\left|D^{\alpha} u\right|_{W_{p}^{k-|\alpha|}(\mathcal{D})} \\
& \leq C_{d, \theta} \frac{d^{k-|\alpha|}}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-|\alpha|-d / p}|u|_{W_{p}^{k}(\mathcal{D})} .
\end{aligned}
$$

Here we used the fact [38] that the chunkiness parameter $\gamma$ can be bounded by $1 \leq \gamma \leq$ $\csc \left(\frac{\theta}{2}\right)$.

We shall use the following local polynomial reproduction from [65], which we introduced in the remarks below Theorem 3.4.3.

Theorem 4.1.2 Let $\Omega \subset \mathbb{R}^{d}$ satisfy an interior cone condition with angle $\theta \in(0, \pi / 2)$ and radius $r, \ell \in \mathbb{N}_{0}$ and $\alpha \in \mathbb{N}_{0}^{d}$ with $|\alpha| \leq \ell$. Then there are constants $c_{0}, c_{1}^{(\alpha)}, c_{2}>0$, such that for all $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ with $h_{X, \Omega} \leq h_{0}:=c_{0} / \ell^{2}$ and all $x \in \Omega$ there exist numbers $\tilde{a}_{1}^{(\alpha)}(x), \ldots, \tilde{a}_{N}^{(\alpha)}(x)$ with

1. $\sum_{j=1}^{N} p\left(x_{j}\right) \tilde{a}_{j}^{(\alpha)}(x)=D^{(\alpha)} p(x)$ for $x \in \Omega$ and $p \in \pi_{\ell}^{d}(\Omega)$
2. $\sum_{j=1}^{N}\left|\tilde{a}_{j}^{(\alpha)}(x)\right| \leq c_{1}^{(\alpha)} h_{X, \Omega}^{-|\alpha|}$ for all $x \in \Omega$,
3. $\tilde{a}_{j}^{(\alpha)}(x)=0$ if $\left\|x-x_{j}\right\|_{2}>c_{2}^{(\alpha)} h_{X, \Omega}$ and $x \in \Omega$.

The condition $\theta \in(0, \pi / 2)$ implies $\sin \theta \in(0,1)$, i.e., $\left(\frac{1}{2(1+\sin \theta)}\right)^{|\alpha|} \leq \frac{1}{2(1+\sin \theta)}$ for all $\alpha \in \mathbb{N}_{0}^{d}$. Therefore, we can choose all the constants independent of $\alpha$, i.e., there exist constants $c_{\theta}$ depending only on $\theta$ such that [65]

$$
\begin{equation*}
c_{1}^{(\alpha)} \leq c_{\theta} 2^{-|\alpha|} \leq c_{\theta}, \quad c_{2}:=c_{\theta} \ell^{2} . \tag{4.1.2}
\end{equation*}
$$

Inserting the bounds of Lemma4.1.1 and Theorem 4.1.2 into (4.1.1) leads to the following local estimate.

Theorem 4.1.3 Suppose $\mathcal{D}$ satisfies an interior cone condition with angle $\theta$ and radius $r$, let $\alpha \in \mathbb{N}_{0}^{d}$ such that $k>|\alpha|+d / p$ for $1<p<\infty$, or $k \geq d$ if $p=1$. Then

$$
\begin{aligned}
\left\|D^{\alpha} u\right\|_{L_{\infty}(\mathcal{D})} \leq & \frac{C_{d, \theta} d^{k}}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-d / p}\left(\delta_{\mathcal{D}}^{-|\alpha|}+h^{-|\alpha|}\right)|u|_{W_{p}^{k}(\mathcal{D})} \\
& +C_{d, \theta} h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
\end{aligned}
$$

holds for all $u \in W_{p}^{k}(\mathcal{D})$.
Corollary 4.1.4 Under the assumptions from Theorem 4.1.3, we get for $1 \leq q \leq \infty$

$$
\begin{aligned}
\left\|D^{\alpha} u\right\|_{L_{q}(\mathcal{D})} & \leq \operatorname{vol}(\mathcal{D})^{1 / q}\left\|D^{\alpha} u\right\|_{L_{\infty}(\mathcal{D})} \leq \delta_{\mathcal{D}}^{d / q}\left\|D^{\alpha} u\right\|_{L_{\infty}(\mathcal{D})} \\
& \leq c_{d, \theta} \frac{d^{k}}{(k-|\alpha|)!} \delta_{\mathcal{D}}^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)}\left(\delta_{\mathcal{D}}^{-|\alpha|}+h^{-|\alpha|}\right)|u|_{W_{p}^{k}(\mathcal{D})}+ \\
& +C_{d, \theta} \delta_{\mathcal{D}}^{d / q} h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
\end{aligned}
$$

Now we consider a 'global' domain $\Omega \subset \mathbb{R}^{d}$ that is bounded, has a Lipschitz boundary and satisfies an interior cone condition with maximum radius $R$ and angle $\phi \in(0, \pi / 2)$. To cover $\Omega$ with smaller star-shaped domains $\left\{\mathcal{D}_{t}\right\}$ we use the construction described in Theorem 3.3.10.

Theorem 4.1.5 Let $\alpha \in \mathbb{N}_{0}^{d}$ and $k \in \mathbb{N}$ be fixed with $|\alpha|<k, k>d / p$ for $1<p<\infty$ or $k \geq d$ for $p=1$ and set $C_{\min }:=\min \left\{\frac{c_{0}}{2}, Q_{\theta, R}\right\}$ with the constant $c_{0}$ from Theorem 4.1.2. Suppose a discrete set $X \subset \Omega$ with fill distance $h \leq C_{\min } / k^{2}$. Then for all $u \in W_{p}^{k}(\Omega)$ the inequality

$$
\begin{align*}
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq & \frac{c^{k} h^{-|\alpha|}}{(k-|\alpha|)!}\left(h k^{2}\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)}+|u|_{W_{p}^{k}(\Omega)} \\
& +c h^{-|\alpha|}\left(h k^{2}\right)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} \tag{4.1.3}
\end{align*}
$$

holds for $1 \leq q \leq \infty$ with generic positive constants $c$, which may depend only on $d, R, \phi, p, q$ and $\alpha$.

Proof: For $u \in W_{p}^{k}(\Omega)$ we may use the decomposition from Theorem 3.3.10, Corollary 4.1.4 and the estimate $\delta_{\mathcal{D}}^{-|\alpha|} \leq C_{\phi, \alpha} h^{-|\alpha|}$, which gives

$$
\begin{aligned}
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} & =\left(\int_{\Omega}\left|D^{\alpha} u(x)\right|^{q} d x\right)^{1 / q} \leq\left(\sum_{\mathcal{D}_{t}}\left\|D^{\alpha} u\right\|_{L_{q}\left(\mathcal{D}_{t}\right)}^{q}\right)^{1 / q} \\
& \leq \frac{C_{d, \theta} d^{k}}{(k-|\alpha|)!}\left(D_{\Phi} h k^{2}\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)} C_{\phi, \alpha} h^{-|\alpha|}\left(\sum_{\mathcal{D}_{t}}|u|_{W_{p}^{k}\left(\mathcal{D}_{t}\right)}^{q}\right)^{1 / q}+ \\
& +\left(D_{\Phi} h k^{2}\right)^{d / q} h^{-|\alpha|}\left(\sum_{\mathcal{D}_{t}}\left\|\left.u\right|_{X \cap \mathcal{D}_{t}}\right\|_{\ell_{\infty}\left(X \cap \mathcal{D}_{t}\right)}^{q}\right)^{1 / q} \\
& \left.\leq \frac{C_{d, \theta, \alpha, p, q}^{k}}{(k-|\alpha|)!} h^{-|\alpha|}\left(h k^{2}\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)-d\left(\frac{1}{q}-\frac{1}{p}\right.}\right)_{+}\left(\sum_{\mathcal{D}_{t}}|u|_{W_{p}^{k}\left(\mathcal{D}_{t}\right)}^{p}\right)^{1 / p}+ \\
& +C h^{-|\alpha|}\left(h k^{2}\right)^{d / q}\left(\sum_{\mathcal{D}_{t}}\left\|\left.u\right|_{X \cap \mathcal{D}_{t}}\right\|_{\ell_{\infty}\left(X \cap \mathcal{D}_{t}\right)}^{q}\right)^{1 / q} \\
& \left.\leq \frac{C^{k}}{(k-|\alpha|)!} h^{-|\alpha|}\left(h k^{2}\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right.}\right)_{+|u|_{W_{p}^{k}(\Omega)}}+ \\
& +C h^{-|\alpha|}\left(h k^{2}\right)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} .
\end{aligned}
$$

We can restate this Theorem measuring the discrete term in the $\ell_{\infty}$-norm.
Corollary 4.1.6 Under the assumptions from Theorem 4.1.5 we get with an analogous calculation

$$
\begin{align*}
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq & \frac{c^{k}}{(k-|\alpha|)!} h^{-|\alpha|}\left(h k^{2}\right)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)}+|u|_{W_{p}^{k}(\Omega)}  \tag{4.1.4}\\
& +c h^{-|\alpha|}|u|_{X} \|_{\ell_{\infty}(X)} .
\end{align*}
$$

We shall now relate $h$ and $k$ to derive main result of this section, i.e., exponential estimates. The actual orders depend on the asymptotic behavior of the constant $E(k)$ from (4.0.1) for $k \rightarrow \infty$.

Theorem 4.1.7 If there are constants $\epsilon, C_{E}>0$ such that $E(k) \leq C_{E}^{k} k^{(1-\epsilon) k}$ for all $k \in \mathbb{N}$, then there are constants $c, h_{0}>0$ depending on $d, p, q, R, \phi, \alpha, \epsilon, C_{E}$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_{0}$, the inequality

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq e^{c \log (c h) / \sqrt{h}}\|u\|_{\mathcal{H}(\Omega)}+c h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)}
$$

holds for all $u \in \mathcal{H}(\Omega)$ and all $1 \leq q \leq \infty$.
If there are constants $C_{E} \geq 0$ and $s \geq 1$ such that $E(k) \leq C_{E}^{k} k^{s k}$ for all $k \in \mathbb{N}$, then there
are constants $c, h_{0}>0$ depending on $d, p, q, R, \phi, \alpha, C_{E}$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_{0}$, the inequality

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq e^{-\frac{c}{h^{1}(s+1)}}\|u\|_{\mathcal{H}(\Omega)}+c h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell q(X)}
$$

holds for all $u \in \mathcal{H}(\Omega)$ and all $1 \leq q \leq \infty$.
Proof: We use Stirling's formula to estimate

$$
\frac{1}{(k-|\alpha|)!} \leq \frac{k^{|\alpha|}}{k!} \leq \frac{k^{|\alpha|} e^{k}}{k^{k}}
$$

If $\|u\|_{W_{p}^{k}(\Omega)} \leq C_{E}^{k} k^{(1-\epsilon) k}\|u\|_{\mathcal{H}(\Omega)}$ holds for all $k \in \mathbb{N}$, we can bound the first term of (4.1.3) for arbitrary $k \in \mathbb{N}$ by

$$
\left(\tilde{c} h k^{2-\epsilon}\right)^{k}\left(h^{-1} k\right)^{|\alpha|}\left(h k^{2}\right)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}+\|u\|_{\mathcal{H}(\Omega)} .
$$

We set $B=\min \left\{c_{\min }, 1 / \tilde{c}\right\}$ and choose $k \in \mathbb{N}$ such that $\frac{B}{2 k^{2}} \leq h \leq \frac{B}{k^{2}}$ holds. Then the first term can be bounded by

$$
c k^{-\epsilon k} h^{-3|\alpha| / 2}\|u\|_{\mathcal{H}(\Omega)} \leq e^{c \log (c h) / \sqrt{h}}\|u\|_{\mathcal{H}(\Omega)}
$$

where the constants $c>0$ may depend on $d, p, q, R, \phi, \alpha, C_{E}$ and $\epsilon$. With this choice of $k$ the second term of (4.1.3) can be bounded by

$$
c h^{-|\alpha|}\left(h k^{2}\right)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)} \leq c h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

If $E(k) \leq C_{E}^{k} k^{s k}$, we can bound the first term of (4.1.3) for arbitrary $k \in \mathbb{N}$ by

$$
\left(\tilde{c} h k^{1+s}\right)^{k}\left(h^{-1} k\right)^{|\alpha|}\left(h k^{2}\right)^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}+\|u\|_{\mathcal{H}(\Omega)} .
$$

We set $B=\min \left\{c_{\text {min }}, \frac{1}{e \tilde{c}}\right\}$ and choose $k \in \mathbb{N}$ such that $\frac{B}{2 k^{1+s}} \leq h \leq \frac{B}{k^{1+s}}$ holds. We point out that the condition $h \leq \frac{c_{\min }}{k^{2}}$ is satisfied since $s \geq 1$. Then $h k^{2} \leq B$, and therefore the first term can be bounded by

$$
c e^{-k} h^{-3|\alpha| / 2}\|u\|_{\mathcal{H}(\Omega)} \leq e^{-\frac{c}{h^{1 /(s+1)}}}\|u\|_{\mathcal{H}(\Omega)}
$$

where the constant $c>0$ may depend on $d, p, q, R, \phi, \alpha, C_{E}$. The second term is again bounded by

$$
c h^{-|\alpha|}\left(h k^{2}\right)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)} \leq c h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)} .
$$

Again we get the following result for the $\ell_{\infty}$-norm.
Corollary 4.1.8 If we use Corollary 4.1.6 instead of Theorem 4.1.5, we get in the case $E(k) \leq C_{E}^{k} k^{(1-\epsilon) k}$ for all $k \in \mathbb{N}$ with constants $\epsilon, c>0$, that for all $u \in \mathcal{H}(\Omega)$

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq e^{c \log (c h) / \sqrt{h}}\|u\|_{\mathcal{H}(\Omega)}+c h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

If $E(k) \leq C_{E}^{k} k^{s k}$ for all $k \in \mathbb{N}$ with a constant $s \geq 1$, we find for all $u \in \mathcal{H}(\Omega)$

$$
\left\|D^{\alpha} u\right\|_{L_{q}(\Omega)} \leq e^{-\frac{1}{h^{1 /(1+s)}}}\|u\|_{\mathcal{H}(\Omega)}+c h^{-|\alpha|}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

### 4.2 Estimates on Compact Cubes

To derive estimates for function values on a compact cube we may use the following local polynomial reproduction from [32]. From now on we shall always assume the cubes to be compact.

Theorem 4.2.1 Let $\Omega$ be a compact cube in $\mathbb{R}^{d}$. There exist constants $c_{0}, c_{2}>0$ depending only on $\Omega$ such that for every $\ell \in \mathbb{N}$ and every $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ with $h_{X, \Omega} \leq c_{0} / \ell$ we can find functions $a_{j}: \Omega \rightarrow \mathbb{R}$ satisfying

- $\sum_{j=1}^{N} a_{j}(x) p\left(x_{j}\right)=p(x)$ for all $x \in \Omega$ and all $p \in \pi_{\ell}\left(\mathbb{R}^{d}\right)$,
- $\sum_{j=1}^{N}\left|a_{j}(x)\right| \leq e^{2 d \gamma_{d}(\ell+1)}$ for all $x \in \Omega$,
- $a_{j}(x)=0$ if $\left\|x-x_{j}\right\|_{2}>c_{2} \ell h_{X, \Omega}$.

The numbers $\gamma_{d}$ are defined recursively by $\gamma_{1}=2$ and $\gamma_{d}=2 d\left(1+\gamma_{d-1}\right)$.
Lemma 4.1.1 gives in the special case $\alpha=0$ the bound

$$
\left\|u-Q^{k} u\right\|_{L_{\infty}(\mathcal{D})} \leq \frac{C_{d} d^{k}}{k!} \delta_{\mathcal{D}}^{k-d / p}|u|_{W_{p}^{k}(\mathcal{D})}
$$

where $C$ now depends only on $d$, since cubes satisfy cone conditions with a fixed angle. If we insert this estimate and the bound from Theorem4.2.1 into (4.1.1), we find

$$
\|u\|_{L_{\infty}(\mathcal{D})} \leq \frac{c^{k} \delta_{\mathcal{D}}^{k-d / p}}{k!}|u|_{W_{p}^{k}(\mathcal{D})}+c^{k}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

and

$$
\|u\|_{L_{q}(\mathcal{D})} \leq \frac{c^{k} \delta_{\mathcal{D}}^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)}}{k!}|u|_{W_{p}^{k}(\mathcal{D})}+c^{k} \delta_{\mathcal{D}}^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

The constant $c \leq C_{d} e^{4 d \gamma_{d}}$ depends now only on $d$. To derive global estimates we use the obvious covering of the big cube $\Omega$ with axially parallel small cubes $\mathcal{D}_{t}$ with the following properties.

- There are constants $D, \tilde{D}>0$ such that $\tilde{D} h k \leq \delta_{D_{t}} \leq D h k$.
- There is a constant $M_{1}=M_{1}(\theta, d)>0$ such that $\sum_{t \in T_{r}} \chi_{\mathcal{D}_{t}} \leq M_{1}$.
- There is a constant $M_{2}(\theta, d)>0$ such that $\# T_{h}<M_{2}(h k)^{-d}$.

As in the previous section, we find the following global estimate.
Theorem 4.2.2 Under the assumptions from above, there exists a positive constant $c$, which depends only on $p, q$, the side length $R$ of the cube $\Omega$ and the space dimension $d$, such that

$$
\|u\|_{L_{q}(\Omega)} \leq \frac{c^{k}}{k!}(h k)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)}+|u|_{W_{p}^{k}(\Omega)}+c^{k}(h k)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)}
$$

holds for all data sets $X \subset \Omega$ and for all $u \in W_{p}^{k}(\Omega)$ with $k>d / p$ for $1<p<\infty$ or for $k \geq d$ for $p=1$ and all $1 \leq q \leq \infty$.

As in the case of domains obeying a cone condition we get the following corollary.
Corollary 4.2.3 Under the assumptions from Theorem 4.2.2 we obtain

$$
\|u\|_{L_{q}(\Omega)} \leq \frac{c^{k}}{k!}(h k)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)_{+}}|u|_{W_{p}^{k}(\Omega)}+c^{k}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)} .
$$

Now we can derive exponential approximation orders. Again, the actual orders depend on the asymptotic behavior of the constant $E(k)$ from (4.0.1) for $k \rightarrow \infty$.

Theorem 4.2.4 If there are constants $C_{E}, \epsilon>0$ such that $E(k) \leq C_{E}^{k} k^{(1-\epsilon) k}$ for all $k \in \mathbb{N}$, then there exist constants $c, h_{0}>0$ such that for all data sets $X$ with fill distance $h \leq h_{0}$ the estimate

$$
\begin{equation*}
\|u\|_{L_{q}(\Omega)} \leq e^{c \log (c h) / h}\|u\|_{\mathcal{H}(\Omega)}+c^{1 / h}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} \tag{4.2.1}
\end{equation*}
$$

holds for all $u \in \mathcal{H}$ and all $1 \leq q \leq \infty$.
If there are constants $C_{E}>0$ and $s \geq 1$ such that $E(k) \leq C_{E}^{k} k^{s k}$ for all $k \in \mathbb{N}$, then there exist constants $c, h_{0}>0$ such that for all data sets $X$ with fill distance $h \leq h_{0}$

$$
\begin{equation*}
\|u\|_{L_{q}(\Omega)} \leq e^{c / \sqrt[s]{h}}\|u\|_{\mathcal{H}(\Omega)}+c^{1 / \sqrt[s]{h}}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} \tag{4.2.2}
\end{equation*}
$$

holds for all $u \in \mathcal{H}$ and all $1 \leq q \leq \infty$.
Remark 4.2.5 The first remark is that the condition $h<h_{0}$ ensures $c h<1$ which implies convergence.
The sampling order is in these cases improved in comparison to Theorem 4.1.7. The price is a remarkably worse Lebesgue constant. The Lebesgue constant grows even exponentially. This requires much more higher orders of consistency, i.e. a much smaller discrete residual of the reconstruction.

Proof: If $\|u\|_{W_{p}^{k}(\Omega)} \leq C_{E}^{k} k^{(1-\epsilon) k}\|u\|_{\mathcal{H}(\Omega)}$ for all $u \in \mathcal{H}(\Omega)$, Theorem 4.2.2 gives

$$
\|u\|_{L_{q}(\Omega)} \leq \frac{C^{k}}{k!}(h k)^{k-d\left(\frac{1}{p}-\frac{1}{q}\right)}+k^{(1-\epsilon) k}\|u\|_{\mathcal{H}(\Omega)}+C^{k}(h k)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)}
$$

with generic positive constants $C$. Using Stirling's formula we can bound the first term by

$$
\frac{C^{k}}{k!}(h k)^{k-d\left(\frac{1}{q}-\frac{1}{p}\right)}+k^{(1-\epsilon) k}\|u\|_{\mathcal{H}(\Omega)} \leq\left(\tilde{c} h k^{(1-\epsilon)}\right)^{k}(h k)^{-d\left(\frac{1}{q}-\frac{1}{p}\right)}+\|u\|_{\mathcal{H}(\Omega)} .
$$

If we set $B:=\min \left\{\frac{c_{0}}{2}, \frac{1}{\bar{c}}\right\}$ with the constant $c_{0}$ from the local polynomial reproduction 4.2.1 and choose $k \in \mathbb{N}$ such that $\frac{B}{2 k} \leq h \leq \frac{B}{k}$, we can further estimate

$$
\begin{align*}
\left(\tilde{c} h k^{(1-\epsilon)}\right)^{k}(h k)^{-d\left(\frac{1}{q}-\frac{1}{p}\right)}+\|u\|_{\mathcal{H}(\Omega)} & \leq k^{-\epsilon k} h^{-d\left(\frac{1}{p}-\frac{1}{q}\right)}+\|u\|_{\mathcal{H}(\Omega)} \\
& \leq e^{c \log (c h) / h}\|u\|_{\mathcal{H}(\Omega)} \tag{4.2.3}
\end{align*}
$$

By the choice of $k$ there exists a constant $c$ such that the second term is bounded by

$$
\begin{equation*}
C_{d, R}^{k}(h k)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} \leq c^{1 / h}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} . \tag{4.2.4}
\end{equation*}
$$

Adding (4.2.3) and (4.2.4) establishes estimate (4.2.1).
If $E(k) \leq C_{E}^{k} k^{s k}$ for all $k$ with a constant $s \geq 1$, we find for the first term from Theorem 4.2.2

$$
\frac{C^{k}}{k!}(h k)^{k-d\left(\frac{1}{q}-\frac{1}{p}\right)}+k^{s k}\|u\|_{\mathcal{H}(\Omega)} \leq\left(\tilde{c} h k^{s}\right)^{k}(h k)^{-d\left(\frac{1}{q}-\frac{1}{p}\right)}+\|u\|_{\mathcal{H}(\Omega)} .
$$

We set $B:=\min \left\{\frac{c_{0}}{2}, \frac{1}{e \tilde{c}}\right\}$ with the constant $c_{0}$ from the local polynomial reproduction 4.2 .1 and choose $k \in \mathbb{N}$ such that $\frac{B}{2 k^{s}} \leq h \leq \frac{B}{k^{s}}$. Note that the condition $h \leq \frac{c_{0}}{2 k}$ is satisfied. We can further estimate

$$
\begin{equation*}
\left(\tilde{c} h k^{s}\right)^{k}(h k)^{-d\left(\frac{1}{q}-\frac{1}{p}\right)}+\|u\|_{\mathcal{H}(\Omega)} \leq C e^{-k}\|u\|_{\mathcal{H}(\Omega)} \leq e^{c / \sqrt[s]{h}}\|u\|_{\mathcal{H}(\Omega)} \tag{4.2.5}
\end{equation*}
$$

By the choice of $k$ there exists a constant $c$ such that the second term is bounded by

$$
\begin{equation*}
C_{d, R}^{k}(h k)^{d / q}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} \leq c^{1 / \sqrt[s]{h}}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} \tag{4.2.6}
\end{equation*}
$$

Adding (4.2.5) and (4.2.6) establishes the second inequality (4.2.2).
We can reformulate this Theorem using the $\ell_{\infty}$-norm.
Corollary 4.2.6 Under the assumptions from Theorem 4.2 .4 with $E(k) \leq C_{E}^{k} k^{(1-\epsilon) k}$ we obtain for all $u \in \mathcal{H}(\Omega)$

$$
\|u\|_{L_{q}(\Omega)} \leq e^{c \log (c h) / h}\|u\|_{\mathcal{H}(\Omega)}+c^{1 / h}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

If $E(k) \leq C_{E}^{k} k^{s k}$, we find for all $u \in \mathcal{H}(\Omega)$

$$
\|u\|_{L_{q}(\Omega)} \leq e^{c / \sqrt[s]{h}}\|u\|_{\mathcal{H}(\Omega)}+c^{1 / \sqrt[s]{h}}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)}
$$

### 4.3 Oversampling Near the Boundary

In this section, we aim at proving better sampling rates if we have some oversampling near the boundary of an arbitrary domain. We first prove a polynomial reproduction for this situation and derive a new sampling inequality afterwards. For that, we need a generalized notation of interior points of a set, which we call the $\epsilon$-interior. Ideally we have $\operatorname{diam}(\Omega) \gg \epsilon$.

Definition 4.3.1 For any set $\Omega \subset \mathbb{R}^{d}$ and any $\epsilon>0$ we define the $\epsilon$-interior of $\Omega$ to be

$$
\Omega^{-\epsilon}:=\{x \in \Omega: \overline{B(x, \epsilon)} \subset \Omega\},
$$

where $\overline{B(x, \epsilon)}$ denotes the closed ball with center $x$ and radius $\epsilon$.

### 4.3.1 Polynomial Reproduction

In the spirit of [65, Theorem 3.8], we derive a polynomial reproduction formula on any well covered domain $\Omega \subset \mathbb{R}^{d}$.

Theorem 4.3.2 Suppose, $\Omega \subset \mathbb{R}^{d}$ is bounded and satisfies an interior cone condition with parameters $(r, \theta)$. Let $m \in \mathbb{N}, \epsilon>r$ be fixed such that diam $(\Omega) \gg \epsilon$. Set
$C_{0}:=\min \left\{\frac{r \sin \theta}{4(1+\sin \theta)},\left(1+\frac{1}{\sin \theta}\right) \sqrt{\epsilon(r+\epsilon)}, \sqrt{\frac{r}{2\left(1+\frac{1}{\sin \theta}\right)}}, \frac{\epsilon}{1+\frac{1}{\sin \theta}}, \frac{\sqrt{\epsilon(r+\epsilon)} \sin \theta}{2(\sin \theta+1)}, 1\right\}$.
Suppose $h>0$ and $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ satisfy

1. $h \leq \frac{C_{0}}{m}$,
2. For every $B(x, h) \subset \Omega$ there is a center $x_{j} \in X \cap B(x, h)$,
3. For every $B\left(x, h^{2}\right) \cap\left\{\Omega \backslash \Omega^{-2 \epsilon}\right\} \neq \emptyset$ there is a center $x_{j} \in X \cap B\left(x, h^{2}\right)$.

Then $Z=\left\{\delta_{x}: x \in X\right\}$ is a norming set for $\pi_{m}(\Omega)$, and the norm of the inverse of the associated sampling operator is bounded by 2.

Remark 4.3.3 The assumptions of Theorem 4.3.2 are satisfied if the data set $X$ has fill distance $h$ in $\Omega$ and fill distance $h^{2}$ close to the boundary, i.e., in $\Omega \backslash \Omega^{-3 \epsilon}$.

Proof: Markov's inequality for univariate polynomials $p \in \pi_{m}(\mathbb{R})$ on the compact interval $[-1,1]$ is given by [9, Theorem 5.1.8]

$$
\begin{equation*}
\left|p^{\prime}(t)\right| \leq m^{2}\|p\|_{L_{\infty}[-1,1]}, \quad t \in[-1,1] \tag{4.3.1}
\end{equation*}
$$

A simple scaling argument shows that for any $r>0$

$$
\begin{equation*}
\left|p^{\prime}(t)\right| \leq \frac{2}{r} m^{2}\|p\|_{L_{\infty}[0, r]}, \quad t \in[0, r] \tag{4.3.2}
\end{equation*}
$$

In the open interval we have the bound [9, Theorem 5.1.7]

$$
\left|p^{\prime}(t)\right| \leq \frac{m}{\sqrt{1-t^{2}}}\|p\|_{L_{\infty}[-1,1]}, \quad t \in(-1,1)
$$

which can for any $r>0$ and $\epsilon>0$ again be rewritten as

$$
\begin{equation*}
\left|p^{\prime}(t)\right| \leq \frac{m}{\sqrt{(t+\epsilon)(r+\epsilon-t)}}\|p\|_{L_{\infty}(-\epsilon, r+\epsilon)}, \quad t \in(-\epsilon, r+\epsilon) \tag{4.3.3}
\end{equation*}
$$

Suppose $p \in \pi_{m}\left(\mathbb{R}^{d}\right)$ with $\|p\|_{L_{\infty}(\Omega)}=1$. Since $\bar{\Omega}$ is compact, there exists an $x \in \bar{\Omega}$ with $|p(x)|=1$. Now we discern good and evil.
1.) Assume $x \in \Omega^{-\epsilon}$.

Since $\Omega$ satisfies an interior cone condition we can find a cone $C_{h}(x):=C(x, \xi, \theta, r)$ which is completely contained in $\Omega$. Since by assumption, $\frac{h}{\sin \theta} \leq \frac{r}{1+\sin \theta}$ we can use [65, Lemma 3.7], to find a ball $B\left(y^{(h)}, h\right) \subset C(x)$ with $y^{(h)}=x+\frac{h}{\sin \theta} \xi$. Again by
assumption we find a data point $x_{j}^{(h)} \in X \cap B(y, h)$, i.e., $\left\|y^{(h)}-x_{j}^{(h)}\right\|_{2} \leq h$ which implies by the triangle inequality $\left\|x-x_{j}^{(h)}\right\|_{2} \leq\left(1+\frac{1}{\sin \theta}\right) h$.
By [65, Lemma 3.7] and by construction the whole line segment

$$
x+t \frac{x_{j}^{(h)}-x}{\left\|x_{j}^{(h)}-x\right\|_{2}} \quad, \quad t \in[-\epsilon, r]
$$

lies in $\Omega$. There are two possibilities left.

- If the whole line segment

$$
x+t \frac{x_{j}^{(h)}-x}{\left\|x_{j}^{(h)}-x\right\|_{2}} \quad, \quad t \in[-\epsilon, r+\epsilon]
$$

lies in $\Omega$, we may employ the bound (4.3.3) to the polynomial

$$
\tilde{p}(t)=p\left(x+t \frac{x_{j}^{(h)}-x}{\left\|x_{j}^{(h)}-x\right\|_{2}}\right)
$$

which gives

$$
\begin{aligned}
\left|p(x)-p\left(x_{j}^{(h)}\right)\right| & \leq \int_{0}^{\left\|x-x_{j}^{(h)}\right\|_{2}}\left|\tilde{p}^{\prime}(t)\right| d t \\
& \leq\left\|x-x_{j}^{(h)}\right\|_{2} \cdot \max _{t \in\left[0,\left\|x-x_{j}^{(h)}\right\|_{2}\right]}\left|\tilde{p}^{\prime}(t)\right| \\
& \leq\left\|x-x_{j}^{(h)}\right\|_{2} \cdot \max _{t \in[0, r]}\left|\tilde{p}^{\prime}(t)\right| \\
& \leq\left(1+\frac{1}{\sin \theta}\right) h \cdot \frac{m}{\sqrt{\epsilon(r+\epsilon)}}\|p\|_{L_{\infty}(\Omega)} \\
& \leq \frac{1}{2} .
\end{aligned}
$$

- If the elongated line segment

$$
x+t \frac{x_{j}^{(h)}-x}{\left\|x_{j}^{(h)}-x\right\|_{2}} \quad, \quad t \in[-\epsilon, r+\epsilon]
$$

is not completely contained in $\Omega$, the point $x_{j}^{(h)}$ and $x$ must lie in $\Omega \backslash \Omega^{-(\epsilon+r)} \subset \Omega \backslash \Omega^{-2 \epsilon}$. As $h \leq 1$ we find in the cone $C(x, r, \theta, \xi) \subset C_{h}(x) \subset$ $\Omega$ an associated smaller ball $B\left(y^{\left(h^{2}\right)}, h^{2}\right) \subset C_{h^{2}}(x)$, where $y^{\left(h^{2}\right)}:=x+\frac{h^{2}}{\sin \theta} \xi$.
Since

$$
\begin{aligned}
\left\|y^{\left(h^{2}\right)}-x_{j}^{(h)}\right\|_{2} & \leq\left\|y^{\left(h^{2}\right)}-y^{(h)}\right\|_{2}+\left\|y^{(h)}-x_{j}^{(h)}\right\|_{2} \\
& \leq h\left(1+\frac{1}{\sin \theta}\right) \leq \epsilon,
\end{aligned}
$$

we have in particular $B\left(y^{\left(h^{2}\right)}, h^{2}\right) \cap\left(\Omega \backslash \Omega^{-2 \epsilon}\right) \neq \emptyset$. By assumption, we find a point $x_{j}^{\left(h^{2}\right)} \in X \cap B\left(y^{\left(h^{2}\right)}, h^{2}\right)$, i.e.,

$$
\left\|x-x_{j}^{\left(h^{2}\right)}\right\|_{2} \leq\left(1+\frac{1}{\sin \theta}\right) h^{2}
$$

In this case we apply Markov's inequality (4.3.2) to the polynomial

$$
\tilde{p}(t)=p\left(x+t \frac{x_{j}^{\left(h^{2}\right)}-x}{\left\|x_{j}^{\left(h^{2}\right)}-x\right\|_{2}}\right)
$$

which gives

$$
\begin{aligned}
\left|p(x)-p\left(x_{j}^{\left(h^{2}\right)}\right)\right| & \leq \int_{0}^{\left\|x-x_{j}^{\left(h^{2}\right)}\right\|_{2}}\left|\tilde{p}^{\prime}(t)\right| d t \\
& \leq\left\|x-x_{j}^{\left(h^{2}\right)}\right\|_{2} \frac{2}{r} m^{2}\|\tilde{p}\|_{L_{\infty}[0, r]} \\
& \leq\left(1+\frac{1}{\sin \theta}\right) h^{2} \frac{2}{r} m^{2} \\
& \leq \frac{1}{2}
\end{aligned}
$$

2.) Assume $x \in \Omega \backslash \Omega^{-\epsilon}$.

Again, we find a cone $C(x):=C(x, \xi, \theta, r)$ which contains the ball $B\left(y^{\left(h^{2}\right)}, h^{2}\right)$ with $y^{\left(h^{2}\right)}:=x+\frac{h^{2}}{\sin \theta} \xi$. Since

$$
\left\|x-y^{\left(h^{2}\right)}\right\|_{2}=\frac{h^{2}}{\sin \theta} \leq \frac{h}{\sin \theta} \leq \frac{\epsilon}{\sin \theta+1} \leq \epsilon
$$

we have $y^{\left(h^{2}\right)} \in \Omega \backslash \Omega^{-2 \epsilon}$ and can proceed as in the second case of 1.).

Note that in this case we do not easily get a local polynomial reproduction, but Theorem 3.4.2 can be applied to get the following global version. If the assumptions of Theorem 4.3.2 are satisfied, for all $x \in \Omega$ there exist numbers $a_{1}(x), \ldots, a_{N}(x)$, which do not depend on $X, h$ or $m$, such that

- $\sum_{j=1}^{N} p\left(x_{j}\right) a_{j}(x)=p(x)$ for all $x \in \Omega$ and $p \in \pi_{m}^{d}(\Omega)$,
- $\sum_{j=1}^{N}\left|a_{j}(x)\right| \leq 2$.


## Remark 4.3.4 Theorem 4.3.2 also holds true even if the domain $\Omega$ is unbounded.

Proof: For any polynomial $p$ with $\|p\|_{L_{\infty}(\Omega)}=1$ we find a point $x \in \Omega$ such that $0.99<$ $p\left(x_{0}\right) \leq 1$. Then we can proceed with this point $x$ as in the proof of Theorem 4.3.2.

Now we can proceed as normal to derive a sampling inequality.

### 4.3.2 Sampling Inequality

We again use the polynomial approximation by generalized Taylor polynomials. For the remainder $R^{k}:=u-Q^{k} u$ there is the following explicit bound, see Lemma 4.1.1.
Lemma 4.3.5 For $u \in W_{p}^{k}(\Omega)$ with $1<p<\infty$ and $k>\frac{d}{p}+1$ or with $p=1$ and $k \geq d$ we get

$$
\left\|u-Q^{k} u\right\|_{L_{\infty}(\Omega)} \leq\left(\frac{C_{d, p, \Omega}}{k}\right)^{k}|u|_{W_{p}^{k}(\Omega)}
$$

where the constant $C_{d, p, \Omega}$ depends only on $d, p$ and $\Omega$, but not on $k$.
Proof: Following Lemma 4.1.1, we denote with $\delta_{\Omega}$ the diameter of $\Omega$. Then we have with Stirling's formula

$$
\left\|u-Q^{k} u\right\|_{L_{\infty}(\Omega)} \leq C_{d, p, \theta} \frac{e^{k} d^{k}}{k!} \delta_{\Omega}^{k-d / p}|u|_{W_{p}^{k}(\Omega)} \leq\left(\frac{C_{d, p, \Omega}}{k}\right)^{k}|u|_{W_{p}^{k}(\Omega)} .
$$

Using again the estimate (4.1.1) for the special case $\alpha=0$ we immediately find the following theorem.
Theorem 4.3.6 Suppose that the discrete set $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$, and the number $h:=h_{X, \Omega} \leq \frac{c_{0}}{k}$ for some $k \in \mathbb{N}$ satisfy the assumptions of Theorem 4.3.2. Then there is a constant $C>0$ which depends only on $d, p$ and $\Omega$, but not on $k$, such that

$$
\|f\|_{L_{\infty}(\Omega)} \leq\left(\frac{C}{k}\right)^{k}|f|_{W_{p}^{k}(\Omega)}+2\left\|\left.f\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

holds for all $f \in W_{p}^{k}(\Omega)$.
We now again consider special spaces $\mathcal{H}(\Omega)$ of smooth functions $\Omega \rightarrow \mathbb{R}$.
Theorem 4.3.7 Suppose that the discrete set $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ and the number $h:=h_{X, \Omega} \leq \frac{c_{0}}{k}$ for some $k \in \mathbb{N}$ satisfy the assumptions of Theorem 4.3.2. Suppose further, that there are constants $C, \epsilon>0$, which do not depend on $k$ or $h$, such that

$$
|f|_{W_{p}^{k}(\Omega)} \leq C^{k} k^{(1-\epsilon) k}\|f\|_{\mathcal{H}(\Omega)}
$$

holds for all $k \in \mathbb{N}$. Then there are constants $c_{1}, c_{2}>0$ which depend on $d, p, \Omega$ and $\epsilon$, but not on $h$ such that

$$
\|f\|_{L_{\infty}(\Omega)} \leq e^{c_{1} \log \left(c_{2} h\right) / h}\|f\|_{\mathcal{H}(\Omega)}+2\left\|\left.f\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

holds for all $f \in \mathcal{H}(\Omega)$.
If there are constants $C, s>0$, which do not depend on $k$ or $h$, such that

$$
|f|_{W_{p}^{k}(\Omega)} \leq C^{k} k^{s k}\|f\|_{\mathcal{H}(\Omega)}
$$

holds for all $k \in \mathbb{N}$, then there is a constant $c>0$ which depends on $d, p, \Omega$ and $s$, but not on $h$ such that

$$
\|f\|_{L_{\infty}(\Omega)} \leq e^{c / h}\|f\|_{\mathcal{H}(\Omega)}+2\left\|\left.f\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

holds for all $f \in \mathcal{H}(\Omega)$.

Proof: We choose $k$ such that $\frac{c_{o}}{2 k} \leq h \leq \frac{c_{0}}{k}$. Then we may apply Theorem 4.3.6, which states for $|f|_{W_{p}^{k}(\Omega)} \leq C^{k} k^{(1-\epsilon) k}\|f\|_{\mathcal{H}(\Omega)}$ that

$$
\|f\|_{L_{\infty}(\Omega)} \leq\left(\frac{C}{k^{\epsilon}}\right)^{k}\|f\|_{\mathcal{H}(\Omega)}+2\left\|\left.f\right|_{X}\right\|_{\ell_{\infty}(X)}
$$

Now we can bound the first term using

$$
\left(C k^{-\epsilon}\right)^{k} \leq\left(\frac{2 c}{c_{0}} h\right)^{\epsilon k} \leq\left(\frac{2 c}{c_{0}} h\right)^{c_{0} \epsilon /(2 h)} \leq e^{c_{1} \log \left(c_{2} h\right) / h}
$$

with $c_{1}:=c_{0} \epsilon / 2$ and $c_{2}:=2 c / c_{0}$.
The second part can be proven similarly.
In our applications we restrict ourselves to the sampling inequalities from Theorems 4.1.7 and 4.2.4, but the results can be easily carried over to the situation with oversampling near the boundary.

### 4.4 Kernels and Native Spaces

In this section, we will provide two famous examples of Hilbert spaces of infinitely smooth functions. In the case of a positive definite radial basis function $K$ that possesses a Fourier transform $\hat{K}$, the native Hilbert space is defined as in Proposition 3.2.17, i.e.,

$$
\mathcal{N}_{K}\left(\mathbb{R}^{d}\right)=\left\{f \in C\left(\mathbb{R}^{d}\right) \cap L_{2}\left(\mathbb{R}^{d}\right):\|f\|_{\mathcal{N}_{K}}^{2}:=\int_{\mathbb{R}^{d}} \frac{|\hat{f}(\omega)|^{2}}{|\hat{K}(\omega)|} d \omega<\infty\right\}
$$

In general, the native Hilbert space $\mathcal{N}_{K}(\Omega)$ on a bounded domain $\Omega$ is defined as

$$
\begin{aligned}
& \mathcal{N}_{K}(\Omega):=\left\{f \in \mathcal{N}_{K}\left(\mathbb{R}^{d}\right):\left.f\right|_{\Omega}=0\right\}^{\perp_{\mathcal{N}_{K}}} \\
:= & \left\{\left.f\right|_{\Omega}: f \in \mathcal{N}_{K}\left(\mathbb{R}^{d}\right),(f, g)_{\mathcal{N}_{K}\left(\mathbb{R}^{d}\right)}=0 \text { for all } g \in \mathcal{N}_{K}\left(\mathbb{R}^{d}\right) \text { s.t. }\left.g\right|_{\Omega}=0\right\} .
\end{aligned}
$$

The Sobolev spaces on $\mathbb{R}^{d}$ are defined via

$$
W_{2}^{k}\left(\mathbb{R}^{d}\right):=\left\{f \in L_{2}\left(\mathbb{R}^{d}\right): \hat{f}(\cdot)\left(1+\|\cdot\|_{2}^{2}\right)^{k / 2} \in L_{2}\left(\mathbb{R}^{d}\right)\right\}
$$

This definition is equivalent to the one given in Proposition 3.3.4. Furthermore we know from [68]

$$
\frac{1}{2^{2 k}}\left\|\hat{f}(\cdot)\left(1+\|\cdot\|_{2}^{2}\right)^{k / 2}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} \leq\|u\|_{W_{2}^{K}\left(\mathbb{R}^{d}\right)} \leq\left\|\hat{f}(\cdot)\left(1+\|\cdot\|_{2}^{2}\right)^{k / 2}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)}
$$

We now show that the native spaces of Gaussian kernels consist of infinitely smooth functions in the sense of (4.0.1).

Theorem 4.4.1 For the Gaussian $K(x)=e^{-c^{2}\|x\|_{2}^{2}}$ we get for all $k \in \mathbb{N}$

$$
\|f\|_{W_{2}^{k}\left(\mathbb{R}^{d}\right)} \leq\left(\frac{8 c^{2}}{e}+2\right)^{\frac{k}{2}} k^{\frac{k}{2}}\|f\|_{\mathcal{N}_{G}\left(\mathbb{R}^{d}\right)} .
$$

Proof: By the definition of the spaces in Section 4.4 it suffices to check for this choice for $E(k)$ whether the inequality

$$
(1+r)^{k} \leq E(k)^{2} e^{\frac{r}{4 c^{2}}}
$$

holds for all $r>0$. Here we substituted $r=\|x\|_{2}^{2}$. We split our analysis into two parts. For $r \leq 1$ we see that $(1+r)^{k} \leq 2^{k}$ and $e^{\frac{r}{4 c^{2}}}>1$, hence $E(k)^{2}=2^{k}$ will work in this case. For $r>1$ things are more complicated. First we make the observation $(1+r)^{k} \leq 2^{k} r^{k}$. If we change variables via $r \mapsto 4 c^{2} r$, it remains to check

$$
2^{k}\left(4 c^{2} r\right)^{k}=\left(8 c^{2}\right)^{k}(r)^{k} \leq E(k)^{2} e^{r} \Leftrightarrow \ln \left(\left(8 c^{2}\right)^{k}\right)+k \ln r \leq r+\ln E(k)^{2} .
$$

We shall look at the function

$$
f:(1, \infty) \rightarrow \mathbb{R}_{+}, \quad r \mapsto r+\ln E(k)^{2}-\ln \left(\left(8 c^{2}\right)^{k}\right)-k \ln r
$$

and compute a minimum of this function. Easy calculation yields

$$
f^{\prime}(r)=1-\frac{k}{r} \quad, \text { and } \quad f^{\prime \prime}(r)=+\frac{k}{r^{2}} \geq 0 .
$$

Since $f^{\prime}(r)=0$ implies $r=k$ we see that the global minimum is attained for $r_{\text {min }}=k$, with $f\left(r_{\text {min }}\right)=f(k)=k+\ln E(k)^{2}-k\left(\ln \left(\left(8 c^{2}\right)\right)+\ln k\right)$. This shows that $E(k)^{2}=$ $\left(\frac{8 c^{2}}{e}\right)^{k} k^{k}$ will be sufficient.
Hence taking the maximum $E(k)^{2}=\max \left\{2^{k},\left(\frac{8 c^{2}}{e}\right)^{k} k^{k}\right\} \leq\left(\frac{8 c^{2}}{e}+2\right)^{k} k^{k}$ will ensure $f \geq 0 \Leftrightarrow\left(8 c^{2}\right)^{k} r^{k} \leq E(k)^{2} e^{r}$.

Similar considerations apply to the native spaces of inverse multiquadrics.
Theorem 4.4.2 For the inverse multiquadrics $K(x)=\left(c^{2}+\|x\|_{2}^{2}\right)^{-\beta}$ with $\beta>\frac{d}{2}$,

$$
\|f\|_{W_{2}^{k}\left(\mathbb{R}^{d}\right)} \leq \max \left\{\frac{2^{k}}{C_{1}}, c_{1}^{k} k^{2 k}\right\}^{1 / 2}\|f\|_{\mathcal{N}_{M}\left(\mathbb{R}^{d}\right)} \quad \text { for all } k \in \mathbb{N}
$$

where the constants $C_{1}, c_{1}>0$ are defined in the proof.
Proof: The Fourier transform of the inverse multiquadrics is given by [65, Theorem 6.1]

$$
\hat{K}(\omega)=\frac{2^{1-\beta}}{\Gamma(\beta)}\left(\frac{\|\omega\|_{2}}{c}\right)^{\beta-d / 2} \mathcal{K}_{d / 2-\beta}\left(c\|\omega\|_{2}\right),
$$

where $\mathcal{K}_{\sigma}$ denotes the modified Bessel function of third kind. As in the case of Gaussians, it suffices to check for this choice for $E(k)$ whether the inequality

$$
\left(1+r_{2}^{2}\right)^{k} \leq E(k)^{2} c(\beta)\left(\frac{r}{c}\right)^{\beta-d / 2} \mathcal{K}_{d / 2-\beta}(c r)
$$

holds for all $r \geq 0$. Again we split our analysis into two parts $r \in[0,1]$, and $r>1$. In the case $r \in[0,1]$ we use that

$$
g: \mathbb{R} \rightarrow \mathbb{R} \quad, r \mapsto\left(\frac{r}{c}\right)^{\beta-d / 2} \mathcal{K}_{d / 2-\beta}(c r)
$$

is a continuous mapping since the singularity of the Bessel function at 0 is of order $\beta-\frac{d}{2}$. Further, the function $g$ is nonincreasing for $r>0$ [65], so the minimum is reached for $r_{\min }^{1}=1$. This minimum is given by

$$
C_{1}:=2^{1-\beta} \Gamma(\beta) c^{d / 2-\beta} \min \left\{\mathcal{K}_{d / 2-\beta}(c), \frac{\Gamma(\beta-d / 2)}{2} 2^{\beta-d / 2}\right\}
$$

Since $\left(1+r^{2}\right)^{k} \leq 2^{k}$ for $r \in[0,1]$ we conclude that $E(k)^{2}=\frac{2^{k}}{C_{1}}$ is sufficient in the first case.
For $r>1$ we use the inequality

$$
E(k)^{2} c(\beta)\left(\frac{r}{c}\right)^{\beta-d / 2} \mathcal{K}_{d / 2-\beta}(c r) \leq E(k)^{2} c(\beta)\left(\frac{\|\omega\|_{2}}{c}\right)^{\beta-d / 2+1 / 2} e^{\frac{\|\omega\|_{2}}{c}-\frac{c(d / 2-\beta)^{2}}{2\|\omega\|_{2}}} .
$$

Since $r>1$ and $\beta>\frac{d}{2}$, we can bound the factors by

$$
e^{-\frac{(d / 2-\beta)^{2}}{2 r}} \geq e^{-\frac{(d / 2-\beta)^{2}}{2}}, \text { and }\left(\frac{r}{c}\right)^{\beta-d / 2}>c^{d / 2-\beta}
$$

If we use again $\left(1+r^{2}\right)^{k} \leq 2^{k} r^{2 k}$, we have to determine $E(k)$ such that for all $r>1$

$$
f(r):=\ln C+\ln E(k)^{2}+r-2 k(\ln (2) / 2+\ln r) \geq 0
$$

Easy calculation yields $f^{\prime}(r)=-\frac{2 k}{r}+1$, and $f^{\prime \prime}(r)=\frac{2 k}{r^{2}}$. This gives $r_{\text {min }}=2 k$, and

$$
f\left(r_{\min }\right)=\ln C+\ln E(k)^{2}+k(2-3 \ln (2))-2 k \ln (k)
$$

Hence, for $\ln E(k)^{2}=c_{1} k+2 k \ln k$ we can show that $f$ is positive. This yields that a constant $E(k)^{2}=\max \left\{\frac{2^{k}}{C_{1}}, c_{1}^{k} k^{2 k}\right\}$ will ensure the claim.

The constant $2^{2 k}$ is absorbed into the $C_{E}^{k}$ term of $E(k)$ in Theorems 4.1.7 and 4.2.4.
For the special case of the Gaussian kernel $G(x)=e^{-c^{2}\|x\|_{2}^{2}}$ this yields

$$
\mathcal{N}_{G}\left(\mathbb{R}^{d}\right)=\left\{f \in C\left(\mathbb{R}^{d}\right) \cap L_{2}\left(\mathbb{R}^{d}\right): \int_{\mathbb{R}^{d}}|\hat{f}(\omega)|^{2} e^{\frac{\|x\|_{2}^{2}}{4 c^{2}}} d \omega<\infty\right\}
$$

Theorem 4.4.1 shows, that for all $k \geq 0$

$$
\mathcal{N}_{G}\left(\mathbb{R}^{d}\right) \subset W_{2}^{k}\left(\mathbb{R}^{d}\right) \text { with }\|f\|_{W_{2}^{k}\left(\mathbb{R}^{d}\right)} \leq C_{E}^{k} k^{k / 2}\|f\|_{\mathcal{N}_{G}\left(\mathbb{R}^{d}\right)}
$$

Using [65, Theorem 10.46], every $f \in \mathcal{N}_{G}(\Omega)$ has an extension $\mathcal{E} f \in \mathcal{N}_{G}\left(\mathbb{R}^{d}\right)$ with $\|\mathcal{E} f\|_{\mathcal{N}_{G}\left(\mathbb{R}^{d}\right)} \leq\|f\|_{\mathcal{N}_{G}(\Omega)}$. Thus we have for $f=\left.\mathcal{E} f\right|_{\Omega} \in \mathcal{N}_{G}(\Omega) \subset W_{2}^{k}(\Omega)$ for all non-negative $k$

$$
\begin{equation*}
\|f\|_{W_{2}^{k}(\Omega)} \leq\|\mathcal{E} f\|_{W_{2}^{k}\left(\mathbb{R}^{d}\right)} \leq C_{E}^{k} k^{k / 2}\|\mathcal{E} f\|_{\mathcal{N}_{G}\left(\mathbb{R}^{d}\right)} \leq C_{E}^{k} k^{k / 2}\|f\|_{\mathcal{N}_{G}(\Omega)} \tag{4.4.1}
\end{equation*}
$$

Similar considerations apply to the inverse multiquadrics

$$
K_{M}(x)=\left(c^{2}+\|x\|_{2}^{2}\right)^{-\beta}, \text { for } \beta>d / 2
$$

The essentially same argument as above (see Theorem 4.4.2) leads to

$$
\begin{equation*}
\mathcal{N}_{M}\left(\mathbb{R}^{d}\right) \subset W_{2}^{\tau}\left(\mathbb{R}^{d}\right), \text { and }\|f\|_{W_{2}^{\tau}(\Omega)} \leq C_{E}^{k} k^{k}\|f\|_{\mathcal{N}_{M}(\Omega)} \tag{4.4.2}
\end{equation*}
$$

Applying Corollary 4.2.6 immediately yields
Corollary 4.4.3 Under the assumptions from Theorem 4.2.4 and with the constant for the Gaussian $E_{(G)}(k)=C_{E}^{k} k^{k / 2}$ (see Theorem 4.4.1), we obtain for all $u \in \mathcal{N}_{G}(\Omega)$

$$
\|u\|_{L_{q}(\Omega)} \leq e^{c \log (c h) / h}\|u\|_{\mathcal{N}_{G}(\Omega)}+c^{1 / h}\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)} .
$$

Analogously with $E_{(M)}(k) \leq C_{E}^{k} k^{k}$ and $s=1$ (see Appendix, Theorem 4.4.2) we find for all $u \in \mathcal{N}_{M}(\Omega)$

$$
\|u\|_{L_{q}(\Omega)} \leq e^{c / h}\|u\|_{\mathcal{N}_{M}(\Omega)}+c^{1 / h}\left\|\left.u\right|_{X}\right\|_{\ell_{q}(X)} .
$$

### 4.5 Applications to Smoothed Interpolation

We shall apply our general results in the case of -possibly regularized- kernel based interpolation, as introduced in Section 2.4. To start with, we briefly recall the situation of symmetric reconstruction methods, as outlined in Section 2.3.2. One is given centers $X=\left\{x_{1}, \ldots x_{N}\right\} \subset \mathbb{R}^{d}$ and data $\left(f_{1}, \ldots f_{N}\right)^{T} \in \mathbb{R}^{N}$ generated by an unknown function $u \in \mathcal{N}_{K}(\Omega)$. One has to solve the system

$$
\begin{equation*}
(\mathbf{K}+\lambda \mathbf{I d}) b=\left.f\right|_{X}, \tag{4.5.1}
\end{equation*}
$$

with $\mathbf{K}=\left(K\left(x_{i}-x_{j}\right)\right)_{i, j=1 \ldots N}$ to build an approximant

$$
s_{\lambda, X, K}(f)(\cdot):=\sum_{x_{j} \in X} b_{j} K\left(\cdot, x_{j}\right) .
$$

We point out that the classical interpolant is a special case, namely for choosing $\lambda=0$. It is known [67] that

$$
\begin{aligned}
\left\|s_{\lambda, X, K}(f)\right\|_{\mathcal{N}_{K}(\Omega)} & \leq 2\|f\|_{\mathcal{N}_{K}(\Omega)} \\
\left\|\left.s_{\lambda, X, K}(f)\right|_{X}-\left.f\right|_{X}\right\|_{\ell_{\infty}(X)} & \leq \sqrt{\lambda}\|f\|_{\mathcal{N}_{K}(\Omega)}
\end{aligned}
$$

holds.

Theorem 4.5.1 If $\Omega$ is a compact cube, then there exists a constant $h_{0}>0$ and constants $c, \tilde{c}, \bar{c}>0$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_{0}$ we get for Gaussian kernels

$$
\left\|f-s_{\lambda, X, K}(f)\right\|_{L_{q}(\Omega)} \leq 3\left(e^{c \log (\tilde{c} h) / h}+\sqrt{\lambda} \bar{c}^{1 / h}\right)\|f\|_{\mathcal{N}_{G}(\Omega)}
$$

for all $f \in \mathcal{N}_{G}(\Omega)$ and $1 \leq q \leq \infty$.
For the inverse multiquadrics we find

$$
\left\|f-s_{\lambda, X, K}(f)\right\|_{L_{q}(\Omega)} \leq 3\left(e^{c / h}+\sqrt{\lambda} c^{1 / h}\right)\|f\|_{\mathcal{N}_{M}(\Omega)}
$$

for all $f \in \mathcal{N}_{M}(\Omega)$ and $1 \leq q \leq \infty$.

Remark 4.5.2 In the case $\lambda=0$, i.e., the standard interpolation we obtain the well known orders for interpolation with Gaussian and inverse multiquadric kernels, respectively [65].

Proof: Proof of Theorem 4.5.1: For the Gaussian kernel we have

$$
\begin{aligned}
\left\|f-s_{\lambda, X, K}(f)\right\|_{L_{q}(\Omega)} & \leq e^{c \log (\tilde{c} h) / h}\left\|f-s_{\lambda, X, K}(f)\right\|_{\mathcal{N}_{G}(\Omega)} \\
& +\bar{c}^{1 / h}\left\|\left.s_{\lambda, X, K}(f)\right|_{X}-\left.f\right|_{X}\right\|_{L_{\infty}(X)} \\
& \leq 3\left(e^{c \log (\tilde{c} h) / h}+\sqrt{\lambda} \bar{c}^{1 / h}\right)\|f\|_{\mathcal{N}_{G}(\Omega)}
\end{aligned}
$$

and analogous considerations apply to the inverse multiquadrics.
From here on we restrict ourselves to the case of Gaussian kernels since all results can be carried over to inverse multiquadrics easily. We write abbreviately $\mathcal{N}_{K}:=\mathcal{N}_{G}$.
Corollary 4.5.3 For the choice $\lambda \leq e^{\frac{2(\operatorname{cog}(\bar{c} h)-\log (\bar{c}))}{h}}$ we get

$$
\left\|f-s_{\lambda, X, K}(f)\right\|_{L_{q}(\Omega)} \leq 3 e^{c \log (\tilde{c} h) / h}\|f\|_{\mathcal{N}_{K}(\Omega)}
$$

Theorem 4.5.4 If the domain $\Omega$ satisfies an interior cone condition, there exist constants $A, B, C, c_{\theta}, h_{0}$ such that for all data sets $X \subset \Omega$ with fill distance $h \leq h_{0}$ and $1 \leq q \leq \infty$

$$
\left\|D^{\alpha}\left(f-s_{\lambda, X, K}(f)\right)\right\|_{L_{q}(\Omega)} \leq\left(A e^{B \log (C h) / \sqrt{h}}+\sqrt{\lambda} c_{\theta} h^{-|\alpha|}\right)\|f\|_{\mathcal{N}_{K}(\Omega)}
$$

Here the constants $A, B, C, c_{\theta}$ depend only on $\theta, \alpha, d, q$.
Corollary 4.5.5 For the choice $\lambda \leq A e^{-2 B \log (C h) / \sqrt{h}} h^{2|\alpha|}$ we get

$$
\left\|f-s_{\lambda, X, K}(f)\right\|_{L_{q}(\Omega)} \leq 3 A e^{B \log (C h) / \sqrt{h}}\|f\|_{\mathcal{N}_{K}(\Omega)}
$$

This shows that we can improve the condition number of the system (4.5.1) at least to the value of $\lambda=A e^{-2 B \log (C h) / \sqrt{h}} h^{2|\alpha|}$ instead of $e^{-1 / q^{2}}$ for the Gaussian kernel and still get good approximation orders. We point out that we get rid of the separation distance $q:=q_{X}:=\frac{1}{2} \min _{1 \leq i<j \leq N}\left\|x_{j}-x_{i}\right\|_{2}$, which can spoil the condition number in case of badly distributed points.
In the next section we present a more general application of sampling inequalities in the deterministic error analysis of kernel based machine learning algorithms.

## Chapter 5

## Kernel Based Learning

Support Vector (SV) machines and related kernel-based algorithms are modern learning systems motivated by results of statistical learning theory [57]. The concept of SV machines is to provide a prediction function that is accurate on the given training data, and that is sparse in the sense that it can be written in terms of a typically small subset of all samples, called the support vectors [50]. Therefore, SV regression and classification algorithms are closely related to regularized problems from classical approximation theory [23], and techniques from functional analysis were applied to derive probabilistic error bounds for SV regression [17].
This chapter provides a theoretical framework to derive deterministic error bounds for some popular SV machines. We show how a sampling inequality from [67] can be used to bound the worst-case generalization error for the $\nu$ - and the $\epsilon$-regression without making any statistical assumptions on the inaccuracy of the training data. In contrast to the literature, our error bounds explicitly depend on the pointwise noise in the data. Thus they can be used for any subsequent probabilistic analysis modelling certain assumptions on the noise distribution.
This chapter is organized as follows. Section 5.1 deals with regularized approximation problems in Hilbert spaces with reproducing kernels and outlines the connection to classical SV regression (SVR) algorithms. We provide a deterministic error analysis for the $\nu$ - and the $\epsilon$-SVR for both exact and inexact training data. Our analytical results showing optimal convergence orders in Sobolev spaces are confirmed by numerical experiments.

### 5.1 Regularized Problems in Native Hilbert Spaces

In native Hilbert spaces for kernels we consider the following learning or recovery problem. We assume that we are given (possibly only approximate) function values $y_{1}, \ldots, y_{N} \in \mathbb{R}$ of an unknown function $f \in \mathcal{N}_{K}$ on some scattered points $X:=\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right\} \subset \Omega$, i.e., $f\left(\mathbf{x}^{(j)}\right) \approx y_{j}$ for $j=1, \ldots, N$. We point out that we shall use a slightly different notation in this section, which comes from the machine learning literature. In particular, bold letters denote vectors, i.e., $\mathbf{v}=\left(v_{1}, \ldots, v_{d}\right)^{T} \in \mathbb{R}^{d}$. The character $C$ does not denote a generic constant but a fixed parameter of the optimization problems. Generic constants are denoted by $\tilde{C}$ or different letters.

To control accuracy and complexity of the reconstruction simultaneously, we use the optimization problem

$$
\begin{equation*}
\min _{\substack{s \in \mathcal{N}_{K}(\Omega) \\ \epsilon \in \mathbb{R}^{+}}} \frac{1}{N} \sum_{j=1}^{N} V_{\epsilon}\left(\left|s\left(\mathbf{x}^{(j)}\right)-y_{j}\right|\right)+\frac{1}{2 C}\|s\|_{\mathcal{N}_{K}}^{2}, \tag{5.1.1}
\end{equation*}
$$

where $C>0$ is a positive parameter, and $V_{\epsilon}$ denotes a positive function, which may be parametrized by a positive real number $\epsilon$. In this section $C$ always denotes a positive parameter rather than a constant. We point out that $V_{\epsilon}$ need not be a classical loss function. Therefore we shall give some proofs of results that are well known in the case of $V_{\epsilon}$ being a loss function [51] .

Theorem 5.1.1 (Representer theorem) If $\left(s_{X, \mathbf{y}}, \epsilon^{*}\right)$ is a solution of the optimization problem (5.1.1), then there exists a vector $\mathbf{w} \in \mathbb{R}^{N}$ such that

$$
s_{X, \mathbf{y}}(\cdot)=\sum_{j=1}^{N} w_{j} K\left(\mathbf{x}^{(j)}, \cdot\right),
$$

that is $s_{X, \mathbf{y}} \in \operatorname{span}\left\{K\left(\mathbf{x}^{(1)}, \cdot\right), \ldots, K\left(\mathbf{x}^{(N)}, \cdot\right)\right\}$.
Proof: Every $s \in \mathcal{N}_{K}(\Omega)$ can be decomposed into two parts

$$
s=s_{\|}+s_{\perp},
$$

where $s_{\|}$is contained in the linear span of $\left\{K\left(\mathbf{x}^{(1)}, \cdot\right), \ldots, K\left(\mathbf{x}^{(N)}, \cdot\right)\right\}$, and $s_{\perp}$ is contained in the orthogonal complement, i.e., $\left\langle s_{\|}, s_{\perp}\right\rangle_{\mathcal{N}_{K}}=0$. By the reproducing property of the kernel $K$ in the native space we have

$$
s\left(\mathbf{x}^{(j)}\right)=\left\langle s_{\|}+s_{\perp}, K\left(\mathbf{x}^{(j)}, \cdot\right)\right\rangle_{\mathcal{N}_{K}}=\left\langle s_{\|}, K\left(\mathbf{x}^{(j)}, \cdot\right)\right\rangle_{\mathcal{N}_{K}} .
$$

Using this identity the problem (5.1.1) can be rewritten as

$$
\min _{\substack{s=s_{\|}+s_{\perp} \\ \epsilon \in \mathbb{R}^{+}}} \frac{1}{N} \sum_{j=1}^{N} V_{\epsilon}\left(\left|\left\langle s_{\|}, K\left(\mathbf{x}^{(j)}, \cdot\right)\right\rangle-y_{j}\right|\right)+\frac{1}{2 C}\left\|s_{\|}\right\|_{\mathcal{N}_{K}}^{2}+\frac{1}{2 C}\left\|s_{\perp}\right\|_{\mathcal{N}_{K}}^{2} .
$$

Therefore a solution $\left(s_{X, \mathbf{y}}, \epsilon^{*}\right)$ of the optimization problem (5.1.1) satisfies $\left(s_{X, \mathbf{y}}\right)_{\perp}=0$, which implies $s_{X, \mathbf{y}} \in \operatorname{span}\left\{K\left(\mathbf{x}^{(1)}, \cdot\right), \ldots, K\left(\mathbf{x}^{(N)}, \cdot\right)\right\}$.

Since the proof of Theorem 5.1.1 does not depend on the minimality with respect to $\epsilon$ this result holds also true if $\epsilon$ is a fixed parameter instead of a primal variable. To be precise we state this result as a corollary.

Corollary 5.1.2 If $s_{X, \mathrm{y}}$ is a solution of the optimization problem

$$
\begin{equation*}
\min _{s \in \mathcal{N}_{K}(\Omega)} \frac{1}{N} \sum_{j=1}^{N} V_{\epsilon}\left(\left|s\left(\mathbf{x}^{(j)}\right)-y_{j}\right|\right)+\frac{1}{2 C}\|s\|_{\mathcal{N}_{K}}^{2}, \tag{5.1.2}
\end{equation*}
$$

with $\epsilon \in \mathbb{R}^{+}$being a fixed parameter, then there exists a vector $\mathbf{w} \in \mathbb{R}^{N}$ such that

$$
s_{X, \mathbf{y}}(\cdot)=\sum_{j=1}^{N} w_{j} K\left(\mathbf{x}^{(j)}, \cdot\right)
$$

i.e., $s_{X, \mathbf{y}} \in \operatorname{span}\left\{K\left(\mathbf{x}^{(1)}, \cdot\right), \ldots, K\left(\mathbf{x}^{(N)}, \cdot\right)\right\}$.

We shall use the representer theorem to reformulate infinite-dimensional optimization problems of the forms (5.1.1) or (5.1.2) in a finite-dimensional setting.

### 5.2 Support Vector Regression

As a first optimization problem of the form (5.1.1) we consider the $\nu$-SVR introduced in [52]. The function

$$
V_{\epsilon}(\mathbf{x})=|\mathbf{x}|_{\epsilon}+\epsilon \nu
$$

is related to Vapnik's $\epsilon$-intensive loss function [57]

$$
|\mathbf{x}|_{\epsilon}=\left\{\begin{array}{cc}
0 & \text { if }|\mathbf{x}| \leq \epsilon \\
|\mathbf{x}|-\epsilon & \text { if }|\mathbf{x}|>\epsilon
\end{array}\right.
$$

but has an additional term with a positive parameter $\nu$. The associated optimization problem takes the form

$$
\begin{equation*}
\min _{\substack{s \in \mathcal{N}_{K}(\Omega) \\ \epsilon \in \mathbb{R}^{+}}} \frac{1}{N} \sum_{j=1}^{N}\left|s\left(\mathbf{x}^{(j)}\right)-y_{j}\right|_{\epsilon}+\epsilon \nu+\frac{1}{2 C}\|s\|_{\mathcal{N}_{K}}^{2} \tag{5.2.1}
\end{equation*}
$$

Theorem 5.2.1 The optimization problem (5.2.1) possesses a solution $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$.
Proof: The problem (5.2.1) is equivalent to the optimization problem

$$
\begin{equation*}
\min _{\substack{s \in \mathcal{N}_{K}(\Omega) \\ \delta \in \mathbb{R}}} \frac{1}{N} \sum_{j=1}^{N}\left|s\left(\mathbf{x}^{(j)}\right)-y_{j}\right|_{\delta^{2}}+\delta^{2} \nu+\frac{1}{2 C}\|s\|_{\mathcal{N}_{K}}^{2} \tag{5.2.2}
\end{equation*}
$$

If we set $\mathcal{H}:=\mathcal{N}_{K} \times \mathbb{R}$, we can define an inner product on $\mathcal{H}$ by

$$
\left\langle h_{1}, h_{2}\right\rangle_{\mathcal{H}}:=\left\langle f_{1}, f_{2}\right\rangle_{\mathcal{N}_{K}}+2 C \nu\left\langle r_{1}, r_{2}\right\rangle_{\mathbb{R}}
$$

for $h_{j}=\left(f_{j}, r_{j}\right), j=1,2$. To make $\mathcal{H}$ a space of functions we use the canonical identification of $\mathbb{R}$ with the space of all constant functions $\mathbb{R} \rightarrow \mathbb{R}$. The Hilbert space $\mathcal{H}$ then has the reproducing kernel $\tilde{K}:=\left(K, \frac{1}{2 C \nu} \mathbf{1}\right)$ where $\mathbf{1}$ denotes the constant function that maps everything to 1 , that is $\tilde{K}((\mathbf{x}, r),(\mathbf{y}, s))=K(\mathbf{x}, \mathbf{y})+1 /(2 C \nu)$ for all $r, s \in \mathbb{R}$. With this notation the problem (5.2.2) can be rewritten as

$$
\begin{equation*}
\min _{(s, \delta) \in \mathcal{H}} Q^{\mathbf{y}}\left(I_{X}(s, \delta)\right)+\frac{1}{2 C}\|(s, \delta)\|_{\mathcal{H}}^{2} \tag{5.2.3}
\end{equation*}
$$

where

$$
I_{X}(s, \delta):=\left(s\left(\mathbf{x}^{(1)}\right), \ldots, s\left(\mathbf{x}^{(N)}\right), \delta\right)^{T} \in \mathbb{R}^{N+1}
$$

and

$$
Q^{\mathbf{y}}: \mathbb{R}^{N+1} \rightarrow \mathbb{R}, \quad Q^{\mathbf{y}}(\mathbf{p}, \delta)=\frac{1}{N} \sum_{j=1}^{N}\left|p_{j}-y_{j}\right|_{\delta^{2}}
$$

Since $Q^{\mathbf{y}}$ is continuous on $\mathbb{R}^{N+1}$ for all $\mathbf{y} \in \mathbb{R}^{N}$, the problem (5.2.3) possesses a solution [34].

If we introduce the slack variables $\xi, \xi^{*} \in \mathbb{R}^{N}$, the representer theorem gives us an equivalent finite-dimensional problem considered in [52].

$$
\begin{align*}
& \min _{\substack{\mathbf{w} \in \mathbb{R}^{\mathbb{N}} \\
\xi^{*}, \xi \in \mathbb{R}^{\mathbf{N}} \\
\epsilon \in \mathbb{R}^{+}}} \frac{1}{2} \mathbf{w}^{T} \mathbf{K} \mathbf{w}+C \cdot\left(\nu \epsilon+\frac{1}{N} \sum_{j=1}^{N}\left(\xi_{j}+\xi_{j}^{*}\right)\right) \\
& \text { subject to } \quad(\mathbf{K w})_{j}-y_{j} \leq \epsilon+\xi_{j}, \\
&(-\mathbf{K w})_{j}+y_{j} \leq \epsilon+\xi_{j}^{*}, \\
& \xi_{j}^{*}, \xi_{j} \geq 0, \epsilon \geq 0 \quad \text { for } 1 \leq j \leq N, \tag{5.2.4}
\end{align*}
$$

where

$$
\mathbf{K}=\left(K\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right)\right)_{i, j=1 \ldots N}
$$

denotes the Gram matrix of the kernel $K$. We will use this equivalent problem for implementation and our numerical tests.
A particularly interesting problem arises if we skip the parameter $\nu$, and let $\epsilon$ be fixed. Then the optimization problem (5.2.4) takes the form

$$
\begin{align*}
\min _{\substack{\mathbf{w} \in \mathbb{R}^{\mathbf{N}} \\
\xi^{*}, \xi \in \mathbb{R}^{\mathbf{N}}}} \frac{1}{2} \mathbf{w}^{T} \mathbf{K} \mathbf{w}+C & \cdot \frac{1}{N} \sum_{j=1}^{N}\left(\xi_{j}+\xi_{j}^{*}\right) \\
\text { subject to }(\mathbf{K w})_{j}-y_{j} & \leq \epsilon+\xi_{j}, \\
(-\mathbf{K w})_{j}+y_{j} & \leq \epsilon+\xi_{j}^{*}, \\
\xi_{j}^{*}, \xi_{j} & \geq 0 \text { for } 1 \leq j \leq N . \tag{5.2.5}
\end{align*}
$$

This problem is called $\epsilon$-SVR [52]. Similarly to the $\nu$-SVR, the problem (5.2.5) can be formulated as a regularized minimization problem in a Hilbert space [17], namely

$$
\begin{equation*}
\min _{s \in \mathcal{N}_{K}(\Omega)} \frac{1}{N} \sum_{j=1}^{N}\left|s\left(\mathbf{x}^{(j)}\right)-y_{j}\right|_{\epsilon}+\frac{1}{2 C}\|s\|_{\mathcal{N}_{K}(\Omega)}^{2} . \tag{5.2.6}
\end{equation*}
$$

Like the $\nu$-SVR, this optimization problem possesses a solution [34, Lemma 1].

### 5.3 A Sampling Inequality

We shall employ a special case of a sampling inequality introduced in [67]. It requires the following assumptions, which we need from now on. Let $\Omega \subset \mathbb{R}^{d}$ be a bounded domain with Lipschitz boundary that satisfies an interior cone condition. We recall the Definition 3.3.5. A domain $\Omega$ is said to satisfy an interior cone condition with radius $r$ and angle $\theta$ if for every $\mathrm{x} \in \Omega$ there is a unit vector $\xi(\mathrm{x})$ such that the cone

$$
C(\mathbf{x}, \xi(\mathbf{x}), \theta, r):=\left\{\mathbf{x}+\lambda \mathbf{y}: \mathbf{y} \in \mathbb{R}^{d},\|\mathbf{y}\|_{2}=1, \mathbf{y}^{T} \xi(\mathbf{x}) \geq \cos (\theta), \lambda \in[0, r]\right\}
$$

is contained in $\Omega$. In particular, a domain that satisfies an interior cone condition cannot have any outward pointing cusps. We shall assume for the rest of this chapter that $\Omega$ satisfies an interior cone condition with radius $R_{\max }$ and angle $\theta$. We will need a slightly unhandy constant, which depends only on the geometry of $\Omega$, namely

$$
\begin{equation*}
C_{\Omega}:=\frac{\sin \left(2 \arcsin \left(\frac{\sin \theta}{4(1+\sin \theta)}\right)\right) \sin \theta}{8\left(1+\sin \left(2 \arcsin \left(\frac{\sin \theta}{4(1+\sin \theta)}\right)\right)\right)(1+\sin \theta)} R_{\max } . \tag{5.3.1}
\end{equation*}
$$

Suppose that $K$ is a radial kernel function such that the native Hilbert space of $K$ is a Sobolev space, i.e., $\mathcal{N}_{K}=W_{2}^{\tau}(\Omega)$. Here we assume that $\lfloor\tau-1\rfloor>d / 2$, where we use the notation $\lfloor t\rfloor:=\max \left\{n \in \mathbb{N}_{0}: n \leq t\right\}$ for $t \geq 0$. Furthermore, let $X=$ $\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right\} \subset \Omega$ be a discrete set with sufficiently small fill distance. Let us explain the relation to the usual error bounds in terms of the number of points $N$. In the case of regularly distributed points we have [65] that $h=N^{-\frac{1}{d}}$. Therefore the limit $h \rightarrow 0$ is equivalent to the limit $N \rightarrow \infty$, which is the more intuitive meaning of asymptotic convergence. However, there is a drawback, since the error bounds in terms of $N$ depend on the space dimension $d$, while error bounds in terms of the fill distance $h$ depend only on the smoothness of the function to be learned. We will comment on this again later for the special error bounds we consider here. We shall use the following result from [67].

Theorem 5.3.1 Suppose $\Omega \subset \mathbb{R}^{d}$ is a bounded domain with Lipschitz boundary that satisfies an interior cone condition. Let $\tau$ be a positive real number with $\lfloor\tau-1\rfloor>\frac{d}{2}$ and $1 \leq q \leq \infty$. Then there exists a positive constant $C>0$ such that for all discrete sets $X \subset \Omega$ with sufficiently small fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$ the inequality

$$
\|u\|_{L_{q}(\Omega)} \leq C \cdot\left(h^{\tau-d(1 / 2-1 / q)}+\|u\|_{W_{2}^{\tau}(\Omega)}+\left\|\left.u\right|_{X}\right\|_{\ell_{\infty}(X)}\right)
$$

holds for all $u \in W_{2}^{\tau}(\Omega)$, where we use the notation $(t)_{+}:=\max \{0, t\}$.
We shall apply this theorem to the residual function $f-s_{X, \mathrm{y}}$ of the function $f \in W_{2}^{\tau}(\Omega)$ to be recovered and a solution $s_{X, \mathbf{y}} \in W_{2}^{\tau}(\Omega)$ of the regression problem. In our applications we shall focus on the two main cases $q=\infty$ and $q=2$. Other cases can be treated analogously. It will turn out that we get optimal convergence rates in the noiseless case. In presence of noise, the resulting error will explicitly be bounded in terms of the noise in the data.

## $5.4 \quad \nu$-SVR with Exact Data

To derive error bounds for the $\nu$-SVR optimization problem (5.2.1) we shall apply Theorem 5.3 .1 to the residual $f-s_{X, \mathbf{y}}^{(\nu)}$, where $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ denotes a solution to the problem (5.2.1) for $X:=\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right\} \subset \Omega$ and $\mathbf{y} \in \mathbb{R}^{N}$. In this section we consider exact data, i.e.,

$$
\begin{equation*}
f\left(\mathbf{x}^{(j)}\right)=y_{j} \quad \text { for } j=1, \ldots, N \tag{5.4.1}
\end{equation*}
$$

for a function $f \in W_{2}^{\tau}(\Omega) \cong \mathcal{N}_{K}$. As pointed out by [67] we first need two stability estimates for the solution $s_{X, \mathbf{y}}^{(\nu)}$.

Lemma 5.4.1 Under the assumption (5.4.1) on the data, we find that for every $X$ a solution $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ to problem (5.2.1) satisfies

$$
\begin{aligned}
\left\|s_{X, \mathbf{y}}^{(\nu)}\right\|_{\mathcal{N}_{K}(\Omega)} & \leq\|f\|_{\mathcal{N}_{K}} \text {, and } \\
\left\|s_{X, \mathbf{y}}^{(\nu)} \mid X-\mathbf{y}\right\|_{\ell_{\infty}(X)} & \leq \frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}+\epsilon^{*} \cdot(1-N \nu)
\end{aligned}
$$

Proof: We denote the objective function of the optimization problem (5.2.1) by

$$
\begin{equation*}
H_{C, \nu}^{\mathbf{y}}(s, \epsilon):=\frac{1}{N} \sum_{j=1}^{N}\left|s\left(\mathbf{x}^{(j)}\right)-y_{j}\right|_{\epsilon}+\nu \epsilon+\frac{1}{2 C}\|s\|_{\mathcal{N}_{K}}^{2} \tag{5.4.2}
\end{equation*}
$$

and the interpolant to $f$ with respect to $X$ and $K$ with $I_{f}$, i.e., $\left.I_{f}\right|_{X}=\mathbf{y}$. With this notation we have

$$
\frac{1}{2 C}\left\|s_{X, \mathbf{y}}^{(\nu)}\right\|_{\mathcal{N}_{K}}^{2} \leq H_{C, \nu}^{\mathbf{y}}\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right) \leq H_{C, \nu}^{\mathrm{y}}\left(I_{f}, 0\right) \leq \frac{1}{2 C}\left\|I_{f}\right\|_{\mathcal{N}_{K}}^{2} \leq \frac{1}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}
$$

since $\left\|I_{f}\right\|_{\mathcal{N}_{K}} \leq\|f\|_{\mathcal{N}_{K}}$ [65], which implies the first claim. Furthermore we have for $i=1, \ldots, N$

$$
\begin{aligned}
\left|s_{X, \mathbf{y}}^{(\nu)}\left(\mathbf{x}^{(i)}\right)-y_{i}\right| & \leq \sum_{j=1}^{N}\left|s_{X, \mathbf{y}}^{(\nu)}\left(\mathbf{x}^{(j)}\right)-y_{j}\right|_{\epsilon^{*}}+\epsilon^{*} \leq N H_{C, \nu}^{\mathbf{y}}\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)+\epsilon^{*}(1-N \nu) \\
& \leq N H_{C, \nu}^{\mathbf{y}}\left(I_{f}, 0\right)+\epsilon^{*}(1-N \nu) \leq \frac{N}{2 C}\left\|I_{f}\right\|_{\mathcal{N}_{K}}^{2}+\epsilon^{*}(1-N \nu) \\
& \leq \frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}+\epsilon^{*}(1-N \nu)
\end{aligned}
$$

which finishes the proof.
With Theorem 5.3.1 we find immediately the following result.
Theorem 5.4.2 Suppose $\Omega \subset \mathbb{R}^{d}$ is a bounded domain with Lipschitz boundary that satisfies an interior cone condition. Let $\tau$ be a positive real number with $\lfloor\tau-1\rfloor>\frac{d}{2}$ and $1 \leq q \leq \infty$. We suppose $f \in W_{2}^{\tau}(\Omega)$ with $f\left(\mathbf{x}^{(i)}\right)=y_{i}$. Let $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ be a solution of
the $\nu$-SVR. Then there is a constant $\tilde{C}>0$, which depends on $\tau, d$ and $\Omega$, but not on $f$ or $X$, such that the approximation error can be bounded by

$$
\left\|f-s_{X, y}^{(\nu)}\right\|_{L_{q}(\Omega)} \leq \tilde{C}\left(2 h^{\tau-d(1 / 2-1 / q)+}\|f\|_{W_{2}^{\tau}(\Omega)}+\frac{N}{2 C}\|f\|_{W_{2}^{\tau}(\Omega)}^{2}+(1-N \nu) \cdot \epsilon^{*}\right)
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$
Proof: Combining Lemma 5.4.1 and Theorem 5.3.1 leads to

$$
\begin{aligned}
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{q}(\Omega)} & \leq \tilde{C}\left(h^{\tau-d(1 / 2-1 / q)+}\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{W_{2}^{\tau}(\Omega)}+\left\|\mathbf{y}-s_{X, \mathbf{y}}^{(\nu)} \mid X\right\|_{\ell_{\infty}(X)}\right) \\
& \leq \tilde{C}\left(h^{\tau-(d / 2-d / q)+}\left(\|f\|_{W_{2}^{\tau}(\Omega)}+\left\|s_{X, \mathbf{y}}^{(\nu)}\right\|_{W_{2}^{\tau}(\Omega)}\right)+\left\|\mathbf{y}-s_{X, \mathbf{y}}^{(\nu)} \mid X\right\|_{\ell_{\infty}(X)}\right) \\
& \leq \tilde{C}\left(2 h^{\tau-(d / 2-d / q)+}\|f\|_{W_{2}^{\tau}(\Omega)}+\frac{N}{2 C}\|f\|_{W_{2}^{\tau}(\Omega)}^{2}+(1-N \nu) \epsilon^{*}\right) .
\end{aligned}
$$

At first glance the term containing $\epsilon^{*}$ seems to be odd because it could be uncontrollable. But according to [12] we can at least assume $\epsilon^{*}$ to be bounded by

$$
\epsilon^{*} \leq \frac{1}{2}\left(\max _{1 \leq i \leq N} y_{i}-\min _{1 \leq i \leq N} y_{i}\right) .
$$

If this inequality is not satisfied, the problem (5.2.4) possesses only the trivial solution $s \equiv$ 0 , which is not interesting. Furthermore, we see the $\epsilon^{*}$-term occurs with a factor $(1-N \nu)$, which can be used to control this term. If we choose $\nu \geq \frac{1}{N}$, the term $(1-N \nu) \epsilon^{*}$ vanishes or is even negative. The parameter $\nu$ is a lower bound on the fraction of support vectors [51], and hence $\nu=1 / N$ means to get at least one support vector, that is a non-trivial solution. Since we are not interested in the case of trivial solutions, the condition $\nu \geq 1 / N$ is a reasonable assumption. We can use the results from Lemma 5.4.1 to derive a more explicit upper bound on $\epsilon^{*}=\epsilon^{*}(C, \nu, f)$ by

$$
0 \leq\left\|s_{X, \mathbf{y}}^{(\nu)} \mid X-\mathbf{y}\right\|_{\ell_{\infty}(X)} \leq \frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}+\epsilon^{*}(1-N \nu)
$$

If we assume $\nu>1 / N$, this leads to

$$
\epsilon^{*}=\epsilon^{*}(C, \nu, f) \leq \frac{N}{2 C(N \nu-1)}\|f\|_{\mathcal{N}_{K}}^{2} .
$$

Note that these bounds cannot be used for a better parameter choice, since we would need to rearrange this inequality and solve for $C$ or $\nu$. This would be possible only if there were lower bounds on $\epsilon^{*}$ as well. Moreover, the parameter $C$ appears in our error bound as a factor $\frac{N}{2 C}$, which implies that we expect convergence only in the case $C \rightarrow \infty$. In this case $\epsilon^{*}$ will be small, as can be deduced from problem (5.2.4).
We shall make these bounds more explicit for the case of quasi-uniformly distributed points. In this case the number of points $N$ and the fill distance $h$ are related to each other by

$$
\begin{equation*}
c_{1} N^{-1 / d} \leq h \leq c_{2} N^{-1 / d} \tag{5.4.3}
\end{equation*}
$$

where $c_{1}$ and $c_{2}$ denote positive constants [65, Proposition 14.1].

Corollary 5.4.3 In case of quasi-uniform exact data we may choose the problem parameters as

$$
C=\frac{N\|f\|_{W_{2}^{\tau}(\Omega)}}{2 h^{\tau}} \approx h^{-(\tau+d)}\|f\|_{W_{2}^{\tau}(\Omega)}, \text { and } \nu \geq \frac{1}{N}
$$

to get

$$
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{2}(\Omega)} \leq \tilde{C} h^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)} \leq \tilde{C} N^{-\frac{\tau}{d}}\|f\|_{W_{2}^{\tau}(\Omega)}
$$

or as

$$
C=\frac{N\|f\|_{W_{2}^{\tau}(\Omega)}}{2 h^{\tau-d / 2}} \approx h^{-(\tau+d / 2)}\|f\|_{W_{2}^{\tau}(\Omega)}, \text { and } \nu \geq \frac{1}{N}
$$

to get

$$
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{\infty}(\Omega)} \leq \tilde{C} h^{\tau-d / 2}\|f\|_{W_{2}^{\tau}(\Omega)} \leq \tilde{C} N^{-\frac{\tau}{d}-\frac{1}{2}}\|f\|_{W_{2}^{\tau}(\Omega)}
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$, with a generic positive constant $\tilde{C}$, which depends on $\tau, d, \Omega$, but not on $f$ or $X$.

Note that these bounds yield arbitrarily high sampling orders, provided that the functions are smooth enough, i.e., $\tau$ is large enough. Therefore they are in this setting better than the usual minimax rate $N^{-\frac{2 \tau}{2 \tau+d}}$ [40]. In the following we shall give our error estimates only in terms of the fill distance $h$ rather than in terms of the number of points $N$. This is due to the fact that the approximation rate in $h$ is independent of the space dimension $d$. However, it should be clear how the approximation rates translate into error estimates in terms of $N$ in the case of quasi-uniform data due to the inequality (5.4.3).
Corollary 5.4 .3 shows, that the solution of the $\nu$-SVR leads to optimal sampling orders in the Sobolev space [48] with respect to the fill distance $h$. These optimal rates are also attained by classical interpolation in the native Hilbert space [65]. The $\nu$-SVR, however, allows for much more flexibility and less complicated solutions. Our numerical results will confirm these convergence rates.

## $5.5 \quad \nu$-SVR with Inexact Data

In this section we denote again by $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ the solution to the problem (5.2.1) for a set of points $X:=\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right\} \subset \Omega$ and $\mathbf{y} \in \mathbb{R}^{N}$, but we allow the given data to be corrupted by some additive error $\mathbf{r}=\left(r_{1}, \ldots, r_{N}\right)$, that means

$$
\begin{equation*}
f\left(\mathbf{x}^{(j)}\right)=y_{j}+r_{j} \quad \text { for } j=1, \ldots, N \tag{5.5.1}
\end{equation*}
$$

where is $f \in W_{2}^{\tau}(\Omega) \cong \mathcal{N}_{K}$. Note that there are no assumptions concerning the error distribution. As in the previous section we have to show a stability estimate of the following form.

Lemma 5.5.1 Under the assumption (5.5.1) on the data $\mathbf{y}$, we find that for every $X$ a
solution $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ to the optimization problem (5.2.1) satisfies for all $\epsilon \geq 0$

$$
\begin{aligned}
\left\|s_{X, \mathbf{y}}^{(\nu)}\right\|_{\mathcal{N}_{K}(\Omega)} & \leq \sqrt{\frac{2 C}{N} \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+2 C \nu \epsilon+\|f\|_{\mathcal{N}_{K}}^{2}} \text {, and } \\
\left\|s_{X, \mathbf{y}}^{(\nu)}-\mathbf{y}\right\|_{\ell_{\infty}(X)} & \leq \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon+(1-N \nu) \epsilon^{*}+\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2} .
\end{aligned}
$$

Proof: Again, we denote the interpolant to $f$ with respect to $X$ and $K$ by $I_{f}$, and use $H_{C, \nu}^{\mathrm{y}}$ as defined in equation (5.4.2). Then we have

$$
\frac{1}{2 C}\left\|s_{X, \mathbf{y}}^{(\nu)}\right\|_{\mathcal{N}_{K}}^{2} \leq H_{C, \nu}^{\mathbf{y}}\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right) \leq H_{C, \nu}^{\mathrm{y}}\left(I_{f}, \epsilon\right) \leq \frac{1}{N} \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu \epsilon+\frac{1}{2 C}\|f\|_{\mathcal{N}_{K}}^{2},
$$

which implies

$$
\left\|s_{X, \mathbf{y}}^{(\nu)}\right\|_{\mathcal{N}_{K}} \leq \sqrt{\frac{2 C}{N} \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+2 C \nu \epsilon+\|f\|_{\mathcal{N}_{K}}^{2}} .
$$

Moreover we have for all $i=1, \ldots, N$

$$
\begin{aligned}
\left|s_{X, \mathbf{y}}^{(\nu)}\left(\mathbf{x}^{(i)}\right)-y_{i}\right| & \leq \sum_{j=1}^{N}\left|s_{X, \mathbf{y}}^{(\nu)}\left(\mathbf{x}^{(j)}\right)-y_{j}\right|_{\epsilon^{*}}+\epsilon^{*} \\
& \leq N H_{C, \nu}^{\mathbf{y}}\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)+(1-N \nu) \epsilon^{*} \\
& \leq \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon+(1-N \nu) \epsilon^{*}+\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2} .
\end{aligned}
$$

Again we can use the results from Lemma 5.5 .1 to derive a more explicit upper bound on $\epsilon^{*}=\epsilon^{*}(C, \nu, f, \epsilon)$. Note that $\epsilon^{*}$ depends now also on the free parameter $\epsilon$.

$$
0 \leq\left\|\left.s_{X, \mathbf{y}}^{(\nu)}\right|_{X}-\mathbf{y}\right\|_{\ell_{\infty}(X)} \leq \frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}+\epsilon^{*}(1-N \nu)+\sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon .
$$

If we assume $\nu>1 / N$, this leads to

$$
\epsilon^{*}(C, \nu, f, \epsilon) \leq \frac{1}{N \nu-1}\left(\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}+\sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon\right) .
$$

Using the sampling inequality as in the case of exact data leads to the following result on $L_{q}$-norms.

Theorem 5.5.2 We suppose $f \in W_{2}^{\tau}(\Omega)$ with $f\left(\mathbf{x}^{(i)}\right)=y_{i}+r_{i}$. Let $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ be a solution of the $\nu$-SVR, i.e., the optimization problem (5.2.1). Then there is a constant $\tilde{C}>0$, which depends on $\tau, d$ and $\Omega$, but not on $f$ or $X$, such that for all $\epsilon>0$ the approximation error can be bounded by

$$
\begin{aligned}
\left\|f-s_{X, y}^{(\nu)}\right\|_{L_{q}(\Omega)} & \leq \tilde{C}\left(h^{\tau-(d / 2-d / q)+}\left(\|f\|_{W_{2}^{\tau}(\Omega)}+\sqrt{\frac{2 C}{N} \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+2 C \nu \epsilon+\|f\|_{W_{2}^{\tau}(\Omega)}^{2}}\right)\right. \\
& \left.+\sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon+\epsilon^{*}(1-N \nu)+\frac{N}{2 C}\|f\|_{W_{2}^{\tau}(\Omega)}^{2}\right)
\end{aligned}
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$.
We now want to assume that the data errors do not exceed the data itself. For this we suppose

$$
\begin{equation*}
\|\mathbf{r}\|_{\ell_{\infty}(X)} \leq \delta \leq\|f\|_{W_{2}^{\tau}(\Omega)} \tag{5.5.2}
\end{equation*}
$$

for a parameter $\delta>0$.
Corollary 5.5.3 If we choose

$$
\begin{aligned}
C & =\frac{N\|f\|_{W_{2}^{\tau}(\Omega)}^{2}}{2 \delta} \approx \frac{h^{-d}}{\delta}\|f\|_{W_{2}^{\tau}(\Omega)}^{2} \\
\epsilon & =\delta, \quad, \quad \text { and } \quad \nu=\frac{1}{N}
\end{aligned}
$$

we get

$$
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{2}(\Omega)} \leq \tilde{C}\left(h^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)}+\delta\right),
$$

and

$$
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{\infty}(\Omega)} \leq \tilde{C}\left(h^{\tau-d / 2}\|f\|_{W_{2}^{\tau}(\Omega)}+\delta\right) .
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$, with a generic positive constant $\tilde{C}$, which depends on $\tau, d$ and $\Omega$, but not on $f$ or $X$.

## $5.6 \quad \epsilon$-SVR with Exact Data

Since our arguments for the $\nu$-SVR apply similarly to the $\epsilon$-SVR, we skip over details and just state the results. Note that in this case, the non-negative parameter $\epsilon$ is fixed in contrast to the free variable in the $\nu$-SVR. Analogously to the notation introduced in the previous sections, we denote by $s_{X, \mathrm{y}}^{(\epsilon)}$ the solution to the problem (5.2.6) for $X:=$ $\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right\} \subset \Omega$ and $\mathbf{y} \in \mathbb{R}^{N}$. The stability estimate takes the following form.
Lemma 5.6.1 Under the assumption (5.4.1) concerning the data, we find that for every $X$ and every fixed $\epsilon \in \mathbb{R}^{+}$a solution $s_{X, \mathbf{y}}^{(\epsilon)}$ to problem (5.2.6) satisfies

$$
\begin{aligned}
\left\|s_{X, \mathbf{y}}^{(\epsilon)}\right\|_{\mathcal{N}_{K}(\Omega)} & \leq\|f\|_{\mathcal{N}_{K}}, \text { and } \\
\left\|\left.s_{X, \mathbf{y}}^{(\epsilon)}\right|_{X}-\mathbf{y}\right\|_{\ell_{\infty}(X)} & \leq \frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}+\epsilon
\end{aligned}
$$

Again this leads to the following result on continuous $L_{q}$-norms.
Theorem 5.6.2 We suppose $f \in W_{2}^{\tau}(\Omega)$ with $f\left(\mathbf{x}^{(i)}\right)=y_{i}$. Let $s_{X, \mathbf{y}}^{(\epsilon)}$ be a solution of the $\epsilon$-SVR, i.e., the optimization problem (5.2.6). Then there is a constant $\tilde{C}>0$, which depends on $\tau, d$ and $\Omega$, but not on $\epsilon$, $f$ or $X$, such that the approximation error can be bounded by

$$
\begin{equation*}
\left\|f-s_{X, y}^{(\epsilon)}\right\|_{L_{q}(\Omega)} \leq \tilde{C}\left(2 h^{\tau-(1 / 2-1 / q)+}+\|f\|_{W_{2}^{\tau}(\Omega)}+\frac{N}{2 C}\|f\|_{W_{2}^{\tau}(\Omega)}^{2}+\epsilon\right) \tag{5.6.1}
\end{equation*}
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$.
Applying the same arguments as in the $\nu$-SVR case we obtain the following corollary.
Corollary 5.6.3 If we choose

$$
C=\frac{N\|f\|_{W_{2}^{\tau}(\Omega)}}{2 h^{\tau}}, \text { respectively } C=\frac{N\|f\|_{W_{2}^{\tau}(\Omega)}}{2 h^{\tau-d / 2}},
$$

the inequality (5.6.1) turns into

$$
\left\|f-s_{X, \mathbf{y}}^{(\epsilon)}\right\|_{L_{2}(\Omega)} \leq \tilde{C}\left(3 h^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)}+\epsilon\right),
$$

respectively

$$
\left\|f-s_{X, \mathbf{y}}^{(\epsilon)}\right\|_{L_{\infty}(\Omega)} \leq \tilde{C}\left(3 h^{\tau-d / 2}\|f\|_{W_{2}^{\tau}(\Omega)}+\epsilon\right)
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$, with a generic positive constant $\tilde{C}$, which depends on $\tau, d, \Omega$ and $\epsilon$, but not on $f$ or $X$.

The rôle of the parameter $C$ is similar to the one in case of the $\nu$-SVR. Unlike in the case of the $\nu$-SVR we are free to choose the parameter $\epsilon$. We see that exact data implies that we should choose $\epsilon \approx 0$. The case $C \rightarrow \infty$ and $\epsilon \rightarrow 0$ leads to exact interpolation, and the well known error bounds [65] are attained.

## $5.7 \quad \epsilon$-SVR with Inexact Data

In this section we denote again by $s_{X, \mathrm{y}}^{(\epsilon)}$ the solution to the problem (5.2.6) for a set of points $X:=\left\{x^{(1)}, \ldots, x^{(N)}\right\} \subset \Omega$ and $\mathbf{y} \in \mathbb{R}^{N}$, but we allow the given data to be corrupted by some additive error according to assumption (5.5.1).

Lemma 5.7.1 Under the assumption (5.5.1) concerning the data, for every $X$ and every fixed $\epsilon \in \mathbb{R}^{+}$a solution $s_{X, \mathbf{y}}^{(\epsilon)}$ to problem (5.2.6) satisfies

$$
\begin{aligned}
\left\|s_{X, \mathbf{y}}^{(\epsilon)}\right\|_{\mathcal{N}_{K}(\Omega)} & \leq \sqrt{\|f\|_{\mathcal{N}_{K}}^{2}+\frac{2 C}{N} \sum_{i=1}^{N}\left|r_{i}\right|_{\epsilon}} \text {, and } \\
\left\|\left.s_{X, \mathbf{y}}^{(\epsilon)}\right|_{X}-\mathbf{y}\right\|_{\ell_{\infty}(X)} & \leq \frac{N}{2 C}\|f\|_{\mathcal{N}_{K}}^{2}+\sum_{i=1}^{N}\left|r_{i}\right|_{\epsilon}+\epsilon
\end{aligned}
$$

These bounds shall now be plugged into the sampling inequality.
Theorem 5.7.2 We suppose $f \in W_{2}^{\tau}(\Omega)$ with $f\left(\mathbf{x}^{(i)}\right)=y_{i}$. Let $s_{X, \mathbf{y}}^{(\epsilon)}$ be a solution of the $\epsilon$-SVR, i.e., the optimization problem (5.2.6). Then there is a constant $\tilde{C}>0$, which depends on $\tau, d$ and $\Omega$, but not on $\epsilon, f$ or $X$, such that the approximation error can be bounded by

$$
\begin{aligned}
\left\|f-s_{X, y}^{(\epsilon)}\right\|_{L_{q}(\Omega)} \leq & \tilde{C}\left(2 h^{\tau-d(1 / 2-1 / q)+}\left(\|f\|_{W_{2}^{\tau}(\Omega)}+\sqrt{\|f\|_{W_{2}^{\tau}(\Omega)}^{2}+\frac{2 C}{N} \sum_{i=1}^{N}\left|r_{i}\right|_{\epsilon}}\right)\right. \\
& \left.+\frac{N}{2 C}\|f\|_{W_{2}^{\tau}(\Omega)}^{2}+\sum_{i=1}^{N}\left|r_{i}\right|_{\epsilon}+\epsilon\right)
\end{aligned}
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$.
If we again assume that the error level $\delta$ does not overrule the native space norm of the generating function, i.e.,

$$
\|\mathbf{r}\|_{\ell_{\infty}(X)} \leq \delta \leq\|f\|_{W_{2}^{\tau}(\Omega)},
$$

we get these convergence orders, for our specific choices of the parameters.
Corollary 5.7.3 Again we assume that the error satisfies (5.5.1). If we choose $\epsilon=\delta$ and $C=h^{-\tau-d} / 2$ then we find for quasi-uniform data

$$
\begin{align*}
\left\|f-s_{X, \mathbf{y}}^{(\epsilon)}\right\|_{L_{2}(\Omega)} & \leq \tilde{C}\left(h^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)}+\delta\right), \text { and }  \tag{5.7.1}\\
\left\|f-s_{X, \mathbf{y}}^{(\epsilon)}\right\|_{L_{\infty}(\Omega)} & \leq \tilde{C}\left(h^{\tau-d / 2}\|f\|_{W_{2}^{\tau}(\Omega)}+\delta\right) \tag{5.7.2}
\end{align*}
$$

for all discrete sets $X \subset \Omega$ with fill distance $h:=h_{X, \Omega} \leq C_{\Omega}\lfloor\tau-1\rfloor^{-2}$, with a generic positive constant $\tilde{C}$, which depends on $\tau, d, \Omega$ and $\epsilon$, but not on $f$ or $X$.

### 5.8 Spectral Convergence Orders for Support Vector Machines

So far, we have considered SV machines with kernels of finite smoothness, which have many practical advantages (see Section 5.9). However, there are also some infinitely smooth kernels that are popular in specific applications. The most important examples beside the Gaussian kernels are infinite polynomial kernels [29], or more general, power series kernels [70]. If we employ the sampling inequalities derived in Chapter 4, we similarly find an error analysis for SV machines with smooth kernels. We restrict ourselves to the case of Gaussians since all results can be carried over to other kernels easily. We write again abbreviately $\mathcal{N}_{K}:=\mathcal{N}_{G}$. To derive error bounds for the $\nu$-SVR optimization problem (5.2.1) we shall apply Theorem 4.2.2 to the residual $f-s_{X, \mathbf{y}}^{(\nu)}$, where $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ denotes the solution to the problem (5.2.1) for a $X:=\left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right\} \subset \Omega$ and $\mathbf{y} \in \mathbb{R}^{N}$. First, we consider exact data satisfying assumptions (5.4.1).

Theorem 5.8.1 Let $\Omega$ be a cube, and let $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ be a solution of the optimization problem (5.2.1). Then the generalization error can be bounded by

$$
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{q}(\Omega)} \leq \tilde{C}\left(e^{c \log (\tilde{c} h) / h}\|f\|_{\mathcal{N}_{K}(\Omega)}+\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}+(1-N \nu)_{+} \epsilon^{*}\right) .
$$

Corollary 5.8.2 In case of quasi-uniform exact data we may choose the problem parameters as

$$
C=\frac{N\|f\|_{\mathcal{N}_{K}(\Omega)}}{e^{c \log (\tilde{c} h) / h}} \quad, \text { and } \quad \nu=\frac{1}{N}
$$

to get

$$
\begin{equation*}
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{q}(\Omega)} \leq C e^{c \log (\tilde{c} h) / h}\|f\|_{\mathcal{N}_{K}(\Omega)} \tag{5.8.1}
\end{equation*}
$$

Now we shall consider general domains satisfying an interior cone condition. Similarly to the case above we get with Theorem 4.1.7

Theorem 5.8.3 Let $\left(s_{X, \mathbf{y}}^{(\nu)}, \epsilon^{*}\right)$ be a solution of (5.2.1). Then the approximation error can be bounded by

$$
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{q}(\Omega)} \leq E\left(A e^{B \log (D h) / \sqrt{h}}\|f\|_{\mathcal{N}_{K}(\Omega)}+\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}-(N \nu-1) \epsilon^{*}\right)
$$

Here the constants $A, B, D, E$ depend on $\theta, \alpha, d$ and $q$. In case of quasi-uniform data we may choose the parameters as

$$
C=\frac{N\|f\|_{\mathcal{N}_{K}(\Omega)}}{A e^{B \log (C h) / \sqrt{h}}} \quad, \text { and } \quad \nu \geq \frac{1}{N}
$$

to get the estimate

$$
\begin{equation*}
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{q}(\Omega)} \leq C A e^{B \log (C h) / \sqrt{h}}\|f\|_{\mathcal{N}_{K}(\Omega)} \tag{5.8.2}
\end{equation*}
$$

We now shall consider the case of inexact given data as described in assumption (5.5.1). Again we have to bound both the native space norm and the discrete norm. Following Lemma 5.5.1 we have

$$
\begin{aligned}
\left\|s_{X, \mathbf{y}}^{(\nu)}\right\|_{\mathcal{N}_{K}(\Omega)} & \leq \sqrt{\frac{2 C}{N} \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+2 C \nu \epsilon+\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}} \\
\left\|s_{X, \mathbf{y}}^{(\nu)}-\mathbf{y}\right\|_{\ell_{\infty}(X)} & \leq \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon-\epsilon^{*} \cdot(N \nu-1)+\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}
\end{aligned}
$$

Theorem 5.8.4 Under the assumptions (5.5.1) and $\Omega \subset \mathbb{R}^{d}$ being a compact cube we have for all $\epsilon>0$

$$
\begin{aligned}
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{q}(\Omega)} & \leq \tilde{C}\left(e^{c \log (\tilde{c} h) / h}\left(\|f\|_{\mathcal{N}_{K}(\Omega)}+\sqrt{\frac{2 C}{N} \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+2 C \nu \epsilon+\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}}\right)\right. \\
& \left.+\sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon-\epsilon^{*}(N \nu-1)+\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}\right) .
\end{aligned}
$$

Corollary 5.8.5 Suppose assumption (5.5.2) holds. If we choose

$$
C=\frac{N\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}}{2 \delta e^{c \log (\tilde{c} h) / h}}, \quad \epsilon=\delta \quad \text { and } \quad \nu=\frac{1}{N}
$$

we get in case of quasi-uniform data for any non-trivial solution

$$
\begin{equation*}
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{2}(\Omega)} \leq C\left(e^{c \log (\tilde{c} h) / h}\|f\|_{\mathcal{N}_{K}(\Omega)}+\delta\right) . \tag{5.8.3}
\end{equation*}
$$

Note that bounds like (5.8.3) allow excellent bounds on the number of training samples required in the worst possible case to get required prediction quality.
Now we shall also here consider a general Lipschitz domain $\Omega$. Analogously to the case above we get
Theorem 5.8.6 Under the assumption (5.5.1) we have for all $\epsilon>0$

$$
\begin{aligned}
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{q}(\Omega)} & \leq \tilde{C}\left(A e^{B \log (C h) / \sqrt{\hbar}}\left(\|f\|_{\mathcal{N}_{K}(\Omega)}+\sqrt{\frac{2 C}{N} \sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+2 C \nu \epsilon+\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}}\right)\right. \\
& \left.+\sum_{j=1}^{N}\left|r_{j}\right|_{\epsilon}+\nu N \epsilon-\epsilon^{*}(N \nu-1)+\frac{N}{2 C}\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}\right) .
\end{aligned}
$$

Here the constants $A, B, C$ depend on $\theta, \alpha, d, q$. Again we get for the choice

$$
C=\frac{N\|f\|_{\mathcal{N}_{K}(\Omega)}^{2}}{2 \delta A e^{B \log (C h) / \sqrt{h}}}, \quad \epsilon=\delta, \quad \nu=\frac{1}{N}
$$

in case of quasi-uniform data for any non-trivial solution $\left(f^{*}, \epsilon^{*}\right)$

$$
\begin{equation*}
\left\|f-s_{X, \mathbf{y}}^{(\nu)}\right\|_{L_{2}(\Omega)} \leq C\left(A e^{B \log (C h) / \sqrt{h}}\|f\|_{\mathcal{N}_{K}(\Omega)}+\delta\right) . \tag{5.8.4}
\end{equation*}
$$

We shall now confirm our analytical results by numerical experiments. For that we restrict ourselves to the case of finite smoothness, i.e., algebraic sampling orders.

### 5.9 Numerical Results

In this section we present some numerical examples to support our analytical results, in particular the rates of convergence in case of exact training data, and the detection of the error levels in case of noisy data.

### 5.9.1 Exact Training Data

Figure 5.1 illustrates the approximation orders in case of exact given data as considered in Sections 5.4 and 5.6. For that, we used regular datasets generated by the respective functions to be reconstructed and employed the $\epsilon$ - and the $\nu$-SVR with the parameter choices provided in Corollaries 5.6 .3 and 5.4.3. We implemented the finite dimensional formulations of the associated optimization problems as described in equations (5.2.4) and (2.4.3). As kernel functions we used Wendland's functions (see e.g. [59, 60, 61]) for two reasons:

On the one hand they yield sparse kernel matrices $\mathbf{K}$ due to their compact support, on the other hand they are easy to implement since they are piecewise polynomials. Furthermore Wendland's functions may be scaled to improve their numerical behavior. An unscaled function $K$ has support $\operatorname{supp}(K) \subset B(0,1) \subset \mathbb{R}^{d}$. The scaling is done in such a way, that the decay of the Fourier transform is preserved, i.e.,

$$
\begin{equation*}
K^{(c)}(\mathbf{x})=c^{-d} K\left(\frac{\mathbf{x}}{c}\right), \quad \mathbf{x} \in \mathbb{R}^{d} \tag{5.9.1}
\end{equation*}
$$

By construction we have $\operatorname{supp}\left(K^{(c)}\right) \subset B(0, c)$, such that small choices of the scaling parameter $c$ imply sparse kernel matrices

$$
\mathbf{K}^{(\mathbf{c})}=\left(K^{(c)}\left(\left\|\mathbf{x}^{(i)}-\mathbf{x}^{(j)}\right\|_{2}\right)\right)_{i, j=1 \ldots N}
$$

On the other hand-side it is known that the constant factors in our error estimates increase with decreasing $c$. This is a typical trade-off situation between good approximation properties and good condition numbers of the kernel matrices $\mathbf{K}^{(\mathbf{c})}$ [65].
The double logarithmic plots in subfigure 5.1(a) visualize the convergence orders in terms of the fill distance $h$. For that, the $L_{\infty}$-approximation error $\left\|f-s_{X, \mathbf{y}}\right\|_{L_{\infty}}$ is plotted versus the fill distance $h$. The convergence rates are given by the slopes of the lines.
In Subfigure 5.1(a) the data was generated by

$$
f(\mathbf{x})=(\mathbf{x}-0.5)_{+}^{2.5+e p s} \in W_{2}^{3}([0,1])
$$

where eps denotes the relative machine precision in the sense of MATLAB. We use the notation $(t)_{+}:=\max \{0, t\}$ for all $t \in \mathbb{R}$. This function $f$ is sampled on a regular grid in the unit interval $I:=[0,1]$ with 20 to 100 points. Note that in this case the fill distance is given by $h=1 / N$. We use two different kernel functions, namely

- $K_{1}(\mathbf{x})=(1-\|\mathbf{x}\|)_{+}^{3}(3\|\mathbf{x}\|+1)$ with native space $W_{2}^{2}([0,1])$, and
- $K_{2}(\mathbf{x})=(1-\|\mathbf{x}\|)_{+}^{5}\left(8\|\mathbf{x}\|^{2}+5\|\mathbf{x}\|+1\right)$ with native space $W_{2}^{3}([0,1])$.

The scaling parameter according to equation (5.9.1) was chosen as $c=0.3$. We employed the $\epsilon$ - and the $\nu$ - SVR with the parameter choices provided in Corollaries 5.6 .3 and 5.4.3. The respective corollaries predicted convergence rates of 1.5 for $K_{1}$, and 2.5 for $K_{2}$. In Subfigure 5.1(a) the plots for the $\epsilon$ - and $\nu$-SVR both show orders 1.9 for $K_{1}$ and 2.7 for $K_{2}$. Subfigure 5.1(b) shows a 2-dimensional example. The data was generated by the smooth function

$$
f(\mathbf{x})=\sin \left(x_{1}+x_{2}\right)
$$

This function $f$ is sampled on a regular grid in the unit interval $I:=[0,1]^{2}$ with 16 to 144 points. Note that in this case the fill distance is given by $h=\frac{1}{\sqrt{N}}$. We used three different kernel functions, namely

- $K_{3}(\mathbf{x})=(1-\|\mathbf{x}\|)_{+}^{4}(4\|\mathbf{x}\|+1)$ with native space $W_{2}^{2}\left([0,1]^{2}\right)$,
- $K_{4}(\mathbf{x})=(1-\|\mathbf{x}\|)_{+}^{6}\left(35\|\mathbf{x}\|^{2}+18\|\mathbf{x}\|+3\right)$ with native space $W_{2}^{3}\left([0,1]^{2}\right)$, and
- $K_{5}(\mathbf{x})=(1-\|\mathbf{x}\|)_{+}^{8}\left(32\|\mathbf{x}\|^{3}+25\|\mathbf{x}\|^{2}+8\|\mathbf{x}\|+1\right)$ with native space $W_{2}^{5}\left([0,1]^{2}\right)$.

The kernel functions were scaled by $c=1$ according to equation (5.9.1). For the sake of simplicity we employed only the $\nu$-SVR with the parameter choices provided in Corollary 5.4.3. The predicted convergence rates in the fill distance $h$ are 1.5 for $K_{3}, 2.5$ for $K_{4}$ and 3 for $K_{5}$. The numerical experiments show orders 1.8 for $K_{3}, 2.4$ for $K_{4}$ and 2.9 for $K_{5}$. Thus, the numerical examples verify our analytical results.

(a) Data generated by $f \in W_{2}^{3}(I)$ on regular grids in $I . \nu$ - and $\epsilon$-SVR yield orders 1.9 for $K_{1} c=0.3$, and 2.7 for $K_{2}$. Scaling parameter $c=0.3$.

(b) Data generated by smooth function on regular grids in $I^{2}$. $\nu$-SVR yields orders 1.8 for $K_{3}, 2.4$ for $K_{4}$, and 2.9 for $K_{5}$. Scaling parameter $c=1$

Figure 5.1: Double logarithmic plots of the $L_{\infty}$-approximation error versus the fill distance $h$ for exact training data.

### 5.9.2 Inexact Data

Figure 5.2 shows examples for the case of noisy data. The plots show the $L_{\infty}$-approximation error $\left\|f-s_{X, \mathbf{y}}\right\|_{L_{\infty}}$ versus the fill distance $h$. For simplicity we concentrated on the case of the $\nu$-SVR in the one dimensional setting. We used the noise model from Section 5.5.1, that is $y=f+r$. In Subfigure 5.2(a) the function

$$
f(\mathrm{x})=(\mathrm{x}-0.5)_{+}^{2.5+e p s}+\sin (5 x) \in W_{2}^{3}([0,1])
$$

is sampled on regular grids of 5 to 30 points in $[0,1]$. The data is disturbed by an error $r$ that is normally distributed with mean zero and variance 0.01 . As kernel function we use $K_{1}$, and the parameters of the $\nu$-SVR are chosen according to Corollary 5.5.3. As predicted in Corollary 5.5 .3 the plot shows that the error reaches the variance and keeps bounded for $h \rightarrow 0$.
In Subfigure 5.2(b) the function

$$
f(\mathrm{x})=(\mathrm{x}-0.5)_{+}^{2.5+e p s}+\sin (5 x) \in W_{2}^{3}([0,1])
$$

is sampled on regular grids of 5 to 80 points in the unit interval $I=[0,1]$. Here, the data is corrupted by an error of $\pm 0.01$, where the sign of the error is chosen randomly with equal
likelihood for plus and minus. As kernel function we use $K_{1}$, and the parameters of the $\nu$-SVR are chosen as in Corollary 5.5.3. The plot shows that the $L_{\infty}$-approximation error converges towards the error level 0.01 for $h \rightarrow 0$.

(a) Data disturbed by random error with mean zero and variance 0.01 . Approximation error reaches the variance and keeps bounded for $h \rightarrow 0$.

(b) Data disturbed by random sign deterministic error $\pm 0.01$. Approximation error converges towards the error level for $h \rightarrow 0$.

Figure 5.2: $L_{\infty}$-approximation error versus fill distance in case of inexact data.

## Chapter 6

## Bernstein Inequality

In this chapter we shall prove Bernstein inequalities for certain radial basis functions. This technical result will be applied in the next chapters in the context of weak recovery. To be precise, we consider radial basis functions $\Phi: \mathbb{R}^{d} \rightarrow \mathbb{R}$ that satisfy for constants $c_{1}(\Phi)$ and $c_{2}(\Phi)$

$$
c_{1}(\Phi)\left(1+\|\omega\|_{2}^{2}\right)^{-\tau} \leq \hat{\Phi}(\omega) \leq c_{2}(\Phi)\left(1+\|\omega\|_{2}^{2}\right)^{-\tau} \quad, \tau>d / 2
$$

for all $\omega \in \mathbb{R}^{d}$. We shall consider a bounded domain $\Omega \subset \mathbb{R}^{d}$ with Lipschitz boundary $\partial \Omega$. Furthermore we assume $\Omega$ to be a star-shaped domain, i.e., there is a point $x_{0} \in \Omega$ such that for all $x \in \Omega$ the line segment $x x_{0} \subset \Omega$ between $x_{0}$ and $x$ is completely contained in $\Omega$. Without loss of generality we may assume $x_{0}=0$. We shall use the following notation

$$
\begin{aligned}
\Omega^{\gamma} & :=(1+\gamma) \cdot \Omega:=\{(1+\gamma) x: x \in \Omega\}, \text { and } \\
\Omega^{-\gamma} & :=(1-\gamma) \cdot \Omega:=\{(1-\gamma) x: x \in \Omega\},
\end{aligned}
$$

for $0 \leq \gamma<1$. Note that by construction $\Omega^{0}=\bar{\Omega}$. We point out that this definition differs from the definition of the $\epsilon$-interior in Section 4.3. For "nice" domains, $\Omega^{\gamma}$ is just a "scaled" copy of $\Omega$. We shall make use of these "scaled" domains just for technical reasons, see Remark 6.5.4. As usual, we consider discrete sets $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ of pairwise distinct points. The distribution of the points of $X$ in $\Omega$ is mainly described by the fill distance or mesh norm $h_{X, \Omega}$ and the separation distance $q_{X}$, which are defined by

$$
\begin{aligned}
h_{X, \Omega} & :=\sup _{x \in \Omega} \min _{x_{j} \in X}\left\|x-x_{j}\right\|_{2} \quad, \text { and } \\
q_{X} & :=\frac{1}{2} \min _{\substack{x_{i}, x_{j} \in X \\
x_{i} \neq x_{j}}}\left\|x_{i}-x_{j}\right\|_{2} .
\end{aligned}
$$

So far, Bernstein inequalities have been known only on the whole $\mathbb{R}^{d}$ [39]. They take the following form.

Theorem 6.0.1 [39] Under the assumptions above, there is a constant $C>0$ such that for all $a_{1}, \ldots, a_{N} \in \mathbb{R}$,

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C q_{X}^{-\tau}\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} .
$$

This bounds a strong continuous norm on a finite dimensional trial space by a large factor depending on the separation distance and a weaker continuous norm. Such inequalities must hold true, since all norms on finite dimensional spaces are equivalent. We simply bound the equivalence constants in terms of the geometry of $X$. In this chapter, the main result is a Bernstein inequality for domains.

Theorem 6.0.2 If the discrete set $X$ lies in $\Omega^{-2 q_{X}}$ there is a constant $C>0$ such that the Bernstein inequality

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C q_{X}^{-\tau}\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}(\Omega)}
$$

holds for arbitrary $a_{1}, \ldots, a_{N} \in \mathbb{R}$.
These kinds of inequalities have various applications. We shall mainly use them to quantify the amount of oversampling in unsymmetric recovery processes (see Chapter 7). They are also a main step towards the application of meshless methods in the setting of generalized finite elements. In this context Bernstein inequalities are called inverse assumptions [8, 33].

### 6.1 Stationary Weak Data

To prove Bernstein inequalities, we shall apply a norming set argument to convolution-type stationary weak data, which we shall analyze in more detail in Chapter 7. We consider functionals $\lambda_{z} \in\left(L_{2}\right)^{*}$ of the form

$$
\lambda_{z}(f):=\int_{\mathbb{R}^{d}} \Psi(z-x) f(x) d x \quad, z \in \mathbb{R}^{d}, f \in L_{2},
$$

where $\Psi: \mathbb{R}^{d} \rightarrow \mathbb{R}$ is a radial kernel, whose Fourier-transform decays like

$$
c_{1}(\Psi)\left(1+\|\omega\|_{2}^{2}\right)^{-\rho} \leq \hat{\Psi}(\omega) \leq c_{2}(\Psi)\left(1+\|\omega\|_{2}^{2}\right)^{-\rho} .
$$

We will always assume $\rho>d / 4$ to ensure that $\Psi \in L_{2}\left(\mathbb{R}^{d}\right)$. Further, we restrict ourselves to radial kernels with compact support. As described in detail in Section 5.9, such kernels exist under weak conditions on $\rho$ and the space dimension $d$ [65, Theorem 10.35]. We explained that we can scale a kernel $\Psi$ with compact support in $B(0,1) \subset \mathbb{R}^{d}$ by a positive number $q$ via

$$
\Psi_{q}(\cdot):=q^{-d} \Psi\left(\frac{\dot{ }}{q}\right)
$$

to get a function supported in $B(0, q) \subset \mathbb{R}^{d}$, with Fourier transform

$$
\widehat{\Psi_{q}}(\omega)=(2 \pi)^{d / 2} \hat{\Psi}(q \omega) .
$$

We shall scale the kernel function $\Psi$ with the separation distance. Therefore, we speak of stationary data. From now on we always consider the associated scaled functionals, i.e.,

$$
\begin{equation*}
\lambda_{z}^{q}(f):=\int_{\mathbb{R}^{d}} \Psi_{q}(z-x) f(x) d x \quad, z \in \mathbb{R}^{d} \tag{6.1.1}
\end{equation*}
$$

which yields the Fourier-transform

$$
\widehat{\lambda_{z}^{q}(f)}(\omega)=(2 \pi)^{d / 2} \hat{\Psi}(q \omega) \hat{f}(\omega) .
$$

### 6.2 Norming Set Argument

In the following we need some technical tools, which we collect at the beginning. To start with, we need an estimate of the following form.

Lemma 6.2.1 For all $x \geq B$, all $\beta>0$ and all $c \in \mathbb{R}$ we have

$$
\frac{1}{\left(1+c^{2} x^{2}\right)^{\beta}} \leq\left(\frac{\frac{1}{B^{2}}+1}{c^{2}}\right)^{\beta} \frac{1}{\left(1+x^{2}\right)^{\beta}} .
$$

Proof: We have the following chain of equivalences:

$$
\begin{aligned}
\frac{1}{\left(1+c^{2} x^{2}\right)^{\beta}} & \leq\left(\frac{\frac{1}{B^{2}}+1}{c^{2}}\right)^{\beta} \frac{1}{\left(1+x^{2}\right)^{\beta}} \\
\Leftrightarrow\left(1+c^{2} x^{2}\right) & \geq\left(\frac{c^{2}}{\frac{1}{B^{2}}+1}\right)\left(1+x^{2}\right) \\
\Leftrightarrow \frac{\left(1+c^{2} x^{2}\right)}{c^{2}\left(1+x^{2}\right)} & \geq \frac{1}{\frac{1}{B^{2}}+1} \\
\Leftrightarrow \frac{\frac{1}{x^{2}}+1}{\frac{1}{c^{2} x^{2}}+1} & \leq \frac{1}{B^{2}}+1,
\end{aligned}
$$

and the last inequality is true for $x \geq B$.
We shall also need the notion of a bandlimited function. We say that a function is bandlimited with bandwidth $\sigma$ if $\hat{f}$ has compact support in $B(0, \sigma)$. The space of all bandlimited functions

$$
B_{\sigma}:=\left\{f \in L_{2}\left(\mathbb{R}^{d}\right): \overline{\operatorname{supp}(\hat{f})} \subset B(0, \sigma)\right\}
$$

is often called Paley-Wiener space.
The scaled functionals from (6.1.1) allow us to define a Sampling Operator by

$$
\begin{aligned}
\Lambda_{X}^{q} & : L_{2}\left(\mathbb{R}^{d}\right) \rightarrow \mathbb{R}^{N}, \\
\Lambda_{X}^{q}(f) & :=\left(\lambda_{x_{1}}^{q}(f), \ldots, \lambda_{x_{N}}^{q}(f)\right)^{T} .
\end{aligned}
$$

The main result of this section states that for every $f \in L_{2}\left(\mathbb{R}^{d}\right)$ for any $\sigma$ large enough, there is a function $f_{\sigma} \in B_{\sigma}$ that interpolates in the sense of generalized interpolation with respect to the functionals $\lambda_{x_{j}}^{q}$, and that is nearly the best-approximation with respect to the $L_{2}\left(\mathbb{R}^{d}\right)$-norm. For that we follow the paths of [39, Chapter 3] and use the following theorem from [36].

Theorem 6.2.2 Let $Y$ be a (possibly complex) Banach space, $V$ a subspace of $Y$, and $Z^{*}$ a finite dimensional subspace of $Y^{*}$, the dual of $Y$. If for every $z^{*} \in Z^{*}$ and some $\beta>1$, $\beta$ independent of $z^{*}$,

$$
\left\|z^{*}\right\|_{Y^{*}} \leq \beta\left\|\left.z^{*}\right|_{V}\right\|_{V^{*}}
$$

then for any $y \in Y$ there exists $v \in V$ such that $v$ interpolates $y$ on $Z^{*}$; that is, $z^{*}(y)=$ $z^{*}(v)$ for all $z^{*} \in Z^{*}$. In addition, $v$ approximates $y$ in the sense that

$$
\|y-v\|_{Y} \leq(1+2 \beta) \operatorname{dist}_{Y}(y, V)
$$

To apply Theorem 6.2.2 to our setting we take $Y=L_{2}\left(\mathbb{R}^{d}\right), V=B_{\sigma}$ and $Z^{*}=$ $\operatorname{span}\left\{\lambda_{x_{j}}^{q}: 1 \leq j \leq N\right\}$.

Theorem 6.2.3 For $f \in L_{2}\left(\mathbb{R}^{d}\right)$ there exists a bandlimited function $f_{\sigma} \in B_{\sigma}$ with

$$
\Lambda_{X}(f)=\Lambda_{X}\left(f_{\sigma}\right), \quad \text { and } \quad\left\|f-f_{\sigma}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} \leq 5 \cdot \operatorname{dist}_{L_{2}}\left(f, B_{\sigma}\right)
$$

for $\sigma=\frac{\kappa}{q_{X}}$, where the constant $\kappa \geq 1$ depends only on $\Psi, \rho$ and $d$, but not on $X$.
Proof: We adapt ideas from [39]. From the definition of the functionals $\lambda_{x_{j}}^{q}$,

$$
\lambda_{x_{j}}^{q}(f)=\int_{\mathbb{R}^{d}} \Psi_{q}\left(x_{j}-x\right) f(x) d x=\left\langle\Psi_{q}\left(x_{j}-\cdot\right), f\right\rangle_{L_{2}\left(\mathbb{R}^{d}\right)},
$$

we see that the Riesz-representer of $\lambda_{x_{j}}^{q}$ is given by $\Psi_{q}\left(x_{j}-\cdot\right)$. We shall use the abbreviatory notations $z^{*}=\sum_{j=1}^{N} a_{j} \lambda_{x_{j}}^{q}$ and $z=\sum_{j=1}^{N} a_{j} \Psi_{q}\left(x_{j}-\cdot\right)$. Then Riesz' theorem states that

$$
\left\|\sum_{j=1}^{N} a_{j} \lambda_{x_{j}}^{q}\right\|_{Y^{*}}=\left\|z^{*}\right\|_{Y^{*}}=\|z\|_{Y}=\left\|\sum_{j=1}^{N} a_{j} \Psi_{q}\left(x_{j}-x\right)\right\|_{Y}
$$

By the definition of the separation distance $q:=q_{X}$ we see that

$$
\begin{aligned}
\|z\|_{L_{2}\left(\mathbb{R}^{d}\right)}^{2} & =\sum_{j, k=1}^{N} a_{j} a_{k} \int_{\mathbb{R}^{d}} \Psi_{q}\left(x-x_{j}\right) \Psi_{q}\left(x-x_{k}\right) d x \\
& =\sum_{j=1}^{N} a_{j}^{2} \int_{\mathbb{R}^{d}}\left(\Psi_{q}\left(x-x_{j}\right)\right)^{2} d x=\|a\|_{2}^{2} \int_{\mathbb{R}^{d}}\left(\Psi_{q}(x)\right)^{2} d x \\
& =\|a\|_{2}^{2} \int_{\mathbb{R}^{d}} q^{-2 d}\left(\Psi\left(\frac{x}{q}\right)\right)^{2} d x=\|a\|_{2}^{2} q^{-d} \int_{\mathbb{R}^{d}}(\Psi(x))^{2} d x .
\end{aligned}
$$

We denote by $z_{\sigma}$ the function given by $\widehat{z_{\sigma}}=\hat{z} \chi_{B(0, \sigma)}$, where $\chi_{B(0, \sigma)}$ is the characteristic function of the ball $B(0, \sigma)$. Then it is easy to see that we have

$$
\left\|\left.z^{*}\right|_{V}\right\|_{Y^{*}}=\left\|z_{\sigma}\right\|_{Y} .
$$

Plancherel's theorem allows us to compute with $\kappa:=q \sigma$, where we use Lemma6.2.1 in the last line

$$
\begin{aligned}
& \left\|z-z_{\sigma}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)}^{2}=C \int_{\|\omega\|_{2} \geq \sigma}\left|\sum_{j=1}^{N} a_{j} e^{-i x_{j}^{T} \omega}\right|^{2}\left(\widehat{\Psi_{q}}(\omega)\right)^{2} d \omega \\
\leq & C \int_{\|\omega\|_{2} \geq \sigma}\left|\sum_{j=1}^{N} a_{j} e^{-i x_{j}^{T} \omega}\right|^{2}\left(1+\|q \omega\|_{2}^{2}\right)^{-2 \rho} d \omega \\
= & C \sigma^{d} \int_{\|\xi\|_{2} \geq 1}\left|\sum_{j=1}^{N} a_{j} e^{-i\left(\sigma x_{j}\right)^{T} \xi}\right|^{2}\left(1+\|q \sigma \xi\|_{2}^{2}\right)^{-2 \rho} d \xi \\
= & C \sigma^{d} \int_{\|\xi\|_{2} \geq 1}\left|\sum_{j=1}^{N} a_{j} e^{-i\left(\sigma x_{j}\right)^{T} \xi}\right|^{2}\left(1+\|\kappa \xi\|_{2}^{2}\right)^{-2 \rho} d \xi \\
\leq & C \sigma^{d}\left(\frac{2}{\kappa^{2}}\right)^{2 \rho} \int_{\|\xi\|_{2} \geq 1}\left|\sum_{j=1}^{N} a_{j} e^{-i\left(\sigma x_{j}\right)^{T} \xi}\right|^{2}\left(1+\|\xi\|_{2}^{2}\right)^{-2 \rho} d \xi,
\end{aligned}
$$

where $C>0$ always denotes a generic constant, which may depend on $\Psi$ and $d$, but not on $q, \sigma$ or $\kappa$. Now we have to bound the integral. We note that

$$
\begin{aligned}
\int_{\mathbb{R}^{d}}\left|\sum_{j=1}^{N} a_{j} e^{-i\left(\sigma x_{j}^{T}\right) \xi}\right|^{2}\left(1+\|\xi\|_{2}^{2}\right)^{-2 \rho} d \xi & =C \sum_{j, k=1}^{N} a_{j} a_{k} \mathfrak{K}_{2 \rho}\left(\sigma\left(x_{j}-x_{k}\right)\right) \\
& \leq C \lambda_{\max }\left(A_{\sigma X, \mathfrak{K}_{2 \rho}}\right)\|a\|_{2}^{2}
\end{aligned}
$$

Here we denote by $\lambda_{\max }(A)$ the largest eigenvalue of a matrix $A$, and use the notations

$$
\mathfrak{K}_{\beta}:=C_{\beta}\|x\|_{2}^{-\beta+\frac{d}{2}} \mathcal{K}_{\frac{d}{2}-\beta}(x) \quad, \text { and } \quad A_{\sigma X, \mathfrak{K}_{2 \rho} \rho}:=\left(\mathfrak{K}_{2 \rho}\left(\sigma\left(x_{j}-x_{k}\right)\right)\right)_{i, j=1, \ldots, N},
$$

where $\mathcal{K}_{\beta}$ denotes a modified Bessel function of the third kind. A result from [39] yields

$$
\begin{aligned}
\lambda_{\max }\left(A_{\sigma X, \mathfrak{K}_{2 \rho}}\right) & \leq \mathfrak{K}_{2 \rho}(0)+\sum_{k=1}^{\infty} 3 d(k+2)^{d-1} \mathfrak{K}_{2 \rho}(\kappa k) \\
& \leq \mathfrak{K}_{2 \rho}(0)+\sum_{k=1}^{\infty} 3 d(k+2)^{d-1} \mathfrak{K}_{2 \rho}(k)=: C(\rho, d)
\end{aligned}
$$

since $\kappa \geq 1$ and $\mathfrak{K}_{2 \rho}(\cdot)$ is exponentially decreasing towards infinity. By the assumption $\rho>d / 4$ we can further estimate

$$
\begin{aligned}
\left\|z-z_{\sigma}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)}^{2} & \leq C\left(\frac{2}{\kappa^{2}}\right)^{2 \rho}\left(\frac{\kappa}{q}\right)^{d}\|a\|_{2}^{2} \\
& \leq C \kappa^{d-4 \rho}\|z\|_{L_{2}\left(\mathbb{R}^{d}\right)}^{2} \leq \frac{1}{4}\|z\|_{L_{2}\left(\mathbb{R}^{d}\right)}^{2}
\end{aligned}
$$

with a generic constant $C>0$, which may depend on $\Psi$ and $d$, but not on $q, \sigma$ or $\kappa$. The last inequality holds for $\kappa$ large enough, i.e., $\kappa>(4 C)^{\frac{1}{4 \rho-d}}$. Finally we can apply Theorem 6.2 .2 with $Y=L_{2}\left(\mathbb{R}^{d}\right), V=B_{\sigma}$ and $Z^{*}=\operatorname{span}\left\{\lambda_{x_{j}}^{q}: 1 \leq j \leq N\right\}$.

### 6.3 Extension Operator and Native Space

We shall need a result that relates the native space norms on different domains. Recall that without loss of generality we assume $0 \in \Omega$. First, we consider a special extension map for scaled domains $c \Omega:=\{c x: x \in \Omega\}$ with a scaling parameter $c>0$. We shall prove this in a slightly more general context. For that, we proceed along the lines of [3, Section 5], but we mainly consider the special case of full norms. We stick to the notation of [3], and repeat their results for the simpler case ignoring $P_{k}$.
Definition 6.3.1 We call two subsets $\Omega$ and $\hat{\Omega}$ affine-equivalent, if there is an affine map $F: \mathbb{R}^{d} \rightarrow \mathbb{R}^{d}$ such that $\Omega=F(\hat{\Omega})$. F has a representation as $F: x \mapsto L(x)+\mathfrak{c}$, where $L \in \mathcal{L}\left(\mathbb{R}^{d}, \mathbb{R}^{d}\right)$ is an invertible bounded linear operator and $\mathfrak{c} \in \mathbb{R}^{d}$ is a constant.

Remark 6.3.2 There is a bijection between the functions defined on $\Omega$ and $\hat{\Omega}$. We denote corresponding functions by $v: \Omega \rightarrow \mathbb{R}$ and $\hat{v}: \hat{\Omega} \rightarrow \mathbb{R}$, that is, $v=\hat{v} \circ F^{-1}$ and $\hat{v}=v \circ F$.

We shall pay special attention to the relation between the (semi-)norms of corresponding functions on $\Omega$ and $\hat{\Omega}$.

Theorem 6.3.3 Let $\Omega$ and $\hat{\Omega}$ be two affine-equivalent open subsets of $\mathbb{R}^{d}$ with Lipschitz boundaries. Let $p \in[1, \infty]$ and $\tau \in \mathbb{N}_{0}$. For any function $v$ belonging to $W_{p}^{\tau}(\Omega)$, the function $\hat{v}=v \circ F$ belongs to $W_{p}^{\tau}(\hat{\Omega})$ and vice versa. Moreover, there is a constant $C=C(d, \tau, p)>0$ such that

- $|\hat{v}|_{W_{p}^{\tau}(\hat{\Omega})}=|v \circ F|_{W_{p}^{\tau}(\hat{\Omega})} \leq C\|L\|^{\tau}|\operatorname{det} L|^{-\frac{1}{p}}|v|_{W_{p}^{\tau}(\Omega)}$,
$\bullet|v|_{W_{p}^{\tau}(\Omega)}=\left|\hat{v} \circ F^{-1}\right|_{W_{p}^{\tau}(\Omega)} \leq C\left\|L^{-1}\right\|^{\tau}|\operatorname{det} L|^{\frac{1}{p}}|\hat{v}|_{W_{p}^{\tau}(\hat{\Omega})}$.
Proof: See [14], or [30, page 101] for the special case $p=2$.
Proceeding as in [3], we can also extend the results for semi-norms to full Sobolev norms.
Corollary 6.3.4 Under the assumptions from Theorem 6.3.3, we get

$$
\begin{aligned}
\|v \circ F\|_{W_{p}^{\tau}(\hat{\Omega})} & \leq \tilde{C}(d, \tau, p)|\operatorname{det} L|^{-\frac{1}{p}} \max _{0 \leq k \leq \tau}\|L\|^{k}\|v\|_{W_{p}^{\tau}(\Omega)}, \text { and } \\
\left\|\hat{v} \circ F^{-1}\right\|_{W_{p}^{\tau}(\Omega)} & \leq \tilde{C}(d, \tau, p)|\operatorname{det} L|^{\frac{1}{p}} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\|\hat{v}\|_{W_{p}^{\tau}(\hat{\Omega})}
\end{aligned}
$$

Proof: We use Theorem 6.3.3 to estimate

$$
\begin{aligned}
\|\hat{v}\|_{W_{p}^{\tau}(\hat{\Omega})}^{p} & =\sum_{|\alpha| \leq \tau} \int_{\hat{\Omega}}\left(D^{\alpha} \hat{v}(x)\right)^{p} d x=\sum_{k \leq \tau}|\hat{v}|_{W_{p}^{k}(\hat{\Omega})}^{p} \\
& \leq \sum_{k \leq \tau}\left(C(d, k, p)|\operatorname{det} L|^{-\frac{1}{p}}\|L\|^{k}|v|_{W_{p}^{k}(\Omega)}\right)^{p} \\
& \leq|\operatorname{det} L|^{-1} \sum_{k \leq \tau} C(d, k, p)^{p}\|L\|^{p k}|v|_{W_{p}^{k}(\Omega)}^{p} \\
& \leq|\operatorname{det} L|^{-1} \max _{0 \leq k \leq \tau} C(d, k, p)^{p} \max _{0 \leq k \leq \tau}\|L\|^{p k}\|v\|_{W_{p}^{\tau}(\Omega)}^{p} .
\end{aligned}
$$

Taking the $p$-th root yields

$$
\|v\|_{W_{p}^{\tau}(\hat{\Omega})} \leq|\operatorname{det} L|^{-\frac{1}{p}} \max _{0 \leq k \leq \tau} C(d, k, p) \max _{0 \leq k \leq \tau}\|L\|^{k}\|v\|_{W_{p}^{\tau}(\Omega)}
$$

The second inequality follows analogously.
Following [3], we can extend these results to the whole Euclidean space.
Corollary 6.3.5 Under the assumptions from Theorem 6.3.3, we have

$$
\begin{aligned}
\left|\hat{v} \circ F^{-1}\right|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} & \leq C\left\|L^{-1}\right\|^{\tau}|\operatorname{det} L|^{\frac{1}{p}}|\hat{v}|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \\
\left\|\hat{v} \circ F^{-1}\right\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} & \leq \tilde{C}(d, \tau)|\operatorname{det} L|^{\frac{1}{p}} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\|\hat{v}\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)}
\end{aligned}
$$

where we can also take $F$ instead of $F^{-1}$ if we replace $\hat{v}$ by $v$.
Proof: The proof is analogous to [3, Proof of Proposition 5.2]. For $\nu \in \mathbb{N}$ we apply Theorem 6.3 .3 to the pair of affine-equivalent sets $B(0, \nu)$ and $F^{-1}(B(0, \nu))$. This yields

$$
\left|\hat{v} \circ F^{-1}\right|_{W_{p}^{\tau}(B(0, \nu))} \leq C(d, \tau, p)\left\|L^{-1}\right\|^{\tau}|\operatorname{det} L|^{\frac{1}{p}}|v|_{W_{p}^{\tau}\left(F^{-1}(B(0, \nu))\right.}
$$

where $C(d, \tau, p)$ does not depend on $\nu \in \mathbb{N}$. Now we use the fact that both,

$$
\bigcup_{\nu \in \mathbb{N}} B(0, \nu)=\mathbb{R}^{d} \quad \text { and } \quad \bigcup_{\nu \in \mathbb{N}} F^{-1}(B(0, \nu))=\mathbb{R}^{d}
$$

Hence we can pass to the limit $\nu \rightarrow \infty$ and get with the Theorem of Monotone Convergence in the case $p<\infty$ (one could also use direct calculations),

$$
\left|\hat{v} \circ F^{-1}\right|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C(d, \tau, p)\left\|L^{-1}\right\|^{\tau}|\operatorname{det} L|^{\frac{1}{p}}|\hat{v}|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} .
$$

The second claim follows by a summation argument as in the proof of Corollary 6.3.4.
Now we can state the main extension result, which is essentially a reformulation of [3, Proposition 5.2] for our specific situation.

Theorem 6.3.6 Let $\hat{\Omega} \subset \mathbb{R}^{d}$ be a bounded domain with Lipschitz boundary. Let $p \in[1, \infty]$, let $\tau \in \mathbb{N}$ be such that $\tau>\left(\frac{d}{p}-\frac{d}{2}\right)_{+}$. Then for any bounded domain $\Omega \subset \mathbb{R}^{d}$ with Lipschitz boundary and affine equivalent to $\hat{\Omega}$, there exists a linear extension operator

$$
\mathfrak{E}_{\Omega}: W_{p}^{\tau}(\Omega) \rightarrow W_{p}^{\tau}\left(\mathbb{R}^{d}\right)
$$

such that

- $\left.\mathfrak{E}_{\Omega} v\right|_{\Omega}=v \quad$ for all $v \in W_{p}^{\tau}(\Omega)$
- $\left\|\mathfrak{E}_{\Omega} v\right\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C(\hat{\Omega}, d, \tau, p) \max _{0 \leq k \leq \tau}\|L\|^{k} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\|v\|_{W_{p}^{\tau}(\Omega)}$.

Proof: The proof is analogous to the second half of [3, Proof of Proposition 5.2]. We use the notation for $F$ and $L$ from above, and assume $\Omega=F(\hat{\Omega})$.
We define two operators

$$
\begin{aligned}
& P_{1}: W_{p}^{\tau}(\Omega) \rightarrow W_{p}^{\tau}(\hat{\Omega}), \quad v \mapsto v \circ F, \\
& P_{2}: W_{p}^{\tau}\left(\mathbb{R}^{d}\right) \rightarrow W_{p}^{\tau}\left(\mathbb{R}^{d}\right), \quad w \mapsto w \circ F^{-1} .
\end{aligned}
$$

Furthermore, we use the extension operator due to Stein [56] on the reference domain $\hat{\Omega}$,

$$
\mathcal{E}_{\hat{\Omega}}: W_{p}^{\tau}(\hat{\Omega}) \rightarrow W_{p}^{\tau}\left(\mathbb{R}^{d}\right)
$$

that satisfies

- $\left.\mathcal{E}_{\hat{\Omega}} \hat{v}\right|_{\hat{\Omega}}=\hat{v} \quad$ for all $\hat{v} \in W_{p}^{\tau}(\hat{\Omega})$, and
- $\left\|\mathcal{E}_{\hat{\Omega}} \hat{v}\right\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C(\hat{\Omega}, d, \tau, p)\|\hat{v}\|_{W_{p}^{\tau}(\hat{\Omega})}$ for all $\hat{v} \in W_{p}^{\tau}(\hat{\Omega})$.

Now we define the operator

$$
P:=P_{2} \circ \mathcal{E}_{\hat{\Omega}} \circ P_{1}: W_{p}^{\tau}(\Omega) \rightarrow W_{p}^{\tau}\left(\mathbb{R}^{d}\right) .
$$

The operator $P$ fulfills an extension property since

$$
\begin{aligned}
P v(x) & =\left(P_{2} \circ \mathcal{E}_{\hat{\Omega}} \circ P_{1}\right) v(x)=\left(\mathcal{E}_{\hat{\Omega}} P_{1} v\right)\left(F^{-1}(x)\right) \\
& =\left(P_{1} v\right)\left(F^{-1}(x)\right)=(v \circ F)\left(F^{-1}(x)\right)=v(x)
\end{aligned}
$$

for almost all $x \in \Omega$ and all $v \in W_{p}^{\tau}(\Omega)$. The last line follows directly from the facts that $F^{-1}(x) \in \hat{\Omega}$ and that $\mathcal{E}_{\hat{\Omega}} P_{1} v=P_{1} v$ almost everywhere in $\hat{\Omega}$.
Now we can calculate with the shortened notation $u=\mathcal{E}_{\hat{\Omega}} P_{1} v$, applying Corollary 6.3.4 and Theorem 6.3.3,

$$
\begin{aligned}
\|P v\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} & =\left\|u \circ F^{-1}\right\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \\
& \leq \tilde{C}(d, \tau, p)|\operatorname{det} L|^{\frac{1}{p}} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\|u\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \\
& =\tilde{C}(d, \tau, p)|\operatorname{det} L|^{\frac{1}{p}} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\left\|\left(\mathcal{E}_{\hat{\Omega}} \circ P_{1}\right) v\right\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \\
\leq & \tilde{C}(d, \tau, p)\left\|\mathcal{E}_{\hat{\Omega}}\right\||\operatorname{det} L|^{\frac{1}{p}} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\left\|P_{1} v\right\|_{W_{p}^{\tau}(\hat{\Omega})} \\
& =\tilde{C}(d, \tau, p)\left\|\mathcal{E}_{\hat{\Omega}}\right\||\operatorname{det} L|^{\frac{1}{p}} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\|v \circ F\|_{W_{p}^{\tau}(\hat{\Omega})} \\
\leq & C(d, \tau, p)\left\|\mathcal{E}_{\hat{\Omega}}\right\||\operatorname{det} L|^{\frac{1}{p}} . \\
& \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}|\operatorname{det} L|^{-\frac{1}{p}} \max _{0 \leq k \leq \tau}\|L\|^{k}\|v\|_{W_{p}^{\tau}(\Omega)} \\
= & C\left\|\mathcal{E}_{\hat{\Omega}}\right\| \max _{0 \leq k \leq \tau}\|L\|^{k} \max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k}\|v\|_{W_{p}^{\tau}(\Omega)} .
\end{aligned}
$$

We can extend this result to the case of fractional order spaces.
Corollary 6.3.7 Theorem 6.3 .6 remains true for real $\tau>0$.
Proof: This follows directly by an operator interpolation argument.
We consider a reference domain $\Omega$ in $\mathbb{R}^{d}$ that is star-shaped with respect to $0 \in \Omega$ and that has a Lipschitz boundary, and as an affine map we choose $F: \mathbb{R}^{d} \rightarrow \mathbb{R}^{d}, x \mapsto(1-q) x$ for a $0<q<1 / 2$. This yields a simple scaling of the domain $\Omega$ to the domain $F(\Omega)=$ $(1-q) \Omega$. The linear map is of the form $F=L=(1-q)$ Id, and since $q<1 / 2$ it is also invertible. We can easily compute

$$
\begin{aligned}
\max _{0 \leq k \leq \tau}\|L\|^{k} & =\max _{0 \leq k \leq \tau}(1-q)^{k}=1 \\
\max _{0 \leq k \leq \tau}\left\|L^{-1}\right\|^{k} & =\max _{0 \leq k \leq \tau}(1-q)^{-k}=\frac{1}{(1-q)^{\tau}} \leq 2^{\tau} .
\end{aligned}
$$

If we now apply Theorem 6.3 .6 to the sets $\Omega$ and $(1-q) \Omega$ we get the following result.
Corollary 6.3.8 Let $\Omega$ be a star-shaped domain in $\mathbb{R}^{d}$ with Lipschitz boundary, furthermore $\tau>0$ and $p \in[1, \infty]$. Then, for $0<q<1 / 2$ there is an extension operator

$$
\mathfrak{E}_{(1-q) \Omega}: W_{p}^{\tau}((1-q) \Omega) \rightarrow W_{p}^{\tau}\left(\mathbb{R}^{d}\right)
$$

which satisfies with a constant $C(\Omega, d, \tau, p)$ independent of $q$

- $\left.\mathfrak{E}_{(1-q) \Omega} v\right|_{(1-q) \Omega}=v \quad$ for all $v \in W_{p}^{\tau}((1-q) \Omega)$, and
- $\left\|\mathfrak{E}_{(1-q) \Omega} v\right\|_{W_{p}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C(\Omega, d, \tau, p) 2^{\tau}\|v\|_{W_{p}^{\tau}((1-q) \Omega)}$.

For technical reasons, we shall impose another assumption on the domain $\Omega$. This is done only to simplify notation, see Remark 6.5.4. From now on, we restrict ourselves to domains $\Omega$ that are star-shaped with respect to $0 \in \Omega$ and that satisfy the following condition. For all $0 \leq q \leq q_{0}$ we have the inclusion

$$
\begin{equation*}
(1-2 q) \Omega=\Omega^{-2 q} \subset\{x \in \Omega: \overline{B(x, 2 q)} \subset \Omega\} \tag{6.3.1}
\end{equation*}
$$

Remark 6.3.9 The assumption (6.3.1) is satisfied for most of the standard domains, such as balls or parallelepipeds.

Like in [65, Corollary 10.48] we can derive the following Corollary.
Corollary 6.3.10 Suppose that $\Phi \in L_{1}\left(\mathbb{R}^{d}\right)$ has a Fourier transform that decays algebraically, i.e., as $\left(1+\|\omega\|_{2}^{2}\right)^{-\tau}$ with $\tau>d / 2$. Then $\mathcal{N}_{\Phi}\left(\Omega^{-2 q_{X}}\right) \cong W_{2}^{\tau}\left(\Omega^{-2 q_{X}}\right)$ with equivalent norms, where the equivalence constants do not depend on $0<q_{X}<1 / 4$.

Proof: The proof is analogue to [65, Corollary 10.48]. We apply the extension operator from Corollary 6.3 .8 to the domain $\Omega^{-2 q_{X}}$ to deduce that the extension constants are independent of $0<q_{X}<1 / 4$.

We can use this Corollary 6.3 .10 to prove the following Lemma. To simplify the notation, we write abbreviately $\lambda_{j}$ for $\lambda_{j}^{q}$ and the like.

Lemma 6.3.11 If $X \subset \Omega^{-2 q} \subset \Omega$, we get

$$
\left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C\left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{W_{2}^{\tau}\left(\Omega^{-q}\right)}
$$

where $C$ may depend on $d, \Omega, \tau$, but not on $q$.
Proof: We see

$$
\begin{aligned}
\lambda_{j}^{x} \lambda_{k}^{y} \Phi(x-y) & =\int_{\mathbb{R}^{d}} \Psi_{q}\left(x-x_{j}\right) \int_{\mathbb{R}^{d}} \Psi_{q}\left(y-x_{k}\right) \Phi(x-y) d y d x \\
& =\int_{\mathbb{R}^{d}} \int_{\mathbb{R}^{d}} \Psi_{q}\left(x-x_{j}\right) \Psi_{q}\left(y-x_{k}\right) \Phi(x-y) d y d x \\
& =\int_{\Omega^{-q}} \int_{\Omega^{-q}} \Psi_{q}\left(x-x_{j}\right) \Psi_{q}\left(y-x_{k}\right) \Phi(x-y) d y d x
\end{aligned}
$$

We can now calculate

$$
\left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{\mathcal{N}_{\Phi}\left(\mathbb{R}^{d}\right)}^{2}=c^{T} A c
$$

and

$$
\left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{\mathcal{N}_{\Phi}\left(\Omega^{-q}\right)}^{2}=c^{T} \tilde{A} c .
$$

Since the entries of

$$
A_{i, j}=\int_{\mathbb{R}^{d}} \Psi_{q}\left(x-x_{j}\right) \int_{\mathbb{R}^{d}} \Psi_{q}\left(y-x_{k}\right) \Phi(x-y) d y d x
$$

and

$$
\left(A_{\Lambda}\right)_{i, j}=\int_{\Omega^{-q}} \Psi_{q}\left(x-x_{j}\right) \int_{\Omega^{-q}} \Psi_{q}\left(y-x_{k}\right) \Phi(x-y) d y d x
$$

coincide, we get with Corollary 6.3.10, which yields that $\mathcal{N}_{\Phi}\left(\Omega^{-q}\right) \cong W_{2}^{\tau}\left(\Omega^{-q}\right)$,

$$
\begin{aligned}
& \left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} \sim\left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{\mathcal{N}_{\Phi}\left(\mathbb{R}^{d}\right)}^{2} \\
& =\left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{\mathcal{N}_{\Phi}\left(\Omega^{-q}\right)}^{2} \sim\left\|\sum_{j=1}^{N} c_{j} \lambda_{j}^{x} \Phi(\cdot-x)\right\|_{W_{2}^{\tau}\left(\Omega^{-q}\right)}^{2}
\end{aligned}
$$

Hence we get the claim where $C$ is given by the extension operator norms.
In this proof we explicitly used the fact that the points are separated from the boundary.

### 6.4 Bernstein Inequality

Now we are ready to state the final result of this section, namely a Bernstein inequality. We shall always denote by $q:=q_{X}$ the separation distance of $X$.

Theorem 6.4.1 Suppose $\Omega \subset \mathbb{R}^{d}$ is star-shaped and bounded with Lipschitz boundary, and $X \subset \Omega^{-2 q} \subset \Omega$. Then there is a constant $C>0$ such that the Bernstein inequality

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C q^{-\tau}\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}(\Omega)}
$$

holds for all $a_{1}, \ldots, a_{N} \in \mathbb{R}$.
Proof: We define the abbreviations

$$
\begin{aligned}
s(\cdot) & :=\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right), \\
\hat{s_{\sigma}} & =\left(\hat{s} \chi_{B(0, \sigma)}\right) \quad, \text { and } \\
s_{\Lambda}(x) & :=\sum_{j=1}^{N} a_{j} \lambda_{x_{j}}^{q} \Phi(\cdot-x) .
\end{aligned}
$$

As a first step we use the result from [39, Lemma 3.3], i.e.,

$$
\|s\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} \leq C\left\|s_{\sigma}\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2},
$$

for $\sigma=\frac{\kappa}{q}$ and a constant $C$ independent of $q$. Using this, we can estimate

$$
\begin{aligned}
& \|s\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} \leq C\left\|s_{\sigma}\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} \\
\leq & C \int_{\|\omega\|_{2} \leq \sigma}\left|\sum_{j=1}^{N} a_{j} e^{-i x_{j}^{T} \omega}\right|^{2} \hat{\Phi}^{2}(\omega)\left(1+\|\omega\|_{2}^{2}\right)^{\tau} d \omega \\
= & C \int_{\|\omega\|_{2} \leq \sigma}\left|\sum_{j=1}^{N} a_{j} e^{-i x_{j}^{T} \omega}\right|^{2} \hat{\Phi}^{2}(\omega) \frac{\left(1+\|q \omega\|_{2}^{2}\right)^{2 \rho}}{\left(1+\|q \omega\|_{2}^{2}\right)^{2 \rho}}\left(1+\|\omega\|_{2}^{2}\right)^{\tau} d \omega \\
\leq & C \int_{\|\omega\|_{2} \leq \sigma}\left|\sum_{j=1}^{N} a_{j} e^{-i x_{j}^{T} \omega}\right|^{2} \hat{\Phi}^{2}(\omega)\left(1+\|q \omega\|_{2}^{2}\right)^{-2 \rho}\left(1+\|\omega\|_{2}^{2}\right)^{\tau} d \omega \\
\leq & C\left\|\left(s_{\Lambda}\right)_{\sigma}\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} \leq C\left\|s_{\Lambda}\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} \\
\leq & C\left\|s_{\Lambda}\right\|_{W_{2}^{\tau}\left(\Omega^{-q}\right)}^{2},
\end{aligned}
$$

where we used $\|\omega\|_{2} \leq \sigma=\frac{\kappa}{q_{X}}$ and Lemma 6.3.11 in the last step. Now we choose an $f_{\sigma} \in B_{\sigma}$ according to Theorem 6.2.3 such that

$$
\begin{equation*}
\lambda_{j}\left(f_{\sigma}\right)=\lambda_{j}\left(\chi_{\Omega^{-q}} s_{\Lambda}\right)=\lambda_{j}\left(s_{\Lambda}\right) . \tag{6.4.1}
\end{equation*}
$$

This function $f_{\sigma}$ has another important property.

$$
\begin{align*}
\left\|f_{\sigma}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} & \leq\left\|f_{\sigma}-\chi_{\Omega^{-q}} s_{\Lambda}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)}+\left\|\chi_{\Omega^{-q}} s_{\Lambda}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} \\
& \leq 5 \operatorname{dist}_{L_{2}}\left(\chi_{\Omega^{-q}} s_{\Lambda}, B_{\sigma}\right)+\left\|\chi_{\Omega^{-q}} s_{\Lambda}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} \\
& \leq 6\left\|\chi_{\Omega^{-q}} s_{\Lambda}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} \leq 6\left\|s_{\Lambda}\right\|_{L_{2}\left(\Omega^{-q}\right)} . \tag{6.4.2}
\end{align*}
$$

The stability property in the native space gives the estimate

$$
\left\|s_{\Lambda}\right\|_{W_{2}^{\tau}\left(\Omega^{-q}\right)}^{2} \leq\|f\|_{W_{2}^{\tau}\left(\Omega^{-q}\right)}^{2}
$$

for any $f \in W_{2}^{\tau}\left(\Omega^{-q}\right)$ satisfying the generalized interpolation condition

$$
\lambda_{j}(f)=\lambda_{j}\left(s_{\Lambda}\right), \quad 1 \leq j \leq N
$$

As a final ingredient we use a Bernstein inequality for bandlimited functions on the whole $\mathbb{R}^{d}$ [37],

$$
\left\|f_{\sigma}\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq \sigma^{\tau}\left\|f_{\sigma}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)} \quad \text { for all } f_{\sigma} \in \mathcal{B}_{\sigma} .
$$

Putting these things together yields with equation (6.4.2) the estimate

$$
\begin{aligned}
\|s\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} & \leq C\left\|s_{\Lambda}\right\|_{W_{2}^{\tau}\left(\Omega^{-q}\right)}^{2} \leq C\left\|f_{\sigma}\right\|_{W_{2}^{\tau}\left(\Omega^{-q}\right)}^{2} \\
& \leq C\left\|f_{\sigma}\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d} d\right.}^{2} \leq C \sigma^{2 \tau}\left\|f_{\sigma}\right\|_{L_{2}\left(\mathbb{R}^{d}\right)}^{2} \\
& \leq C \sigma^{2 \tau}\left\|s_{\Lambda}\right\|_{L_{2}\left(\Omega^{-q}\right)}^{2} .
\end{aligned}
$$

Direct evaluation of the $\|\cdot\|_{L_{2}}$-norm yields

$$
\begin{aligned}
& \left\|s_{\Lambda}\right\|_{L_{2}\left(\Omega^{-q}\right)}^{2}=\int_{\Omega^{-q}}\left(\sum_{j=1}^{N} a_{j} \lambda_{j}^{x} \Phi(y-x)\right)^{2} d y \\
= & \int_{\Omega^{-q}}\left(\sum_{j=1}^{N} a_{j} \int_{\mathbb{R}^{d}} \Psi_{q}\left(x-x_{j}\right) \Phi(y-x) d x\right)^{2} d y \\
= & \int_{\Omega^{-q}}\left(\sum_{j=1}^{N} a_{j} \int_{\mathbb{R}^{d}} \Psi_{q}(x-y) \Phi\left(x-x_{j}\right) d x\right)^{2} d y, x \mapsto y-x+x_{j} \\
= & \int_{\Omega^{-q}}\left(\int_{\mathbb{R}^{d}} \Psi_{q}(x-y) \sum_{j=1}^{N} a_{j} \Phi\left(x-x_{j}\right) d x\right)^{2} d y .
\end{aligned}
$$

Now we can use Cauchy-Schwarz' inequality to manipulate the last expression, which gives

$$
\begin{aligned}
& \int_{\Omega^{-q}}\left(\int_{\mathbb{R}^{d}} \Psi_{q}(x-y) s(x) d x\right)^{2} d y=\int_{\Omega^{-q}}\left(\int_{B(y, q)} \Psi_{q}(x-y) s(x) d x\right)^{2} d y \\
\leq & \left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} \int_{\Omega^{-q}}\left(\int_{B(y, q)} s(x) d x\right)^{2} d y \\
\leq & \left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} \int_{\Omega^{-q}}\left(\int_{B(y, q)}|s(x)| \cdot 1 d x\right)^{2} d y \\
\leq & \left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} \int_{\Omega^{-q}}\left(\sqrt{\int_{B(y, q)}|s(x)|^{2} d x} \sqrt{\int_{B(y, q)} 1 d \tilde{x}}\right)^{2} d y \\
= & \left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} \int_{\Omega^{-q}}\left(\int_{B(y, q)}|s(x)|^{2} d x \int_{B(y, q)} 1 d \tilde{x}\right) d y \\
\leq & C\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} q^{d} \int_{\Omega^{-q}} \int_{B(y, q)}|s(x)|^{2} d x d y \\
\leq & C\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)} q^{d} \int_{\Omega^{-q}} \int_{\Omega} \chi_{B(0, q)}(x-y)|s(x)|^{2} d x d y \\
\leq & C\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} q^{d} \int_{\Omega} \int_{\Omega^{-q}} \chi_{B(0, q)}(x-y)|s(x)|^{2} d y d x .
\end{aligned}
$$

Putting all the things together we get with Fubini's Theorem and the substitution $x-y \mapsto z$ in the third line:

$$
\begin{aligned}
\|s\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} & =C \sigma^{2 \tau}\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} q^{d} \int_{\Omega}|s(x)|^{2}\left(\int_{\Omega^{-q}} \chi_{B(0, q)}(x-y) d y\right) d x \\
& \leq C \sigma^{2 \tau}\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} q^{d} \int_{\Omega}|s(x)|^{2}\left(\int_{\mathbb{R}^{d}} \chi_{B(0, q)}(x-y) d y\right) d x \\
& =C \sigma^{2 \tau}\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} q^{d} \int_{\Omega}|s(x)|^{2}\left(\int_{\mathbb{R}^{d}} \chi_{B(0, q)}(z) d z\right) d x \\
& =C \sigma^{2 \tau}\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} q^{d} \int_{\Omega}|s(x)|^{2} \underbrace{\left(\int_{\mathbb{R}^{d}} \chi_{B(0, q)}(z) d z\right)}_{q^{d}} d x \\
& \leq C \sigma^{2 \tau}\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)}^{2} q^{2 d} \int_{\Omega}|s(x)|^{2} d x .
\end{aligned}
$$

Finally we can use the fact that $\left\|\Psi_{q}\right\|_{L_{\infty}\left(\mathbb{R}^{d}\right)} \approx q^{-d}$ to get the desired result

$$
\begin{aligned}
\|s\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)}^{2} & \leq C \sigma^{2 \tau}\|s\|_{L_{2}(\Omega)}^{2} \\
& \leq C q^{-2 \tau}\|s\|_{L_{2}(\Omega)}^{2} .
\end{aligned}
$$

We can easily derive the following corollary.
Corollary 6.4.2 Suppose $\Omega \subset \mathbb{R}^{d}$ is star-shaped and $X \subset \Omega$, then we have for all $a_{1}, \ldots, a_{N} \in \mathbb{R}$ the estimate

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C q^{-\tau}\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}\left(\Omega^{2 q}\right)} .
$$

### 6.5 Interpolation

If $s_{f, X}:=\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)$ is the interpolant with respect to the points from $X$ to a function $f \in W_{2}^{\tau}(\Omega)$, we can show that the norms $\left\|s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)}$ and $\left\|s_{f, X}\right\|_{L_{2}(\Omega)}$ are equivalent if the fill distance of $X$ in $\Omega$ is sufficiently small.

Theorem 6.5.1 Under the assumptions from above there is an $h_{0}(f)$ such that for all $s_{X, f}$ associated to sets $X$ with fill distance $h_{X, \Omega} \leq h_{0}(f)$ the inequalities

$$
\left\|s_{f, X}\right\|_{L_{2}(\Omega)} \leq\left\|s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)} \leq C\left\|s_{f, X}\right\|_{L_{2}(\Omega)}
$$

hold.
Remark 6.5.2 Please note that $h_{0}(f)$ depends on $f$ and is therefore not known a priori. However this estimate indicates that for all "sufficiently dense" sets $X$ the norms $\left\|s_{f, X}\right\|_{L_{2}(\Omega)}$ and $\left\|s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)}$ are comparable.

Proof: The first inequality is trivial. For the second one, we apply the normal sampling inequalities to derive

$$
\begin{aligned}
\left\|\mathcal{E}_{\Omega} s_{f, X}-\mathcal{E}_{\Omega} f\right\|_{L_{2}\left(\Omega^{2 q}\right)} & \leq c_{1} h_{X, \Omega^{2 q}}^{\tau}\left\|\mathcal{E}_{\Omega} s_{f, X}-\mathcal{E}_{\Omega} f\right\|_{W_{2}^{\tau}\left(\Omega^{2 q}\right)} \\
& \leq c_{1} h_{X, \Omega^{2 q}}^{\tau}\left\|\mathcal{E}_{\Omega}\left(s_{f, X}-f\right)\right\|_{W_{2}^{\tau}\left(\Omega^{2 q}\right)} \\
& \leq c_{1}\left\|\mathcal{E}_{\Omega}\right\| h_{X, \Omega^{2 q}}^{\tau}\left\|s_{f, X}-f\right\|_{W_{2}^{\tau}(\Omega)} \\
& \leq C_{1}(\Omega) h_{X, \Omega^{2 q}}^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)} .
\end{aligned}
$$

Analogously we have

$$
\left\|s_{f, X}-\mathcal{E}_{\Omega} f\right\|_{L_{2}\left(\Omega^{2 q}\right)} \leq C_{2}(\Omega) h_{X, \Omega^{2 q}}^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)}
$$

The triangle inequality now yields

$$
\begin{aligned}
\left\|s_{f, X}-\mathcal{E}_{\Omega} s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)} & \leq\left\|s_{f, X}-\mathcal{E}_{\Omega} f\right\|_{L_{2}\left(\Omega^{2 q}\right)}+\left\|\mathcal{E}_{\Omega} s_{f, X}-\mathcal{E}_{\Omega} f\right\|_{L_{2}\left(\Omega^{2 q}\right)} \\
& \leq\left(C_{1}(\Omega)+C_{2}(\Omega)\right) h_{X, \Omega^{2 q}}\|f\|_{W_{2}^{\tau}(\Omega)}
\end{aligned}
$$

First we point out that

$$
h_{X, \Omega^{2 q}} \leq 4 h_{X, \Omega} .
$$

Since $\|f-s\|_{L_{2}(\Omega)} \leq c h_{X, \Omega}^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)} \rightarrow 0$ for $h_{X, \Omega} \rightarrow 0$ we find a $h_{0}(f, \Omega)$ such that

$$
\left|\|f\|_{L_{2}(\Omega)}-\left\|s_{f, X}\right\|_{L_{2}(\Omega)}\right| \leq \frac{1}{2}\|f\|_{L_{2}(\Omega)} \text { for all } X \text { with } h_{X, \Omega} \leq h_{0}(f, \Omega) .
$$

This implies $\|f\|_{L_{2}(\Omega)} \leq 2\left\|s_{f, X}\right\|_{L_{2}(\Omega)}$. We therefore conclude that

$$
\left\|s_{f, X}\right\|_{L_{2}(\Omega)}>0 \quad \text { if } f \neq 0 .
$$

This allows us to find an $h_{1}(f, \Omega)$ such that

$$
\left(C_{1}(\Omega)+C_{2}(\Omega)\right) h_{X, \Omega^{2 q}}^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)} \leq \frac{1}{2}\left\|s_{f, X}\right\|_{L_{2}(\Omega)},
$$

for all sets X with $h_{X, \Omega} \leq h_{1}(f, \Omega)$. Combining these two inequalities yields

$$
\left\|s_{f, X}-\mathcal{E}_{\Omega} s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)} \leq\left\|s_{f, X}\right\|_{L_{2}(\Omega)}
$$

for all sets X with $h_{X, \Omega} \leq \max \left\{h_{0}(f, \Omega), h_{1}(f, \Omega)\right\}$. In the end we get

$$
\begin{aligned}
\left\|s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)} & \leq\left\|s_{f, X}-\mathcal{E}_{\Omega} s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)}+\left\|\mathcal{E}_{\Omega} s_{f, X}\right\|_{L_{2}\left(\Omega^{2 q}\right)} \\
& \leq\left(1+\left\|\mathcal{E}_{\Omega}\right\|\right)\left\|s_{f, X}\right\|_{L_{2}(\Omega)}
\end{aligned}
$$

for all point sets X with $h_{X, \Omega} \leq \max \left\{h_{0}(f, \Omega), h_{1}(f, \Omega)\right\}$.
Theorem6.5.1 indicates that the assumptions in Theorem6.4.1 are not sharp. If we combine Corollary 6.4.2 and Theorem 6.5.1, we immediately get

Corollary 6.5.3 Let the fill distance of $X \subset \Omega$ be sufficiently small and define $s_{f, X}(\cdot)=\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)$, where $A_{\Phi, X} a=f$ and $f \in W_{2}^{\tau}(\Omega)$. Then there is a constant $h_{0}(f)>0$ such that for $h_{X, \Omega} \leq h_{0}(f)$ we get

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C q^{-\tau}\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}(\Omega)} .
$$

Proof: Corollary 6.4 .2 yields

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{W_{2}^{\tau}\left(\mathbb{R}^{d}\right)} \leq C q^{-\tau}\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}\left(\Omega^{2 q}\right)} .
$$

Using Theorem 6.5.1 we get

$$
\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}\left(\Omega^{2 q}\right)} \leq C\left\|\sum_{j=1}^{N} a_{j} \Phi\left(\cdot-x_{j}\right)\right\|_{L_{2}(\Omega)}
$$

Combining these results yields the claim.
Unfortunately, we can not give a precise condition on the fill distance to ensure the last result.

Remark 6.5.4 Careful inspection of the proofs shows that the assumption (6.3.1) can be replaced by the less restrictive condition

$$
X \subset\{x \in \Omega: \overline{B(x, 2 q)} \subset \Omega\},
$$

which does not assume any scaling property of $\Omega$ (see Definition 4.3.1). This shows that we can allow more general domains $\Omega$.

## Chapter 7

## Stationary Weak Data

In the spirit of [47] this chapter deals with weak recovery of functions from weak function spaces. The main difficulty is that we cannot use point values as data for the reconstruction, since they might not be well defined in weak spaces. Like in [47], we shall use convolutiontype data of the form

$$
\begin{equation*}
\lambda_{j}(u)=\int_{\Omega} K\left(x-x_{j}\right) u(x) d x \tag{7.0.1}
\end{equation*}
$$

to build a sampling operator for weak data. Here $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ is a discrete set of points, and $K: \mathbb{R}^{d} \rightarrow \mathbb{R}$ is called a test kernel. Unlike in [47] we shall use stationary data, i.e., the support of the test kernel $K(\cdot, \cdot)$ shall be scaled with the mesh-norm of $X$ in $\Omega$. This implies that $\lambda_{j}(f)$ is some weighted local mean of $f$. This is a connection to Finite Volume Methods. In the usual Finite Volume Method, one simply chooses $K\left(\cdot, x_{j}\right) \equiv 1$ on supp $K\left(\cdot, x_{j}\right)$. The analysis of this special case has been done in [64], and we shall generalize results from [64] in this chapter.
The main improvement compared to [47] is, that the local data of the form (7.0.1) allows us to use only those functionals that are supported in $\Omega$. In [47] one has to use blurred domains $\tilde{\Omega}$ instead of $\Omega$.
We shall treat both recovery methods, the symmetric and the unsymmetric method. In [47] the later method is called unsymmetric local Petrov Galerkin Method because it is a major step toward the analysis of kernel based meshless methods for solving partial differential equations. We shall present error bounds and convergence rates for both, the symmetric and the unsymmetric method. The convergence proof of the unsymmetric methods uses the results from Chapter 6 to get appropriate stability estimates. The first step, however, is to derive sampling inequalities for the local data (7.0.1). To assure that the sampling operator $S_{\Lambda}=\left(\lambda_{1}, \ldots, \lambda_{N}\right)^{T}$ contains enough information we have to make certain assumptions on the kernel $K$. To be precise we shall impose the following conditions on $K$.

1. $\int_{\Omega} K\left(x-x_{j}\right) d x=1$ for all $x_{j} \in X$,
2. $\operatorname{supp}\left(K\left(\cdot-x_{j}\right)\right)=: V_{j} \subset \Omega$,
3. $c_{1} h_{X, \Omega} \leq \operatorname{diam}\left(V_{j}\right)=: k_{j} \leq c_{2} h_{X, \Omega}$, and
4. $\left\|K\left(\cdot-x_{j}\right)\right\|_{L_{p}(\Omega)} \leq C k_{j}^{-d / q} \leq C h_{X, \Omega}^{-d / q}, \quad \frac{1}{p}+\frac{1}{q}=1$.

Roughly speaking, $K\left(\cdot-x_{j}\right)$ should be a good approximation to the delta distribution $\delta_{x_{j}}$. We point out that especially the second assumption is different from the situation in [47], since we shall deal with stationary testing. This means that the support of our test kernel is scaled with the mesh-norm $h_{X, \Omega}$. The first step now is to show that kernels with these properties exist. We shall show that all appropriately scaled mollifiers, see [65] for instance for a definition, satisfy the conditions.

Lemma 7.0.5 Assume $\Psi: \mathbb{R}^{d} \rightarrow \mathbb{R}$ satisfies

1. $\operatorname{supp}(\Psi)=B(0,1)$,
2. $\sup _{x \in B(0,1)} \Psi(x)=C$, and
3. $\int_{B(0,1)} \Psi(x) d x=1$.

Now we scale $\Psi$ by a parameter $\delta>0$ via

$$
\Psi_{\delta}(\cdot)=\delta^{-d} \Psi\left(\frac{\dot{\zeta}}{\delta}\right) .
$$

Then, by definition $\Psi=\Psi_{1}$. Under these conditions, for $\frac{1}{p}+\frac{1}{q}=1$ we have

$$
\omega(d)^{-1 / p} \delta^{-d / q} \leq\left\|\Psi_{\delta}(\cdot)\right\|_{L_{p}(B(0, \delta))} \leq C \omega(d)^{1 / p} \delta^{-d / q}
$$

where $\omega(d)$ denotes the volume of the unit ball in $\mathbb{R}^{d}$, and the constant $C$ comes from Condition 2.

Proof:

$$
\begin{aligned}
\int_{\mathbb{R}^{d}}\left|\Psi_{\delta}(x)\right|^{p} d x & =\delta^{-d p} \int_{\mathbb{R}^{d}}\left|\Psi\left(\frac{x}{\delta}\right)\right|^{p} d x \\
& =\delta^{-d p} \int_{B(0, \delta)} \underbrace{\left|\Psi_{1}\left(\frac{x}{\delta}\right)\right|^{p}}_{\leq C^{p}} d x \\
& \leq \delta^{-d p} C^{p} \int_{B(0, \delta)} 1 d x=C^{p} \omega(d) \delta^{-d p+d}=C^{p} \omega(d) \delta^{d(1-p)} .
\end{aligned}
$$

Here, the constant $C$ is defined as in Assumption 2 of Lemma 7.0.5 by

$$
C:=\sup _{x \in B(0, \delta)}\left|\Psi_{1}\left(\frac{x}{\delta}\right)\right| .
$$

Now, since $\frac{1}{q}+\frac{1}{p}=1$ implies $1-p=-p / q$, we have the bound

$$
\left\|\Psi_{\delta}(\cdot)\right\|_{L_{p}(B(0, \delta))}^{p} \leq C^{p} \omega(d) \delta^{-d p / q} .
$$

Extracting the $p$-th root yields the second inequality.
The first one follows directly from Hölder's inequality since

$$
\begin{aligned}
1 & =\int_{B(0, \delta)} \Psi_{\delta}(x) d x \leq \int_{B(0, \delta)}\left|\Psi_{\delta}(x)\right| d x \leq \int_{B(0, \delta)}\left|\Psi_{\delta}(x)\right| \cdot 1 d x \\
& \leq \delta^{d / q} \omega(d)^{\frac{1}{p}}\left\|\Psi_{\delta}(x)\right\|_{L_{p}(B(0, \delta))}
\end{aligned}
$$

Multiplying by $\delta^{-d / q} \omega(d)^{-1 / p}$ gives the first inequality.
If we now consider the case that $V_{j}$ is contained in some ball $B$ with radius $\delta=c h_{X, \Omega}$, without loss of generality $B=B(0, \delta)$, this lemma shows that any mollifier scaled by a multiple of $h_{X, \Omega}$ satisfies our assumptions on the kernel.

### 7.1 Local Estimates

As in Chapter 4, we first obtain estimates on local domains $\mathcal{D} \subset \Omega$ and use a covering argument to get global results. We assume a domain $\mathcal{D}$ that is star-shaped with respect to a ball $B_{r}\left(x_{c}\right)$, and that is contained in a ball $B_{R}\left(x_{c}\right)$. In this case we know [38] that $\mathcal{D}$ satisfies an interior cone condition as well. We again denote the associated chunkiness parameter with

$$
\gamma=\frac{\delta_{\mathcal{D}}}{\rho_{\max }}
$$

where $\rho_{\text {max }}=\sup \{\rho: \mathcal{D}$ is star-shaped with respect to a ball of radius $\rho\}$, and $\delta_{\mathcal{D}}$ denotes the diameter of $\mathcal{D}$.

### 7.1.1 Norming Set Argument

Essentially following [64], we need a polynomial reproduction of degree $k$ with respect to a discrete set $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \mathcal{D}$, i.e., a family of functions $\left\{a_{j}^{(\alpha)}: j=1, \ldots, N\right\}$ satisfying

$$
D^{\alpha} q(x)=\sum_{j=1}^{N} a_{j}^{(\alpha)}(x) q\left(x_{j}\right)
$$

for every $\alpha \in \mathbb{N}_{0}^{d}$ with $|\alpha| \leq k$, all $x \in \mathcal{D}$ and all $q \in \pi_{k}(\mathcal{D})$. To bound the norm of the reproduction coefficients $\left\{a_{j}^{(\alpha)}: j=1, \ldots, N\right\}$, we can apply the rather general norming set approach [26] explained in Section 3.4. To do so, one has to show bounded invertibility for the actual sampling operator. This operator is defined by

$$
T: \pi_{k}\left(\mathbb{R}^{d}\right) \rightarrow \mathbb{R}^{N} \quad \text { with } \quad T(p)=\left(\lambda_{1}(p), \ldots, \lambda_{N}(p)\right)^{T}
$$

where the functionals $\lambda_{j}$ are defined by (7.0.1).

Lemma 7.1.1 Let $\mathcal{D} \subset \mathbb{R}^{d}$ be a compact domain that satisfies an interior cone condition with angle $\theta$ and radius $r>0$. Suppose $\mathcal{D}$ is covered by subregions $V_{j}$, i.e., $\mathcal{D} \subset \cup_{j=1}^{N} V_{j}$. Suppose that every ball $B\left(x_{0}, h\right) \subset \mathcal{D}$ contains at least one $V_{j}$. Then, provided that

$$
\begin{equation*}
h \leq \frac{1}{C} \cdot \frac{r \sin \theta}{4 k^{2}(1+\sin \theta)}, \tag{7.1.1}
\end{equation*}
$$

where the constant $C=C(K)>0$ depends only on the kernel $K$, the mapping $T: \pi_{k}\left(\mathbb{R}^{d}\right) \rightarrow \mathbb{R}^{N}$ is injective. The operator norm is bounded from below by

$$
\|T(p)\|_{\infty}=\max _{1 \leq j \leq N}\left|\lambda_{j}(p)\right| \geq \frac{1}{2}\|p\|_{L_{\infty}(\mathcal{D})}, \quad p \in \pi_{k}\left(\mathbb{R}^{d}\right) .
$$

Proof: Suppose $p \in \pi_{k}(\mathcal{D})$ with $\|p\|_{\infty}=1$ is given. Then there exists a point $x_{0} \in \mathcal{D}$ with $\left|p\left(x_{0}\right)\right|=1$. For this point we can choose a cone $C\left(x_{0}\right)$ with angle $\theta$ and radius $r>0$ that is completely contained in $\mathcal{D}$. By the essentially same argument as in [64, Lemma 4.8] we find a volume $V_{j}$, such that

$$
V_{j} \subseteq B(y, h) \subseteq C\left(x_{0}\right) \subseteq \mathcal{D},
$$

with $y=x_{0}+\frac{h}{\sin \theta} \xi$, and $h \leq \frac{r \sin \theta}{1+\sin \theta}$. Moreover, using Markov's inequality [64, Lemma 4.8], we have the bound

$$
\left|p\left(x_{0}\right)-p(x)\right| \leq h \cdot \frac{2 k^{2}(1+\sin \theta)}{r \sin \theta}, \quad x \in V_{j} .
$$

Now we can estimate as follows

$$
\begin{aligned}
\left|p\left(x_{0}\right)-\lambda_{j}(p)\right| & =\left|\int_{\Omega} K\left(x, x_{j}\right) p\left(x_{0}\right) d x-\int_{\Omega} K\left(x, x_{j}\right) p(x) d x\right| \\
& \leq \int_{\Omega}\left|K\left(x, x_{j}\right)\right| \cdot\left|p\left(x_{0}\right)-p(x)\right| d x \\
& \leq \int_{\Omega}\left|K\left(x, x_{j}\right)\right| d x \cdot\left\|p\left(x_{0}\right)-p\right\|_{L_{\infty}\left(V_{j}\right)} \\
& \leq C \cdot h \frac{2 k^{2}(1+\sin \theta)}{r \sin \theta} \\
& \leq \frac{1}{2}
\end{aligned}
$$

where the constant $C$ comes from the condition $\left\|K\left(x, x_{j}\right)\right\|_{L_{1}(\Omega)} \leq C$. This implies $\lambda_{j}(p) \geq \frac{1}{2}\|p\|_{L_{\infty}(\mathcal{D})}$ for this particular $j$.

An important observation is that one has used only those functionals whose associated kernels have support in $\mathcal{D} \subset \Omega$. Hence, we are in an analogous situation to [64]. This is a huge advantage of these stationary methods in contrast to the situation in [47].

Remark 7.1.2 Only those $V_{j}$ 's with $V_{j} \subset \mathcal{D}$ are used to derive Lemma 7.1.1.

As stated in the following lemma from [64], we get a stable polynomial reproduction.
Lemma 7.1.3 Suppose $\mathcal{D}$ is compact and satisfies a cone condition with angle $\theta$ and radius $r>0$. Then, for every $|\alpha| \leq k$ and every $x \in \mathcal{D}$ there exist numbers $a_{1}^{(\alpha)}(x), \ldots, a_{N}^{(\alpha)}(x)$ such that for every $p \in \pi_{k}\left(\mathbb{R}^{d}\right)$

- $\left\|D^{\alpha} p\right\|_{L_{\infty}(\mathcal{D})} \leq\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|}\|p\|_{L_{\infty}(\mathcal{D})}$,
- $D^{\alpha} p(x)=\sum_{j=1}^{N} \lambda_{j}(p) a_{j}^{(\alpha)}(x)$,
- $\sum_{j=1}^{N}\left|a_{j}^{(\alpha)}(x)\right| \leq 2\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|}$,
holds.


### 7.1.2 Local Polynomial Approximation

In this section we essentially repeat the arguments from [38], but extend them in the way that we deal with error representation for approximation by averaged Taylor polynomials [11] not in the $L_{\infty}$-norm, but in a more general $L_{p}$-norm. The Taylor polynomials are defined in Section 3.3.2. To deal with fractional Sobolev spaces, we need a version of the BrambleHilbert Lemma (see Lemma 3.3.11) that applies to $u \in W_{p}^{k+s}(\mathcal{D})$ with $0<s \leq 1$. To start with, we have the following estimate, which is the $L_{p}$-analogon to [38, Lemma 2.6].
Lemma 7.1.4 Suppose $p \geq 1$ and $k \geq 1$. For all $u \in W_{p}^{k}(\mathcal{D})$ and $P \in \pi_{k}\left(\mathbb{R}^{d}\right)$, the residual of the averaged Taylor-approximation is bounded by

$$
\begin{equation*}
\left\|R^{k+1} u\right\|_{L_{p}(\mathcal{D})} \leq C_{k, d, p, \gamma} \delta_{\mathcal{D}}^{k}|u-P|_{W_{p}^{k}(\mathcal{D})} \tag{7.1.2}
\end{equation*}
$$

Proof: We follow [38, Proof of Lemma 2.6]. We use a decomposition of the form

$$
R^{k+1} u=R^{k+1}(u-P)=R^{k}(u-P)+\left(Q^{k}-Q^{k+1}\right)(u-P),
$$

where $P \in \pi_{k}\left(\mathbb{R}^{d}\right)$ denotes an arbitrary polynomial. By the triangle inequality and the Bramble-Hilbert Lemma 3.3.11, we obtain

$$
\begin{equation*}
\left\|R^{k+1} u\right\|_{L_{p}(\mathcal{D})} \leq C_{k, d, p, \gamma} \delta_{\mathcal{D}}^{k}|u-P|_{W_{p}^{k}(\mathcal{D})}+\left\|\left(Q^{k}-Q^{k+1}\right)(u-P)\right\|_{L_{p}(\mathcal{D})} \tag{7.1.3}
\end{equation*}
$$

The second term can be bounded as follows. By the definition of $Q^{k}$, using the fact that $\max \phi \leq C \operatorname{diam}(B)^{-d}=: C \rho^{-d}$ and the identity $\sum_{|\alpha|_{d}=k} \frac{1}{\alpha!}=\frac{d^{k}}{k!}$ we get

$$
\begin{align*}
& \left\|\left(Q^{k}-Q^{k+1}\right)(u-P)\right\|_{L_{p}(\mathcal{D})}^{p} \leq \\
\leq & \int_{\mathcal{D}}\left(\sum_{|\alpha|=k} \int_{B_{\rho}} \frac{\phi(y)\|x-y\|_{2}^{k}\left|D^{\alpha}(u-P)(y)\right|}{\alpha!} d y\right)^{p} d x \\
\leq & C \operatorname{vol}(\mathcal{D}) \delta_{\mathcal{D}}^{k p} \rho^{-d p} \frac{d^{k p}}{k!^{p}}\left(\max _{|\alpha|=k} \int_{B_{\rho}}\left|D^{\alpha}(u-P)(y)\right| d y\right)^{p} \\
\leq & C \delta_{\mathcal{D}}^{d} \delta_{\mathcal{D}}^{k p} \rho^{-d p} \frac{d^{k p}}{k!p}\left(\max _{|\alpha|=k} \int_{B_{\rho}}\left|D^{\alpha}(u-P)(y)\right| d y\right)^{p} . \tag{7.1.4}
\end{align*}
$$

From [38, Proof of Lemma 2.6] we see that for any $\alpha \in \mathbb{N}_{0}^{d}$ with $|\alpha|=k$

$$
\int_{B_{\rho}}\left|D^{\alpha}(u-P)(y)\right| d y \leq \omega(d)^{1-1 / p} \rho^{d-d / p}|u-P|_{W_{p}^{k}(\mathcal{D})} .
$$

Inserting this into the estimate (7.1.4) and using $\delta_{\mathcal{D}} / \rho \leq 2 \gamma$, we arrive at the estimate

$$
\left\|\left(Q^{k}-Q^{k+1}\right)(u-P)\right\|_{L_{p}(\mathcal{D})} \leq C \omega(d)^{1-1 / p} \frac{2^{d / p} d^{k}}{k!} \gamma^{d / p} \delta_{\mathcal{D}}^{k}|u-P|_{W_{p}^{k}(\mathcal{D})}
$$

Consequently, combining the inequality above with (7.1.3) yields (7.1.2).
Now we can treat the fractional case as in [38, Prop. 2.7].
Proposition 7.1.5 Let $0<s \leq 1$. If $u \in W_{p}^{k+s}(\mathcal{D})$, then

$$
\begin{equation*}
\left\|R^{k+1} u\right\|_{L_{p}(\mathcal{D})} \leq C_{k, d, p, \gamma} \delta_{\mathcal{D}}^{k+s}|u|_{W_{p}^{k+s}(\mathcal{D})} \tag{7.1.5}
\end{equation*}
$$

Proof: We follow [38, Proof of Proposition 2.7]. The case $s=1$ is a consequence of Lemma 3.3.11, so we may assume that $s<1$. Let $P=Q^{k+1} u \in \pi_{k}\left(\mathbb{R}^{d}\right)$. As in [39, equation (7)] we have for all $\alpha \in \mathbb{N}_{0}^{d}$ with $|\alpha|=k$

$$
D^{\alpha} Q^{k+1} u(x)=Q^{1} D^{\alpha} u(x)=\int_{B_{\rho}} \phi(y) D^{\alpha} u(y) d y \in \mathbb{R}
$$

Since $\int_{B_{\rho}} \phi(y) d y=1$, we note that

$$
\int_{\mathcal{D}}\left|D^{\alpha} u(x)-D^{\alpha} Q^{k+1} u(x)\right|^{p} d x=\int_{\mathcal{D}}\left|\int_{B_{\rho}} \phi(y)\left(D^{\alpha} u(x)-D^{\alpha} u(y)\right) d y\right|^{p} d x .
$$

Now we can compute as follows,

$$
\begin{aligned}
& \int_{\mathcal{D}}\left(\int_{B_{\rho}}|\phi(y)| \cdot|x-y|^{s+d / p} \frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|}{|x-y|^{s+d / p}} d y\right)^{p} d x \\
\leq & C \rho^{-d p} \delta_{\mathcal{D}}^{s p+d} \int_{\mathcal{D}}\left(\int_{B_{\rho}} \frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|}{|x-y|^{s+d / p}} d y\right)^{p} d x .
\end{aligned}
$$

We may use Hölder's inequality, which yields

$$
\begin{aligned}
& \int_{\mathcal{D}}\left(\int_{B_{\rho}} 1 \cdot \frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|}{|x-y|^{s+d / p}} d y\right)^{p} d x \\
\leq & C \int_{\mathcal{D}}\left[\rho^{d / q}\left(\int_{B_{\rho}} \frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|^{p}}{|x-y|^{s p+d}} d y\right)^{1 / p}\right]^{p} d x \\
\leq & C \rho^{d p / q} \int_{\mathcal{D}} \int_{\mathcal{D}}\left(\frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|}{|x-y|^{s+d / p}}\right)^{p} d y d x .
\end{aligned}
$$

This implies

$$
\begin{aligned}
& \int_{\mathcal{D}}\left|D^{\alpha} u(x)-D^{\alpha} Q^{k+1} u(x)\right|^{p} \\
& \leq C \rho^{-d p+d p / q} \delta_{\mathcal{D}}^{s p+d} \int_{\mathcal{D}} \int_{\mathcal{D}}\left(\frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|}{|x-y|^{s+d / p}}\right)^{p} d y d x \text {. }
\end{aligned}
$$

Now we can use $\delta_{\mathcal{D}}=2 \gamma \rho$ to derive $\rho^{-d p+d p / q} \delta_{\mathcal{D}}^{d}=(2 \gamma)^{d} \rho^{d p\left(-1+\frac{1}{p}+\frac{1}{q}\right)}=(2 \gamma)^{d}$, which gives

$$
\begin{aligned}
& \int_{\mathcal{D}}\left|D^{\alpha} u(x)-D^{\alpha} Q^{k+1} u(x)\right|^{p} d x \\
& \leq C(2 \gamma)^{d} \delta_{\mathcal{D}}^{s p} \int_{\mathcal{D}} \int_{\mathcal{D}}\left(\frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|}{|x-y|^{s+d / p}}\right)^{p} d y d x .
\end{aligned}
$$

If we sum over all $\alpha \in \mathbb{N}_{0}^{d}$ with $|\alpha|=k$, we find with $P=Q^{k+1} u$

$$
|u-P|_{W_{p}^{k}(\mathcal{D})}^{p} \leq C_{d, p, \gamma}^{p} \delta_{\mathcal{D}}^{s p} \sum_{|\alpha|=k} \int_{\mathcal{D}} \int_{\mathcal{D}} \frac{\left|D^{\alpha} u(x)-D^{\alpha} u(y)\right|^{p}}{|x-y|^{d+s p}} d y d x .
$$

The summed double integral on the right hand is just $|u|_{W_{p}^{k+s}(\mathcal{D})}^{p}$. Taking the $p^{t h}$ root of both sides, we obtain

$$
|u-P|_{W_{p}^{k}(\mathcal{D})} \leq C_{d, p, \gamma} \delta_{\mathcal{D}}^{s}|u|_{W_{p}^{k+s}(\mathcal{D})}
$$

Lemma 7.1.4 yields the result.
Having the applications in mind, we also need a generalization of Proposition 7.1.5 to derivatives. Following [38, Cor. 2.8], such estimates can be easily obtained using the identity

$$
D^{\alpha} Q^{k+1}=Q^{k-|\alpha|} D^{\alpha}
$$

from [11].
Corollary 7.1.6 Let $0<s \leq 1$. For $u \in W_{p}^{k+s}(\mathcal{D})$ we have

$$
\left\|D^{\alpha} u-D^{\alpha} Q^{k+1} u\right\|_{L_{p}(\mathcal{D})} \leq C_{k, n, p, \gamma} \gamma_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})}
$$

provided that $|\alpha|<k$ and $p>1$.
Proof: Using the identity $D^{\alpha} Q^{k+1}=Q^{k-|\alpha|} D^{\alpha}$ this result follows directly from Proposition 7.1.5 and the obvious bound

$$
\left|D^{\alpha} u\right|_{W_{p}^{k+s-|\alpha|}(\mathcal{D})} \leq|u|_{W_{p}^{k+s}(\mathcal{D})} .
$$

Now we have the main ingredients to show a local sampling inequality. We point out that, having the applications to recovery from weak data in mind, we assumed only low regularity of the function $u$. Hence, we may handle convolution-type data of functions in the sense of (7.0.1), even if point-evaluations are not well-defined, that is, the functions to be recovered are not continuous.

### 7.1.3 Local Sampling Inequality

Following [38,67,64], we can combine the local polynomial reproduction with the approximation error estimates for polynomials to get the following result.

Theorem 7.1.7 Let $k$ be a positive integer, $1 \leq p<\infty, 0<s \leq 1$, and let $\alpha$ be a multiindex satisfying $k>|\alpha|$.
Suppose $\mathcal{D} \subseteq B\left(x_{c}, R\right)$ is star-shaped with respect to $B\left(x_{c}, r\right)$ and covered with volumes $V_{j}$ such that for every $x_{0} \in \mathcal{D}$ the ball $B\left(x_{0}, h\right) \subset \mathcal{D}$ with radius $h$ contains at least one of these volumes $V_{j}$. Let $\delta_{\mathcal{D}}$ denote the diameter of $\mathcal{D}$. If $h$ satisfies condition (7.1.1), then we have the bound

$$
\begin{aligned}
\left\|D^{\alpha} u\right\|_{L_{p}(\mathcal{D})} & \leq C\left(1+\delta_{\mathcal{D}} / h\right)^{d / p} \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})} \\
& +C \delta_{\mathcal{D}}^{-|\alpha|+d / p} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|
\end{aligned}
$$

for every $u \in W_{p}^{k+s}(\mathcal{D})$ with a constant $C=C(|\alpha|, k, d, p, \theta)$.
Proof: Corollary 7.1.6 yields

$$
\left\|D^{\alpha} u-D^{\alpha} Q^{k+1} u\right\|_{L_{p}(\mathcal{D})} \leq C \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})} .
$$

We have for every $P \in \pi_{k}\left(\mathbb{R}^{d}\right)$ and all $x \in \mathcal{D}$

$$
\begin{aligned}
\left|D^{\alpha} P(x)\right| & =\left|\sum_{j=1}^{N} \lambda_{j}(P) a_{j}^{(\alpha)}(x)\right| \\
& =\left|\sum_{j=1}^{N}\left[\lambda_{j}(P)-\lambda_{j}(u)+\lambda_{j}(u)\right] a_{j}^{(\alpha)}(x)\right| \\
& \leq \sum_{j=1}^{N}\left|a_{j}^{(\alpha)}(x)\right| \int_{V_{j}}\left|K\left(y, x_{j}\right)\right| \cdot|P(x)-u(x)| d y \\
& +\sum_{j=1}^{N}\left|a_{j}^{(\alpha)}(x)\right| \cdot\left|\lambda_{j}(u)\right| \\
& \leq C\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|} \max _{1 \leq j \leq N}\left\|K\left(\cdot, x_{j}\right)\right\|_{L_{q}\left(V_{j}\right)} \cdot\|u-P\|_{L_{p}(\mathcal{D})} \\
& +C\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right| \\
& \leq C\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|}\left(h^{-d / p}\|u-P\|_{L_{p}(\mathcal{D})}+\max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|\right)
\end{aligned}
$$

with $\frac{1}{p}+\frac{1}{q}=1$. Specifying $P=Q^{k+1} u$ yields

$$
\left|D^{\alpha} P(x)\right| \leq C\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|}\left(C h^{-d / p} \delta_{\mathcal{D}}^{k+s}|u|_{W_{p}^{k+s}(\mathcal{D})}+\max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|\right) .
$$

From this we get

$$
\begin{aligned}
&\left\|D^{\alpha} Q^{k+1} u\right\|_{L_{p}(\mathcal{D})} \leq \\
& \leq C \operatorname{vol}(\mathcal{D})^{\frac{1}{p}}\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|}\left(C h^{-d / p} \delta_{\mathcal{D}}^{k+s}|u|_{W_{p}^{k+s}(\mathcal{D})}+\max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|\right) \\
& \leq C\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|} \delta_{\mathcal{D}}^{d / p}\left(h^{-d / p} \delta_{\mathcal{D}}^{k+s}|u|_{W_{p}^{k+s}(\mathcal{D})}+\max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|\right) \\
& \leq C\left(\frac{2 k^{2}}{r \sin \theta}\right)^{|\alpha|}\left(\left(\delta_{\mathcal{D}} / h\right)^{d / p} \delta_{\mathcal{D}}^{k+s}|u|_{W_{p}^{k+s}(\mathcal{D})}+\delta_{\mathcal{D}}^{d / p} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|\right) \\
& \leq C\left(\frac{2 k^{2} \delta_{\mathcal{D}}}{r \sin \theta}\right)^{|\alpha|}\left(\delta_{\mathcal{D}} / h\right)^{d / p} \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})}+ \\
&+ C\left(\frac{2 k^{2} \delta_{\mathcal{D}}}{r \sin \theta}\right)^{|\alpha|} \delta_{\mathcal{D}}^{d / p-|\alpha|} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right| .
\end{aligned}
$$

At this point we use the following chain of inequalities

$$
1 \leq \gamma \leq \frac{\delta_{\mathcal{D}}}{r} \leq 2 \frac{R}{r}=\frac{1}{\sin (\theta / 2)}
$$

from [64] to see that

$$
\left(\frac{2 k^{2} \delta_{\mathcal{D}}}{r \sin \theta}\right)^{|\alpha|} \leq C(\theta, \alpha, k, d) \leq C
$$

Using the triangle inequality we finally obtain

$$
\begin{aligned}
& \left\|D^{\alpha} u\right\|_{L_{p}(\mathcal{D})} \leq\left\|D^{\alpha} u-D^{\alpha} Q^{k+1} u\right\|_{L_{p}(\mathcal{D})}+\left\|D^{\alpha} Q^{k+1} u\right\|_{L_{p}(\mathcal{D})} \\
\leq & C \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})}+2 C\left(\delta_{\mathcal{D}} / h\right)^{d / p} \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})} \\
+ & 2 C \delta_{\mathcal{D}}^{-|\alpha|+d / p} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right| \\
\leq & C\left(1+\left(\delta_{\mathcal{D}} / h\right)^{d / p}\right) \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})} \\
+ & C \delta_{\mathcal{D}}^{-|\alpha|+d / p} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|
\end{aligned}
$$

The previous result is also true for Sobolev semi-norms.
Corollary 7.1.8 Under the assumptions of Theorem 7.1.7 we find

$$
\begin{aligned}
|u|_{W_{p}^{|\alpha|}(\mathcal{D})} & \leq C_{k, d, p,|\alpha|}\left(1+\left(\delta_{\mathcal{D}} / h\right)^{d / p}\right) \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})} \\
& +C \delta_{\mathcal{D}}^{-|\alpha|+d / p} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|
\end{aligned}
$$

for all $u \in W_{p}^{k}(\Omega)$.

Proof: Since $\#\left\{\beta \in \mathbb{N}_{0}^{d}:|\beta|=|\alpha|\right\}=\binom{|\alpha|+d-1}{d-1}=\mathcal{O}\left(|\alpha|^{d-1}\right)$ we find that

$$
\begin{aligned}
|u|_{W_{p}^{|\alpha|}(\mathcal{D})} & \leq\binom{|\alpha|+d-1}{d-1} \max _{|\beta|=|\alpha|}\left\|D^{\beta} u\right\|_{L_{p}(\mathcal{D})} \\
& \leq C_{d, q,|\alpha|} \max _{|\beta|=|\alpha|}\left\|D^{\beta} u\right\|_{L_{p}(\mathcal{D})} \\
& \leq C_{k, d, p, q,|\alpha|}\left(1+\left(\delta_{\mathcal{D}} / h\right)^{d / p}\right) \delta_{\mathcal{D}}^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\mathcal{D})} \\
& +C \delta_{\mathcal{D}}^{-|\alpha|+d / p} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right| .
\end{aligned}
$$

Now we may use a covering argument to derive global estimates.

### 7.2 Global Estimates

Let us suppose our global region $\Omega$ is bounded and satisfies an interior cone condition with radius $R_{\max }$ and angle $\theta$. Let $h=h_{X, \Omega}$ be the fill distance of the discrete set $X \subseteq \Omega$. Essentially following [67, 38], we shall again use the covering $\Omega=\bigcup_{t \in T_{h}} \mathcal{D}_{t}$ described in Theorem 3.3.10. Please note that this construction allows us to cancel the term $\delta_{\mathcal{D}} / h$, which appeared in the local estimates. After these preparatory steps, we are in the position to state and prove the main result of this section.
Theorem 7.2.1 Suppose $\Omega$ is bounded and satisfies an interior cone condition with radius $r$ and angle $\theta$. Let $k$ be a positive integer, $0<s \leq 1,1 \leq p<\infty$, and let $\alpha$ be a multi-index satisfying $k>|\alpha|$. Suppose that every ball $B\left(x_{0}, h\right) \subset \Omega$ of radius $h \leq Q(k, \theta) R_{\max }$ contains at least one volume $V_{j}$. Then for $u \in W_{p}^{k+s}(\Omega)$,

$$
\begin{equation*}
|u|_{W_{p}^{|\alpha|}(\Omega)} \leq C h^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\Omega)}+\tilde{C} h^{-|\alpha|} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right| . \tag{7.2.1}
\end{equation*}
$$

Proof: We find

$$
\begin{aligned}
|u|_{W_{p}^{|\alpha|}(\Omega)}= & \left(\sum_{|\beta|=|\alpha|} \int_{\Omega}\left|D^{\beta} u(x)\right|^{p} d x\right)^{1 / p} \\
\leq & \left(\sum_{t \in T_{h}} \sum_{|\beta|=|\alpha|} \int_{\mathcal{D}_{t}}\left|D^{\beta} u(x)\right|^{p} d x\right)^{1 / p}=\left(\sum_{t \in T_{h}}|u|_{W_{p}^{|\alpha|}\left(\mathcal{D}_{t}\right)}^{p}\right)^{1 / p} \\
\leq & \left(\sum_{t \in T_{h}}\left(C h^{k+s-|\alpha|}|u|_{W_{p}^{k+s}\left(\mathcal{D}_{t}\right)}+\tilde{C} h^{\frac{d}{p}-|\alpha|} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|\right)^{p}\right)^{\frac{1}{p}} \\
\leq & \left(\sum_{t \in T_{h}}\left(C h^{k+s-|\alpha|}|u|_{W_{p}^{k+s}\left(\mathcal{D}_{t}\right)}\right)^{p}\right)^{\frac{1}{p}} \\
& +\left(\# T_{h}\right)^{\frac{1}{p}} \tilde{C} h^{\frac{d}{p}-|\alpha|} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|
\end{aligned}
$$

where $\# T_{h}$ denotes the number of elements in $T_{h}$. As in [38] using again Theorem 3.3.10 we get

$$
\sum_{t \in T_{h}}|u|_{W_{p}^{k+s}\left(\mathcal{D}_{t}\right)}^{p} \leq M_{1}|u|_{W_{p}^{k+s}(\Omega)}^{p} .
$$

Combining the results yields

$$
\begin{aligned}
|u|_{W_{p}^{|\alpha|}(\Omega)} \leq & C M_{1}^{1 / p} h^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\Omega)} \\
& +\tilde{C}\left(\# T_{h}\right)^{\frac{1}{p}} h^{-|\alpha|+\frac{d}{p}} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|
\end{aligned}
$$

According to Theorem 3.3.10 with $R=Q(k, \theta) h$ one finds $\# T_{h}<C h^{-d}$, and therefore

$$
|u|_{W_{p}^{|\alpha|}(\Omega)} \leq C h^{k+s-|\alpha|}|u|_{W_{p}^{k+s}(\Omega)}+\tilde{C} h^{-|\alpha|} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right| .
$$

We can apply these results in the framework of optimal recovery.

### 7.3 Error Estimates for Optimal Recovery

In this section we shall show briefly how our results can be used to derive error estimates in the framework of generalized optimal recovery. A more detailed overview can be found in [58] and [65]. We start with a short introduction to kernel-based optimal recovery. From now on, we assume that $\Omega \subset \mathbb{R}^{d}$ is bounded by a Lipschitz boundary and satisfies an interior cone condition with radius $r$ and angle $\theta$. Here, $\Phi: \mathbb{R}^{d} \rightarrow \mathbb{R}$ is a positive definite translation invariant kernel. We assume that the kernel $\Phi$ satisfies

$$
\hat{\Phi}(\omega) \sim\left(1+\|\omega\|_{2}^{2}\right)^{-\tau} .
$$

This means that $\Phi$ is the reproducing kernel of the Sobolev space $W_{2}^{\tau}(\Omega)$ with an equivalent norm. For general functionals from the dual space we can derive an easy representation using the reproducing property of the kernel. If $\lambda \in\left(W_{2}^{\tau}(\Omega)\right)^{*}$, its Riesz-representer $R_{\lambda}$ takes the form $R_{\lambda}(x)=\lambda \Phi(\cdot-x)$. This can be seen by

$$
R_{\lambda}(x)=\left(R_{\lambda}, \Phi(\cdot-x)\right)_{W_{2}^{\tau}(\Omega)}=\lambda \Phi(\cdot-x) .
$$

For a given $f \in W_{2}^{\tau}(\Omega)$ and the functionals $\lambda_{j}$ from the previous section we can formulate the smoothed optimal recovery problem

$$
\begin{equation*}
\min _{s \in W_{2}^{\tau}(\Omega)} \sum_{j=1}^{N}\left|\lambda_{j}(s-f)\right|^{2}+\nu\|s\|_{W_{2}^{\tau}(\Omega)}^{2} \tag{7.3.1}
\end{equation*}
$$

where $\nu \geq 0$ is called the smoothing parameter. We note that again the special case $\nu=0$ corresponds to finding a generalized interpolant, i.e., a function $s_{f} \in W_{2}^{\tau}(\Omega)$ that satisfies the generalized interpolation conditions

$$
\lambda_{j}(f)=\lambda_{j}\left(s_{f}\right) \text { for all } j=1, \ldots, N .
$$

It is well known [65] that there always exists a solution $s_{f}^{(\nu)} \in \operatorname{span}\left\{\lambda_{j}^{x} \Phi(\cdot-x)\right\}$ to the optimization problem (7.3.1). The coefficients $a \in \mathbb{R}^{N}$ can be found by solving the linear system

$$
\left(A_{\Phi, \Lambda}+\nu \mathrm{Id}\right) a=f_{\Lambda}
$$

where

$$
A_{\Phi, \Lambda}:=\left(\lambda_{j} \lambda_{k} \Phi(\cdot-\cdot)\right)_{j, k=1, \ldots, N}, f_{\Lambda}=\left(\lambda_{1}(f), \ldots, \lambda_{N}(f)\right)^{T}
$$

Furthermore, it is known [65] that $s_{f}:=s_{f}^{(0)}$ is a best approximation, which implies

$$
\left\|s_{f}\right\|_{W_{2}^{\tau}(\Omega)} \leq\|f\|_{W_{2}^{\tau}(\Omega)}
$$

Now we are prepared to formulate our error estimate for the smoothed optimal recovery problem (7.3.1).

Theorem 7.3.1 Under the conditions from above we find a positive constant $C=C(\Omega)>$ 0 such that for all $f \in W_{2}^{\tau}(\Omega)$, all $\nu \geq 0$ and all discrete sets $X=\left\{x_{1}, \ldots, x_{N}\right\} \subset \Omega$ with sufficiently small fill distance $h$,

$$
\left\|f-s_{f}^{(\nu)}\right\|_{L_{2}(\Omega)} \leq C\left(h^{\tau}+\sqrt{\nu}\right)\|f\|_{W_{2}^{\tau}(\Omega)}
$$

where $s_{f}^{(\nu)}$ denotes the solution to (7.3.1).
Proof: As in [67] we derive the following two inequalities

$$
\begin{aligned}
\left\|s_{f}^{(\nu)}\right\|_{W_{2}^{\tau}(\Omega)} & \leq\|f\|_{W_{2}^{\tau}(\Omega)}, \\
\max _{1 \leq j \leq N}\left|\lambda_{j}\left(f-s_{f}^{(\nu)}\right)\right| & \leq \sqrt{\nu}\|f\|_{W_{2}^{\tau}(\Omega)} .
\end{aligned}
$$

Now we can apply our sampling inequality (7.2.1) to obtain

$$
\begin{aligned}
& \left\|f-s_{f}^{(\nu)}\right\|_{L_{2}(\Omega)} \leq C h^{\tau}\left\|f-s_{f}^{(\nu)}\right\|_{W_{2}^{\tau}(\Omega)}+\tilde{C} \max _{1 \leq j \leq N}\left|\lambda_{j}\left(f-s_{f}^{(\nu)}\right)\right| \\
\leq & C h^{\tau}\left(\|f\|_{W_{2}^{\tau}(\Omega)}+\left\|s_{f}^{(\nu)}\right\|_{W_{2}^{\tau}(\Omega)}\right)+\tilde{C} \max _{1 \leq j \leq N}\left|\lambda_{j}\left(f-s_{f}^{(\nu)}\right)\right| \\
\leq & C\left(h^{\tau}+\sqrt{\nu}\right)\|f\|_{W_{2}^{\tau}(\Omega)} .
\end{aligned}
$$

Of course we can also couple $\nu$ and $\tau$ to derive a sampling order.
Corollary 7.3.2 For the special choice $\nu=h^{2 \tau}$ we even find optimal order, i.e.,

$$
\left\|f-s_{f}^{\left(h^{2 \tau}\right)}\right\|_{L_{2}(\Omega)} \leq C h^{\tau}\|f\|_{W_{2}^{\tau}(\Omega)} .
$$

Now we can apply this result in the framework of unsymmetric weak recovery.

### 7.4 Unsymmetric Weak Recovery

In this section we show how suitable sampling inequalities can be used to derive upper bounds for a weak unsymmetric recovery process from [47]. We briefly recall this process. Since we are going to apply a non-symmetric method, we have to distinguish carefully between the tRial and the teSt side. As outlined in Section 2.3.1 we shall denote the test kernel with $S: \mathbb{R}^{d} \times \mathbb{R}^{d} \rightarrow \mathbb{R}$ and the discretization by $X_{s}=\left\{x_{1}, \ldots, x_{N_{s}}\right\} \subset \Omega$ with the associated fill distance $h_{s}$. On the $\mathbf{t}$ ial side we denote the kernel with $R: \mathbb{R}^{d} \times \mathbb{R}^{d} \rightarrow \mathbb{R}$ and the discretization with $Y_{r}=\left\{y_{1}, \ldots, y_{N_{r}}\right\} \subset \Omega$ with fill distance $h_{r}$. According to [45, Chapters 3.4 and 3.5] we use weak data of the form given in (2.1.7), where we replace $K$ by $S$, namely

$$
\begin{equation*}
\lambda_{j}(u)=\int_{\Omega} S\left(x, x_{j}\right) u(x) d x, \text { for } x_{j} \in X_{s} \tag{7.4.1}
\end{equation*}
$$

where the kernel $S: \mathbb{R}^{d} \times \mathbb{R}^{d} \rightarrow \mathbb{R}$ needs to fulfill the following properties,

1. $\int_{\Omega} S\left(x, x_{j}\right) d x=1$ for all $x_{j} \in X_{s}$,
2. $\operatorname{supp}\left(S\left(\cdot, x_{j}\right)\right)=V_{j}, \operatorname{diam}\left(V_{j}\right)=\delta_{V_{j}} \approx h_{s}$,
3. $\left\|S\left(\cdot, x_{j}\right)\right\|_{L_{p}(\Omega)} \leq C \delta_{V_{j}}^{-d / q} \approx h_{s}^{-d / q}, \quad \frac{1}{p}+\frac{1}{q}=1$.

Under these conditions we may apply Theorem 7.2.1, i.e., there is a sampling inequality of the form

$$
|u|_{W_{p}^{|\alpha|}(\Omega)} \leq C h_{s}^{\tau+d / p-|\alpha|}|u|_{W_{p}^{k+s}(\Omega)}+\tilde{C} h_{s}^{-|\alpha|} \max _{1 \leq j \leq N}\left|\lambda_{j}(u)\right|
$$

for all $u \in W_{2}^{\tau}(\Omega)$ and all discrete test sets $X_{s}$. Now we consider the following problem. An unknown function $f \in W_{2}^{1}(\Omega)$ has to be recovered approximately from its data $\left(\lambda_{1}(f), \ldots, \lambda_{N}(f)\right)^{T}$. From the previous chapters we know that there is a good but unknown approximation. It is given by the best approximation from the trial space $s_{f} \in V_{R, Y_{r}}=\operatorname{span}\left\{R\left(\cdot-y_{j}\right) \mid y_{j} \in Y_{r}\right\}$ to the function $f \in W_{2}^{1}(\Omega)$. In Chapter 8 we obtain under certain conditions an error estimate of the form

$$
\begin{equation*}
\left\|f-s_{f}\right\|_{L_{2}(\Omega)} \leq h_{r}\|f\|_{W_{2}^{1}(\Omega)}, \tag{7.4.2}
\end{equation*}
$$

showing linear approximation order, which is optimal [48]. Unfortunately, this best approximation is in general numerically unavailable. The idea from [47] is to solve the system

$$
\text { find } s_{r} \in V_{R, Y_{r}}: \lambda_{j}\left(f-s_{r}\right)=0 \text { for all } 1 \leq j \leq N_{s}
$$

to produce an approximation. Since this, however, is an unsymmetric kernel method, we first have to prove that this system is solvable. To do so, we proceed along the lines of [47], which we briefly outlined in Section 2.3.1. We assume an inverse inequality of the form

$$
\|s\|_{W_{2}^{1}(\Omega)} \leq C \gamma\left(Y_{r}\right)\|s\|_{L_{2}(\Omega)} \quad \text { for all } s \in V_{R, Y_{r}}
$$

Unfortunately, the value of $\gamma\left(Y_{r}\right)$ is in general not known. There is a result in this direction in Chapter 6, namely if $R$ is a radial basis function with algebraic smoothness $\tau>d / 2$ and
if $Y_{r}$ is separated from the boundary then $\gamma\left(Y_{r}\right) \approx q_{r}^{-\tau}$. However, we can always make sure that the test-meshnorm is small enough to stabilize the reconstruction, i.e., we shall assume

$$
\begin{equation*}
h_{s} \gamma\left(Y_{r}\right) \leq \frac{1}{2 C}, C>1 \tag{7.4.3}
\end{equation*}
$$

Now we can prove the full-rank of the unsymmetric reconstruction of a function $f \in$ $W_{2}^{1}(\Omega)$ by an approximation $s_{r} \in V_{R, Y_{r}}$.

Lemma 7.4.1 Under the condition (7.4.3) the system

$$
\begin{equation*}
\lambda_{j}\left(f-s_{r}\right)=0, \quad \text { for all } 1 \leq j \leq N_{s} \tag{7.4.4}
\end{equation*}
$$

is of full rank.
Proof: We have for $s_{r}=\sum_{\ell=1}^{N_{r}} a_{\ell} \Phi\left(\cdot-y_{\ell}\right)$ the equivalence

$$
\begin{align*}
\lambda_{j}\left(f-s_{r}\right) & =0 \text { for all } 1 \leq j \leq N_{s} \Leftrightarrow \\
\lambda_{j}(f) & =\sum_{\ell=1}^{N_{r}} a_{\ell} \lambda_{j}\left(\Phi\left(\cdot-y_{\ell}\right)\right) \text { for all } 1 \leq j \leq N_{s} \tag{7.4.5}
\end{align*}
$$

If we introduce the notation

$$
\begin{aligned}
A & :=\left(A_{\ell, j}\right)_{\substack{1 \leq \ell \leq N_{r} \\
1 \leq j \leq N_{s}}}:=\left(\int_{\Omega} R\left(x-y_{\ell}\right) S\left(x-x_{j}\right) d x\right)_{\substack{1 \leq \ell \leq N_{r} \\
1 \leq j \leq N_{s}}} \in \mathbb{R}^{N_{r} \times N_{s}} \\
F & :=\left(\begin{array}{c}
\lambda_{1}(f) \\
\vdots \\
\lambda_{N_{s}}(f)
\end{array}\right) \in \mathbb{R}^{N_{s}}
\end{aligned}
$$

equation (7.4.5) leads to the unsymmetric linear system

$$
\begin{equation*}
A a=F \tag{7.4.6}
\end{equation*}
$$

Now we can compute

$$
\begin{aligned}
\left\|u_{r}\right\|_{L_{2}(\Omega)} & \leq C\left(h_{s}\left\|u_{r}\right\|_{W_{2}^{1}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(u_{r}\right)\right|\right) \\
& \leq C\left(h_{s} \gamma\left(X_{r}\right)\left\|u_{r}\right\|_{L_{2}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(u_{r}\right)\right|\right) \\
& \leq \frac{1}{2}\left\|u_{r}\right\|_{L_{2}(\Omega)}+C \max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(u_{r}\right)\right|
\end{aligned}
$$

This yields

$$
\left\|u_{r}\right\|_{L_{2}(\Omega)} \leq 2 C \max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(u_{r}\right)\right|
$$

Therefore, $\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(u_{r}\right)\right|=0$ implies $\left\|u_{r}\right\|_{L_{2}(\Omega)}=0$. Since $u_{r} \in V_{R, Y_{r}}$ we have $u_{r} \equiv 0$. Thus rank of $A=\# Y_{r} \leq \# X_{s}$, which means that the system is of full rank.

Equation (7.4.2) implies that we can theoretically solve the system (7.4.4) to some prescribed accuracy $\delta(r, s)$. This is due to the fact $s_{f}$ is a good candidate for a solution, and by choosing $h_{s}$ small enough we can achieve every prescribed value $\delta(r, s)$. Practically we can apply some residual minimization techniques to find an approximate solution, such that the residual is smaller than $\delta(r, s)$. We denote this approximate solution by $u_{r, s} \in V_{R, X_{r}}$, i.e., we have

$$
\left|\lambda_{j}\left(f-u_{r, s}\right)\right| \leq \delta(r, s) \text { for all } 1 \leq j \leq N_{s}
$$

The exact error bound is given in the following theorem.
Theorem 7.4.2 We denote by $u_{r, s} \in V_{R, Y_{r}}$ the approximate solution of the system (7.4.4) to the accuracy $\delta(r, s)$. Then we have an error bound of the form

$$
\left\|f-u_{r, s}\right\|_{L_{2}(\Omega)} \leq C\left(2 h_{s}+\frac{1}{C_{1}} h_{r}\right)\|f\|_{W_{2}^{1}(\Omega)}+\delta(r, s) .
$$

If we choose $\delta(r, s)=\left(2 h_{s}+\frac{1}{C_{1}} h_{r}\right)\|f\|_{W_{2}^{1}(\Omega)}$, we get a final bound of the form

$$
\left\|f-u_{r, s}\right\|_{L_{2}(\Omega)} \leq C\left(2 h_{s}+\frac{1}{C_{1}} h_{r}\right)\|f\|_{W_{2}^{1}(\Omega)} .
$$

Proof: Let us denote by $u_{r}$ the unknown best trial approximation. Then we can compute

$$
\begin{aligned}
& \frac{1}{C}\left\|f-u_{r, s}\right\|_{L_{2}(\Omega)} \leq h_{s}\left\|f-u_{r, s}\right\|_{W_{2}^{1}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right| \\
\leq & h_{s}\|f\|_{W_{2}^{1}(\Omega)}+h_{s}\left\|u_{r, s}\right\|_{W_{2}^{1}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right| \\
\leq & h_{s}\|f\|_{W_{2}^{1}(\Omega)}+h_{s}\left\|u_{r}\right\|_{W_{2}^{1}(\Omega)}+ \\
& +h_{s}\left\|u_{r, s}-u_{r}\right\|_{W_{2}^{1}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right| \\
\leq & 2 h_{s}\|f\|_{W_{2}^{1}(\Omega)}+h_{s} \gamma\left(X_{r}\right)\left\|u_{r, s}-u_{r}\right\|_{L_{2}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right| \\
\leq & 2 h_{s}\|f\|_{W_{2}^{1}(\Omega)}+\frac{1}{2 C_{1}}\left\|u_{r, s}-u_{r}\right\|_{L_{2}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right| \\
\leq & 2 h_{s}\|f\|_{W_{2}^{1}(\Omega)}+\frac{1}{2 C_{1}}\left(\left\|u_{r, s}-f\right\|_{L_{2}(\Omega)}+\left\|u_{r}-f\right\|_{L_{2}(\Omega)}\right)+ \\
+ & \max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right| \\
\leq & 2 h_{s}\|f\|_{W_{2}^{1}(\Omega)}+\frac{1}{2 C_{1}}\left\|u_{r, s}-f\right\|_{L_{2}(\Omega)}+\frac{1}{2 C} h_{r}\|f\|_{W_{2}^{1}(\Omega)} \\
+ & \max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right| .
\end{aligned}
$$

This shows

$$
\left\|f-u_{r, s}\right\|_{L_{2}(\Omega)} \leq C\left(2 h_{s}+\frac{1}{2 C_{1}} h_{r}\right)\|f\|_{W_{2}^{1}(\Omega)}+\max _{1 \leq j \leq N_{s}}\left|\lambda_{j}\left(f-u_{r, s}\right)\right|
$$

giving the expected error bound

$$
\left\|f-u_{r, s}\right\|_{L_{2}(\Omega)} \leq C\left(2 h_{s}+\frac{1}{2 C_{1}} h_{r}\right)\|f\|_{W_{2}^{1}(\Omega)}+\delta(r, s)
$$

The second inequality follows from the special choice for $\delta(r, s)$.
This generalizes the results of [47] in two ways. Our error analysis does not assume $f$ to be known on a slightly larger domain $\tilde{\Omega}$. Furthermore, we obtain results for the $L_{2}$-norm, not for negative order Sobolev norms.

## Chapter 8

## Galerkin Methods

In this chapter we shall combine the well-known theory of variational problems with the idea of sampling inequalities. Galerkin methods are a powerful tool for solving elliptic partial differential equations. We shall extend the theory given in [62] and show a new application in the theory of weak recovery [47]. Unfortunately, the error estimates in [62] are given in the $W_{2}^{1}(\Omega)$-norm, and they require the solution $u$ to be in a native space of a radial basis function, which automatically forces $u$ to be continuous. In contrast, we shall give error estimates in the $L_{2}$-norm, hence we do not assume $u$ to be continuous.

### 8.1 Model Problem: Elliptic Partial Differential Equations

We shall concentrate on problems of the following form. Let $\Omega$ be a bounded domain with boundary $\partial \Omega \in C^{1}$. We shall consider partial differential equations of the form

$$
\begin{align*}
-\sum_{i, j=1}^{d} \frac{\partial}{\partial x_{i}}\left(a_{i, j}(x) \frac{\partial u}{\partial x_{j}(x)}\right)+c(x) u(x) & =f(x), \quad x \in \Omega \\
\left.\sum_{i, j=1}^{d} a_{i, j}\right|_{\partial \Omega}(x) \frac{\partial u(x)}{\partial x_{j}} \nu_{i}(x)+h(x) u(x) & =g(x), \quad x \in \partial \Omega, \tag{8.1.1}
\end{align*}
$$

where we assume $a_{i, j}, c \in L_{\infty}(\Omega), f \in L_{2}(\Omega), a_{i, j} \mid \partial \Omega, h \in L_{\infty}(\partial \Omega)$ and $g \in L_{2}(\partial \Omega)$ for $i, j=1, \ldots, d$. Further, we denote by $\nu$ the outward pointing unit normal vector to the boundary $\partial \Omega$.
The weak formulation of (8.1.1) is given in terms of the bilinear form

$$
\begin{equation*}
a_{\text {pde }}(u, v):=\int_{\Omega}\left(\sum_{i, j=1}^{d} a_{i, j} \frac{\partial u(x)}{\partial x_{j}} \frac{\partial v(x)}{\partial x_{i}}+c u v\right) d x+\int_{\partial \Omega} h u v d S \tag{8.1.2}
\end{equation*}
$$

and the linear form

$$
\begin{equation*}
F_{p d e}(v):=\int_{\Omega} f v d x+\int_{\partial \Omega} g v d S \tag{8.1.3}
\end{equation*}
$$

for a function $v \in W_{2}^{1}(\Omega)$. The matrix $A_{p d e}(x):=\left(a_{i, j}(x)\right)$ is assumed to be uniformly elliptic on $\Omega$, i.e., there is a constant $\gamma_{p d e}$ such that

$$
\gamma_{p d e}\|\alpha\|_{2}^{2} \leq \alpha^{T} A(x) \alpha
$$

for all $\alpha \in \mathbb{R}^{d}$ and all $x \in \Omega$. This gives rise to the variational problem

$$
\begin{equation*}
\text { find } u \in W_{2}^{1}(\Omega): a_{\text {pde }}(u, v)=F_{p d e}(v) \text { for all } v \in W_{2}^{1}(\Omega) . \tag{8.1.4}
\end{equation*}
$$

If further $c, h \geq 0$, and if at least one of them is uniformly bounded from zero on a subset of nonzero measure of $\Omega$ or $\partial \Omega$, we know that $a_{\text {pde }}(\cdot, \cdot): W_{2}^{1}(\Omega) \times W_{2}^{1}(\Omega) \rightarrow \mathbb{R}$ is strictly coercive, and that $F: W_{2}^{1}(\Omega) \rightarrow \mathbb{R}$ is continuous. We point out that the boundary conditions are part of the weak formulation. This unfortunately excludes Dirichlet boundary conditions at the moment.

### 8.2 A Norming Set Approach for Variational Problems

Since our analysis does not depend on the specific form of the variational problem in Section 8.1.1, we can concentrate on the essential features of (8.1.4). We consider the problem

$$
\begin{equation*}
\text { find } u \in W_{2}^{\tau}(\Omega): a(u, v)=f(v) \text { for all } v \in W_{2}^{\tau}(\Omega), \tag{8.2.1}
\end{equation*}
$$

for $a: W_{2}^{\tau}(\Omega) \times W_{2}^{\tau}(\Omega) \rightarrow \mathbb{R}$ being continuous, i.e., there is a constant $K_{a}>0$, s.t.

$$
\begin{equation*}
|a(u, v)| \leq K_{a}\|u\|_{W_{2}^{\tau}(\Omega)}\|v\|_{W_{2}^{\tau}(\Omega)} \quad \text { for all } u, v \in W_{2}^{\tau}(\Omega), \tag{8.2.2}
\end{equation*}
$$

and strictly coercive, i.e., there is a constant $\gamma>0$, s.t.

$$
\begin{equation*}
a(v, v) \geq \gamma\|v\|_{W_{2}^{\tau}(\Omega)}^{2} \quad \text { for all } v \in W_{2}^{\tau}(\Omega) \tag{8.2.3}
\end{equation*}
$$

and $f$ being a continuous linear form $f: W_{2}^{\tau}(\Omega) \rightarrow \mathbb{R}$. In our applications, we shall concentrate on the cases $\tau=1$ or $\tau>d / 2$. This models exactly the situation of [62]. The Lax-Milgram theorem [16] yields that (8.2.1) has a unique solution $u_{a, f} \in W_{2}^{\tau}(\Omega)$. Since we shall use techniques from the finite-element literature (see for instance [10]) we have to make some common assumptions. These are satisfied in the model-problem (8.1.1), but have to be stated for the general setting. We assume that the adjoined problem

$$
\text { find } \begin{align*}
w_{g} \in W_{2}^{\tau}(\Omega): a^{*}\left(w_{g}, \psi\right) & :=a\left(\psi, w_{g}\right) \\
& =(g, \psi)_{L_{2}(\Omega)} \tag{8.2.4}
\end{align*}
$$

has a unique solution for all $\psi \in W_{2}^{\tau}(\Omega)$ and $g \in L_{2}(\Omega)$. Since we want to get some convergence orders we have to impose that the solution $w_{g} \in W_{2}^{\sigma}(\Omega)$ lies in a continuously embedded subspace $W_{2}^{\sigma}(\Omega) \subset W_{2}^{\tau}(\Omega)$, i.e., $\sigma>\tau$ and satisfies the regularity assumption

$$
\begin{equation*}
\left\|w_{g}\right\|_{W_{2}^{\sigma}(\Omega)} \leq C\|g\|_{W_{2}^{\tau}(\Omega)} . \tag{8.2.5}
\end{equation*}
$$

Now we can apply a standard duality argument [10] to derive the estimate

$$
\left\|u-u_{n}\right\|_{L_{2}(\Omega)} \leq K^{a}\left\|u-u_{n}\right\|_{W_{2}^{\tau}(\Omega)} \sup _{g \in L_{2}(\Omega)} \frac{\inf _{\phi \in V}\left\|w_{g}-\phi\right\|_{W_{2}^{\tau}(\Omega)}}{\|g\|_{L_{2}(\Omega)}}
$$

where $u_{n}$ is the solution to the finite dimensional problem (8.2.6) given below. This technique using the adjoined problem is known as Aubin-Nitsche approach.
As a discretization we have to choose a sequence of $n$-dimensional subspaces $V_{n} \subset W_{2}^{\tau}(\Omega)$ with bases $\left\{\phi_{j}^{(n)}\right\}_{j=1, \ldots, n}$. The space $V_{n}$ can be the space $V_{\Phi, X}$ for instance. Under the assumptions (8.2.2) and (8.2.3) the finite dimensional variational problem

$$
\begin{equation*}
\text { find } u_{n} \in V \text { s.t. } a\left(u_{n}, v_{n}\right)=f\left(v_{n}\right) \quad \text { for all } v_{n} \in V_{n} \tag{8.2.6}
\end{equation*}
$$

is uniquely solvable for all $f \in\left(W_{2}^{\tau}(\Omega)\right)^{*}$. The Cea Lemma says that the solution of (8.2.6) is nearly optimal.

Lemma 8.2.1 Let $u_{n}$ be a solution to (8.2.6) and $u$ a solution to (8.2.1), then we have

$$
\left\|u-u_{n}\right\|_{W_{2}^{\tau}(\Omega)} \leq C \inf _{v_{n} \in V_{n}}\left\|u-v_{n}\right\|_{W_{2}^{\tau}(\Omega)}
$$

Proof: See [30, Lemma 1.26].
Since we shall use weak data of the form $a\left(\cdot, \phi_{j}^{(n)}\right)$ we have to ensure that this data contains enough information. We shall even show that this data induces a norm on $V_{n}$. First we define a projection map by

$$
\begin{align*}
P: \quad \mathbb{R}^{n} & \rightarrow V_{n} \\
u & =\left(u_{1}, \ldots, u_{n}\right)^{T} \mapsto u^{n}:=\sum_{j=1}^{n} u_{j} \phi_{j}^{(n)} . \tag{8.2.7}
\end{align*}
$$

This is obviously an isomorphism between $\mathbb{R}^{n}$ and $V$. We have to assume from now on that the norm of its inverse is bounded, i.e., $\left\|P^{-1}\right\|=: \Gamma^{V_{n}, \phi}<\infty$. The constant $\Gamma^{V_{n}, \phi}$ is in some important cases explicitly known. We consider two equivalent norms on $V_{n}$, namely the usual $\|\cdot\|_{W_{2}^{\tau}}$-norm and the norm $\|\cdot\|_{a}:=\sqrt{a(\cdot, \cdot)}$ induced by the bilinear form $a(\cdot, \cdot)$. The equivalence constants are given by

$$
\begin{equation*}
\gamma\|u\|_{W_{2}^{\tau}(\Omega)}^{2} \leq\|u\|_{a}^{2}:=a(u, u) \leq K\|u\|_{W_{2}^{\tau}(\Omega)}^{2} \quad \text { for all } u \in V_{n}, \tag{8.2.8}
\end{equation*}
$$

where $K$ and $\gamma$ are given in equations (8.2.2) and (8.2.3).
Lemma 8.2.2 Let $\left\{\phi_{j}^{(n), 1}\right\}_{j=1, \ldots, n}$ be an orthonormal base of $V_{n}$ with respect to the $a(\cdot, \cdot)$ inner product, and let $\left\{\phi_{j}^{(n), 2}\right\}_{j=1, \ldots, n}$ be an orthonormal base of $V_{n}$ with respect to the usual $(\cdot, \cdot)_{W_{2}^{\tau}(\Omega)}$ inner product. We define

$$
\begin{align*}
P_{i}:\left(\mathbb{R}^{n},\|\cdot\|_{2}\right) & \rightarrow\left(V_{n},\|\cdot\|_{W_{2}^{\tau}(\Omega)}\right) \\
u=\left(u_{1}, \ldots, u_{n}\right)^{T} & \mapsto \quad u^{(n)}:=\sum_{j=1}^{n} u_{j} \phi_{j}^{(n), i} \quad \text { ffor } i=1,2 . \tag{8.2.9}
\end{align*}
$$

Then $\sqrt{\gamma} \leq\left\|P_{1}^{-1}\right\| \leq \sqrt{K}$ and $\left\|P_{2}^{-1}\right\|=1$.

Proof: We can calculate as follows. Let $u_{i}^{(n)}:=\sum_{j=1}^{n} u_{j} \phi_{j}^{(n), i}$.

$$
\left\|P_{2}^{-1}\right\|:=\sup _{u_{2}^{n} \in V_{n} \backslash\{0\}} \frac{\left\|P_{2}^{-1}\left(u_{2}^{n}\right)\right\|_{2}}{\left\|u_{2}^{n}\right\|_{W_{2}^{\tau}(\Omega)}}=\frac{\|u\|_{2}}{\left\|u_{2}^{n}\right\|_{W_{2}^{\tau}(\Omega)}}
$$

Since the $\phi$ 's are orthonormal we can conclude

$$
\begin{aligned}
\left\|u_{2}^{n}\right\|_{W_{2}^{\tau}(\Omega)}^{2} & =\left(\sum_{j=1}^{n} u_{j} \phi_{j}^{(n), 2}, \sum_{j=1}^{n} u_{j} \phi_{j}^{(n), 2}\right)_{W_{2}^{\tau}(\Omega)} \\
& =\sum_{j, k=1}^{n} u_{j} u_{k}\left(\phi_{j}^{(n), 2}, \phi_{j}^{(n), 2}\right)_{W_{2}^{\tau}(\Omega)}=\|u\|_{2}^{2}
\end{aligned}
$$

Hence $\left\|P_{2}^{-1}\right\|=1$. In the first case we can analogously derive

$$
\left\|P_{1}^{-1}\right\|:=\sup _{u_{1}^{(n)} \in V_{n} \backslash\{0\}} \frac{\left\|P_{1}^{-1}\left(u_{2}^{(n)}\right)\right\|_{2}}{\left\|u_{2}^{(n)}\right\|_{W_{2}^{\tau}(\Omega)}}=\frac{\|u\|_{2}}{\left\|u_{1}^{n}\right\|_{W_{2}^{\tau}(\Omega)}}
$$

Now we use the equivalence (8.2.8) to deduce

$$
\gamma\|u\|_{2}^{2}=\gamma\left\|u_{1}^{(n)}\right\|_{a}^{2} \leq\left\|u_{1}^{(n)}\right\|_{W_{2}^{\tau}(\Omega)}^{2} \leq K\left\|u_{1}^{(n)}\right\|_{a}^{2}=K\|u\|_{2}^{2}
$$

Hence $\sqrt{\gamma} \leq\left\|P_{1}^{-1}\right\| \leq \sqrt{K}$.
We denote the dual space of $V_{n}$ by $V_{n}^{*}$. The basis $\left\{\phi_{j}^{(n)}\right\}_{j=1, \ldots, n}$ of $V_{n}$ defines via the Riesz-map a finite subset of $V_{n}^{*}$,

$$
Z:=\left\{a\left(\cdot, \phi_{1}\right), \ldots, a\left(\cdot, \phi_{n}\right)\right\} \subset V^{*}
$$

This set $Z$ gives rise to a sampling operator

$$
\begin{aligned}
T: V_{n} & \rightarrow T\left(V_{n}\right) \subset \mathbb{R}^{n} \\
u^{(n)} & \mapsto\left(z\left(u^{(n)}\right)\right)_{z \in Z}
\end{aligned}
$$

taking the form

$$
T\left(u^{(n)}\right)=\left(\begin{array}{c}
a\left(u^{(n)}, \phi_{1}\right)  \tag{8.2.10}\\
\vdots \\
a\left(u^{(n)}, \phi_{n}\right)
\end{array}\right):=f \in \mathbb{R}^{n}
$$

We collect some important properties of the operator $T$.
Lemma 8.2.3 $T$ is injective and the norm of its inverse on the range of $T$ is bounded by $\Gamma^{V, \phi} / \gamma$, where $\Gamma^{V, \phi}$ is a constant depending both, on the space $V$ and on the arbitrary basis $\{\phi\}$ of $V$.

Proof: $T$ is injective by the Lax-Milgram lemma, since we know that (8.2.6) is uniquely solvable. Therefore $T: V_{n} \rightarrow T\left(V_{n}\right)$ is bijective, and we can compute $\left\|T^{-1}\right\|$. To do so, we define $u:=P^{-1} u^{n}$ and find

$$
\begin{aligned}
\gamma\|u\|_{2} \cdot\left\|u^{(n)}\right\|_{W_{2}^{\tau}(\Omega)} & =\gamma \cdot\left\|P^{-1} u^{(n)}\right\|_{2} \cdot\left\|u^{(n)}\right\|_{W_{2}^{\tau}(\Omega)} \leq \gamma\left\|P^{-1}\right\| \cdot\left\|u^{(n)}\right\|_{W_{2}^{\tau}(\Omega)}^{2} \\
& \leq \gamma \cdot \Gamma^{V, \phi} \cdot\left\|u^{(n)}\right\|_{W_{2}^{\tau}(\Omega)}^{2} \leq \Gamma^{V, \phi} \cdot a\left(u^{(n)}, u^{(n)}\right) \\
& =\Gamma^{V, \phi} \cdot \sum_{i=1}^{n} u_{i} a\left(u^{(n)}, \phi_{i}\right)=\Gamma^{V, \phi} \cdot \sum_{i=1}^{n} u_{i} f_{i} \\
& \leq \Gamma^{V, \phi} \cdot\|u\|_{2} \cdot\|f\|_{2} .
\end{aligned}
$$

This implies

$$
\left\|T^{-1}\right\|:=\sup _{\mathbb{R}^{\ni} f \neq 0} \frac{\left\|T^{-1} f\right\|_{W_{2}^{\tau}(\Omega)}}{\|f\|_{\mathbb{R}^{n}}}=\frac{\left\|u^{n}\right\|_{W_{2}^{\tau}(\Omega)}}{\|f\|_{\mathbb{R}^{n}}} \leq \frac{\Gamma^{V, \phi}}{\gamma} .
$$

Please note that we have chosen the $\|\cdot\|_{2}$-norm as the norm on $\mathbb{R}^{n}$.
The constant $\Gamma^{V, \phi}$ is known in the two cases of Lemma 8.2.2.
Remark 8.2.4 We get $\sqrt{\gamma} \leq \Gamma^{V, \phi_{1}}=\left\|P_{1}^{-1}\right\| \leq \sqrt{K}$ and $\Gamma^{V, \phi_{2}}=1$.
Now we can invoke the usual general result from the norming set theory (see Theorem 3.4.2) to get the following result.

Theorem 8.2.5 For every $\Psi \in V^{*}$ there exists a vector $b=b(\Psi) \in \mathbb{R}^{n}$ depending only on $\Psi$, such that for every $v \in V$

$$
\begin{aligned}
\Psi(v) & =\sum_{j=1}^{n} b_{j}(\Psi) z_{j}=\sum_{j=1}^{n} b_{j}(\Psi) a\left(v, \phi_{j}\right) \quad \text { with } \\
\|b(\Psi)\|_{\left(\mathbb{R}^{n}\right)^{*}} & \leq \frac{\Gamma^{V, \phi}}{\gamma}\|\Psi\|_{V^{*}} .
\end{aligned}
$$

In the following sections we shall always assume that $V$ can be continuously embedded into the space of continuous functions, $V \subset C(\Omega)$. Therefore point evaluations are well defined on $V$, and we may consider $\Psi=\delta_{y} \in V^{*}$ for any $y \in \Omega$.

### 8.3 Sampling Inequality for Galerkin Data

From now on we need some technical assumptions. We shall assume a radial basis function $\Phi$ that is regular enough to satisfy

$$
\hat{\Phi}(\omega) \sim\left(1+\|\omega\|_{2}^{2}\right)^{-\xi} \quad \text { for a } \xi>d / 2,
$$

i.e., $\Phi$ is the reproducing kernel in a Hilbert space that is norm-equivalent to the classical Sobolev space $W_{2}^{\xi}(\Omega)$. As usual, the approximation quality of the space

$$
V_{\Phi, X}:=\operatorname{span}\left\{\Phi\left(\cdot-x_{j}\right): x_{j} \in X\right\}
$$

depends on the choice of the discrete set $X$. It is well-known [65] that the error estimates will depend on the fill distance

$$
h_{X, \Omega}:=h:=\sup _{x \in \Omega} \min _{x_{j} \in X}\left\|x-x_{j}\right\|_{2} .
$$

However, we shall as well need a stability assumption of the form

$$
\begin{equation*}
\|s\|_{W_{2}^{\xi}(\Omega)} \leq C\left(V_{\Phi, X}, \xi\right)\|s\|_{L_{2}(\Omega)} \quad \text { for all } s \in V_{\Phi, X} \tag{8.3.1}
\end{equation*}
$$

The constant $C\left(V_{\Phi, X}, \xi\right)$ exists because of the norm-equivalence on finite dimensional spaces. As shown in Chapter 6, there is strong evidence that the equivalence constant will depend mainly on the separation distance

$$
q_{X}=\min _{\substack{x_{i}, x_{j} \in X \\ x_{i} \neq x_{j}}} \frac{1}{2}\left\|x_{i}-x_{j}\right\|_{2} .
$$

Theorem 8.2 .5 indicates that we have to bound the norm of the point evaluation functional. This is done by Sobolev's inequality [16]

$$
\left|\delta_{x}(s)\right|=|s(x)| \leq\|s\|_{W_{2}^{\beta}(\Omega)}
$$

for some $\beta>d / 2$. Now we can use an inverse estimate to establish

$$
\|s\|_{W_{2}^{\beta}(\Omega)} \leq C\left(V_{\Phi, X}, \beta, \xi\right)\|s\|_{W_{2}^{\xi}(\Omega)}
$$

where $C\left(V_{\Phi, X}, \beta, \xi\right)$ might be obtained from $C\left(V_{\Phi, X}, 0, \xi\right)$ by an operator-interpolation argument (see, e.g., [11]). These two inequalities give

$$
\begin{equation*}
\left|\delta_{x}(s)\right| \leq C\left(V_{\Phi, X}, \beta, \xi\right)\|s\|_{W_{2}^{\xi}(\Omega)} \tag{8.3.2}
\end{equation*}
$$

As we already mentioned, $C\left(V_{\Phi, X}, \beta, \xi\right)$ depends mainly on the separation distance $q:=$ $q_{X}$, so we shall simply refer to it as $C(q)$. We point out that this finally yields a bound of the form

$$
\begin{align*}
v(x) & =\sum_{j=1}^{n} b_{j}(x) z_{j} \quad \text { with }  \tag{8.3.3}\\
\|b(v)\|_{\left(\mathbb{R}^{n}\right)^{*}} & \leq \frac{\Gamma^{V, \phi}}{\gamma} C(q) \tag{8.3.4}
\end{align*}
$$

In the following we use the notation

$$
\|a(u, \phi)\|_{\ell_{2}(n)}:=\left(\sum_{j=1}^{n} a\left(u, \phi_{j}\right)^{2}\right)^{1 / 2}
$$

for an arbitrary vector $\phi=\left(\phi_{1}, \ldots, \phi_{n}\right)^{T}$.

Theorem 8.3.1 Assume that the solution $w_{g}$ of the adjoined problem (8.2.4) fulfills $w_{g} \in$ $W_{2}^{\sigma}(\Omega)$, where we have to impose $\sigma=\xi$. Let $\left\{\phi_{j}\right\}_{j=1, \ldots, n}$ be a basis of $V_{\Phi, X}$. Then there exists a constant $C>0$ with the following property. There is a constant $h_{o}>0$ such that for all sets $X \subset \Omega$ with fill distance $h \leq h_{0}$ and all functions $u \in W_{2}^{\tau}(\Omega)$ the inequality

$$
\|u\|_{L_{2}(\Omega)} \leq C\left(h^{\sigma-\tau}\|u\|_{W_{2}^{\tau}(\Omega)}+C(q) \Gamma^{V, \phi} \gamma^{-1}\|a(u, \phi)\|_{\ell_{2}(n)}\right)
$$

holds.
Proof: We find for an arbitrary $p \in V_{\Phi, X}$ and any $u \in W_{2}^{\tau}(\Omega)$

$$
\begin{aligned}
\|u\|_{L_{2}(\Omega)} & \leq\|u-p\|_{L_{2}(\mathcal{D})}+\|p\|_{L_{2}(\Omega)} \\
& \leq\|u-p\|_{L_{2}(\Omega)}+\operatorname{vol}(\Omega)^{1 / 2}\left|\delta_{\tilde{x}} p\right|
\end{aligned}
$$

where $\tilde{x} \in \Omega$ is the extremum of $p$, i.e., $|p(\tilde{x})|=\max _{x \in \Omega}|p(x)|$. This point exists, since $\Omega$ is compact. Now we can use the reproduction formula from Theorem 8.2 .5 to get

$$
\begin{aligned}
\|u\|_{L_{2}(\Omega)} & \leq\|u-p\|_{L_{2}(\Omega)}+C\left|\sum_{j=1}^{n} b_{j}(\tilde{x}) a\left(p, \phi_{j}\right)\right| \\
& \leq\|u-p\|_{L_{2}(\Omega)}+C\left|\sum_{j=1}^{n} b_{j}(\tilde{x})\left(a\left(u-p, \phi_{j}\right)+a\left(u, \phi_{j}\right)\right)\right| \\
& \leq\|u-p\|_{L_{2}(\Omega)}+C\|b(\tilde{x})\|_{2}\|a(u-p, \phi)\|_{\ell_{2}(n)} \\
& +C\|b(\tilde{x})\|_{2} \cdot\|a(u, \phi)\|_{\ell_{2}(n)} .
\end{aligned}
$$

Since $a(\cdot, \cdot)$ is coercive we can choose $\tilde{p} \in V_{\Phi, X}$, such that

$$
a\left(u, \phi_{j}\right)=a\left(\tilde{p}, \phi_{j}\right) \quad \text { for all } 1 \leq j \leq n .
$$

Since the $\phi_{j}$ 's form a basis for $V_{\Phi, X}$, the estimate reduces to

$$
\|u\|_{L_{2}(\Omega)} \leq\|u-\tilde{p}\|_{L_{2}(\Omega)}+\Gamma^{V, \phi} \gamma^{-1} C(q)\|a(u, \phi)\|_{\ell_{2}(n)},
$$

where we used the bound for the point-evaluation functional from (8.3.2). We can use the Aubin-Nitsche approach to get

$$
\|u-\tilde{p}\|_{L_{2}(\Omega)} \leq K^{a}\|u-\tilde{p}\|_{W_{2}^{\tau}(\Omega)} \sup _{g \in L_{2}(\Omega)} \frac{\inf _{\phi \in V_{\Phi, X}}\left\|w_{g}-\phi\right\|_{W_{2}^{\tau}(\Omega)}}{\|g\|_{L_{2}(\Omega)}} .
$$

Since our stability assumption yields $w_{g} \in W_{2}^{\sigma}(\Omega) \subset C(\Omega)$ we can apply the well-known estimates for interpolation in the native space [65] to get

$$
\inf _{\phi \in V_{\Phi, X}}\left\|w_{g}-\phi\right\|_{W_{2}^{\tau}(\Omega)} \leq C h^{\sigma-\tau}\left\|w_{g}\right\|_{W_{2}^{\sigma}(\Omega)} \leq C h^{\sigma-\tau}\|g\|_{L_{2}(\Omega)} .
$$

The Cea Lemma 8.2.1 yields again

$$
\left\|u-p^{*}\right\|_{W_{2}^{\tau}(\Omega)} \leq C\|u\|_{W_{2}^{\tau}(\Omega)} .
$$

Putting things together we find the claim.
We can use Remark 8.2.4 to present an explicit bound on the constant in front of the discrete term.

Corollary 8.3.2 Under the conditions from Theorem 8.3.1 and the additional requirement that $\left\{\phi_{j}^{(n), 1}\right\}_{j=1}^{n}$ is an orthonormal base of $V_{\Phi, X}$ with respect to the $a(\cdot, \cdot)$ inner product, we get

$$
\begin{equation*}
\|u\|_{L_{2}(\Omega)} \leq C\left(h^{\sigma-\tau}\|u\|_{W_{2}^{\tau}(\Omega)}+C(q) \sqrt{K} \gamma^{-1}\|a(u, \phi)\|_{\ell_{2}(n)}\right) . \tag{8.3.5}
\end{equation*}
$$

If we additionally require that $\left\{\phi_{j}^{(n), 2}\right\}_{j=1}^{n}$ is an orthonormal base of $V_{n}$ with respect to the $(\cdot, \cdot)_{W_{2}^{\tau}(\Omega)}$ inner product, then we get

$$
\begin{equation*}
\|u\|_{L_{2}(\Omega)} \leq C\left(h^{\sigma-\tau}\|u\|_{W_{2}^{\tau}(\Omega)}+C(q) \gamma^{-1}\|a(u, \phi)\|_{\ell_{2}(n)}\right) . \tag{8.3.6}
\end{equation*}
$$

The discrete term seems to be rather unintuitive. But it has an interpretation for orthonormal bases.
Remark 8.3.3 We assume $\xi>\max \tau, d / 2$. If $\left\{\phi_{j}^{(n), 1}\right\}_{j=1}^{n}$ is an orthonormal base of $V_{\Phi, X}$ with respect to the $a(\cdot, \cdot)$ inner product, we get that

$$
\|a(u, \phi)\|_{\ell_{2}(n)} \longrightarrow\|u\|_{a} \quad \text { for } h_{X, \Omega} \rightarrow 0,
$$

where $\|\cdot\|_{a}$ is equivalent to the energy norm.
Proof: Since $\Phi$ is the reproducing kernel of the Sobolev space $W_{2}^{\xi}(\Omega)$ with $\xi>\tau$ we get by appropriate sampling inequalities that $V_{\Phi, X}$ becomes dense in $W_{2}^{\tau}(\Omega)$ for $h_{X, \Omega} \rightarrow 0$. Since the $\left\{\phi_{j}^{(n), 1}\right\}_{j=1}^{n}$ are orthonormal, the claim follows by the Parseval equation.
Let us clarify the main result of Theorem 8.3.1 a bit further. We consider again the weak formulation of a Laplace equation with Neumann boundary conditions, i.e.,

$$
a(u, v)=\int_{\Omega} \nabla u(x) \nabla v(x) d x \quad \text { for } u, v \in W_{2}^{1}(\Omega)
$$

Then the sampling operator consists of smoothed derivative data

$$
\lambda_{j}(f)=\int_{\Omega} \nabla f(x) \nabla \phi_{j}(x) d x
$$

Another approach is to apply the sampling inequality directly to the partial differential equation (8.1.1). If $\Omega \subset \mathbb{R}^{3}$ satisfies the conditions from above, and if $f$ satisfies the regularity assumptions, we know that the solution $u \in W_{2}^{2}(\Omega)$ satisfies the regularity assumption with $\sigma=\xi=2>3 / 2$. We construct an approximate solution $u_{X} \in V_{\Phi, X}$ by solving

$$
a\left(u, \phi_{j}\right)=a\left(u_{X}, \phi_{j}\right) \quad \text { for all } 1 \leq j \leq n .
$$

Now we can apply our result to $u-u_{X} \in W_{2}^{2}(\Omega)$. This yields

$$
\left\|u-u_{X}\right\|_{L_{2}(\Omega)} \leq C h^{1}\|u\|_{W_{2}^{1}(\Omega)} .
$$

To get error estimates in higher norms we use results from [62].

### 8.4 Error Estimates for Best Approximation

From now on we shall assume that $\Omega \subset \mathbb{R}^{d}$ with $d \leq 3$ is bounded with a Lipschitz boundary and satisfies an interior cone condition with radius $r$ and angle $\theta$. We shall give an application of the sampling inequality derived in the previous section. We shall consider the problem to find the best approximation $s_{f} \in V_{\Phi, X}=\operatorname{span}\left\{\Phi\left(\cdot-x_{j}\right) \mid x_{j} \in X\right\}$ to a given function $f \in W_{2}^{1}(\Omega)$. We restrict ourselves to the case $\xi=2$. The best approximation in a Hilbert space is characterized by the orthogonality equations

$$
\left\langle f-s_{f}, s\right\rangle_{W_{2}^{1}(\Omega)}=0 \quad \text { for all } s \in V_{\Phi, X} .
$$

To apply the results from the previous section we define the bilinear form

$$
\begin{aligned}
a: W_{2}^{1}(\Omega) \times W_{2}^{1}(\Omega) & \rightarrow \mathbb{R} \\
a(u, v) & :=\langle u, v\rangle_{W_{2}^{1}(\Omega)} .
\end{aligned}
$$

Obviously, $a(\cdot, \cdot)$ is symmetric, continuous and $W_{2}^{1}(\Omega)$-coercive. Further, the orthogonality equation can be interpreted as a variational problem,

$$
\text { find } s_{f} \text { s.t. } a\left(s_{f}, s\right)=a(f, s)=: f(s) \text { for all } s \in V_{\Phi, X}
$$

Now we are able to apply the results from the previous section to this special bilinear form.
Proposition 8.4.1 Let $\Omega \subset \mathbb{R}^{d}$ with $d \leq 3$ be a bounded domain satisfying an interior cone condition. Under the conditions from Theorem 8.3.1 we find for the best approximation $s_{f}$ from $V_{\Phi, \Omega, X}$ to a function $f \in W_{2}^{1}(\Omega)$ the following error estimate

$$
\left\|f-s_{f}\right\|_{L_{2}(\Omega)} \leq h_{X, \Omega}\|f\|_{W_{2}^{1}(\Omega)} .
$$

Proof: To do our error analysis we have to check whether the problem

$$
\begin{aligned}
& \text { find } w_{g} \text { s.t. } a^{*}\left(w_{g}, \Psi\right)=a\left(w_{g}, \Psi\right)=\left\langle w_{g}, \Psi\right\rangle_{W_{2}^{1}(\Omega)} \\
& =\langle g, \Psi\rangle_{L_{2}(\Omega)} \quad \text { for all } \Psi \in W_{2}^{1}(\Omega)
\end{aligned}
$$

possesses a solution $w_{g} \in W_{2}^{1}(\Omega)$ for all $g \in L_{2}(\Omega)$. The existence of a solution is a direct consequence of the Lax-Milgram theorem [10], since we deal with a symmetric coercive bilinear form. Now we have to check the higher regularity, which we need for our estimates. To see this, we compute, provided $u, v$ are smooth enough, as follows.

$$
\begin{aligned}
\langle u, v\rangle_{W_{2}^{1}(\Omega)} & =\langle u, v\rangle_{L_{2}(\Omega)}+\langle\nabla u, \nabla v\rangle_{L_{2}(\Omega)} \\
& =\langle u, v\rangle_{L_{2}(\Omega)}-\langle u, \Delta v\rangle_{L_{2}(\Omega)}+\int_{\partial \Omega} \nabla v u \nu d S,
\end{aligned}
$$

where $\nu$ denotes the outward pointing normal vector. This calculation shows that $w_{g}$ would solve the Helmholtz equation. We are free to choose a boundary condition. In this case the natural boundary conditions [24]

$$
\begin{aligned}
-\Delta w_{g}+w_{g} & =g \text { in } \Omega \\
\nabla w_{g} \nu & =0 \text { on } \partial \Omega
\end{aligned}
$$

are useful. Standard regularity estimates for elliptic partial differential equations [24] say that a solution $w_{g}$ to the variational problem lies in $W_{2}^{2}(\Omega)$, and that it depends continuously on the data, i.e., $\left\|w_{g}\right\|_{W_{2}^{2}(\Omega)} \leq C\|g\|_{L_{2}(\Omega)}$. Now we can apply the results from the previous section to get the estimate

$$
\|u\|_{L_{2}(\Omega)} \leq h_{X, \Omega}\|u\|_{W_{2}^{1}(\Omega)}+C(\Psi) \frac{C(\Omega)}{\gamma}\|a(u, \phi)\|_{2(\Omega)}
$$

If we insert the residual $u=f-s_{f}$, we find

$$
\left\|f-s_{f}\right\|_{L_{2}(\Omega)} \leq h_{X, \Omega}\|f\|_{W_{2}^{1}(\Omega)}
$$

This shows that one can use the techniques from the finite element literature also in the context of meshless kernel methods. Furthermore, the sampling inequality provides also error bounds, even if we do not solve the linear system with the stiffness matrix exactly.

### 8.5 Sampling Inequality with Polynomials and Galerkin Data

This section is completely independent from the previous ones. In this section, we shall choose $V$ to be a polynomial space, to be more precise, $V=\pi_{k}(\Omega)$ for some $k \in \mathbb{N}$. This means that we do a kind of generalized finite element theory with $\pi_{k}(\Omega)$. To derive an upper bound for $\left\|\delta_{y}\right\|_{V^{*}}$ we first have to work locally on nice "small " domains. The first step in deriving upper bounds is to consider only local regions $\mathcal{D}$ that are star-shaped with respect to a ball. A bounded domain $\Omega$ bounded region that is star-shaped with respect to a ball satisfies a uniform interior cone condition.
In the case that $V$ consists of polynomials, a bound on $\left\|\delta_{y}\right\|$ can be found in [64]. If $V=\pi_{k}\left(\mathbb{R}^{d}\right)$ is the space of all algebraic polynomials of degree not exceeding $k$, we know [38]

$$
\left\|D^{\alpha} p\right\|_{L_{\infty}(\mathcal{D})} \leq\left(\frac{2 k^{2}}{r \sin (\theta)}\right)^{|\alpha|}\|p\|_{L_{\infty}(\mathcal{D})}
$$

for arbitrary $p \in V$. Unfortunately, this estimate does not use the $W_{2}^{\tau}$-norm. Therefore we have to modify this result. To do so, we use a result from [11, Lemma 4.5.3], i.e., for $1 \leq p, q \leq \infty$ and $0 \leq m \leq \ell$ we have

$$
\|v\|_{W_{p}^{\ell}(\mathcal{D})} \leq C \rho_{\mathcal{D}}^{m-\ell+d / p-d / q}\|v\|_{W_{q}^{m}(\mathcal{D})} \text { for all } v \in \pi_{k}\left(\mathbb{R}^{d}\right)
$$

where $\rho_{\mathcal{D}}=\operatorname{diam}(\mathcal{D})<1$ denotes the diameter of the domain $\mathcal{D}$.
Lemma 8.5.1 If $\mathcal{D}$ is star-shaped with respect to a ball with radius $r$, for every $\alpha \in \mathbb{N}_{0}^{d}$ and every real number $\beta>d / 2$ and $\tau \leq \beta$,

$$
\left|\delta_{y} D^{\alpha} p\right| \leq C\left(\frac{2 k^{2}}{r \sin (\theta)}\right)^{|\alpha|} \rho_{\mathcal{D}}^{\tau-\beta}\|p\|_{W_{2}^{\tau}(\mathcal{D})}
$$

for all $p \in V=\pi_{k}(\mathcal{D})$.

Proof: Putting the inequalities together, yields for $\beta>d / 2$ and any $y \in \mathcal{D}$

$$
\begin{aligned}
\left|\delta_{y} D^{\alpha} p\right| & \leq\left\|D^{\alpha} p\right\|_{L_{\infty}(\mathcal{D})} \leq\left(\frac{2 k^{2}}{r \sin (\theta)}\right)^{|\alpha|}\|p\|_{L_{\infty}(\mathcal{D})} \\
& \leq\left(\frac{2 k^{2}}{r \sin (\theta)}\right)^{|\alpha|}\|p\|_{W_{2}^{\beta}(\mathcal{D})} \\
& \leq C\left(\frac{2 k^{2}}{r \sin (\theta)}\right)^{|\alpha|} \rho_{\mathcal{D}}^{\tau-\beta}\|p\|_{W_{2}^{\tau}(\mathcal{D})}
\end{aligned}
$$

where we used Sobolev's inequality in the fourth step.
Since we are going to do an $L_{2}$-theory, we shall assume $\alpha=0$.
We assume that the domain $\mathcal{D}$ is compact, i.e., bounded and closed, so that for any polynomial $p \in \pi_{k}\left(\mathbb{R}^{d}\right)$ there is a point $\tilde{x} \in \mathcal{D}$ such that

$$
\max _{x \in \mathcal{D}}|p(x)|=|p(\tilde{x})|=\left|\delta_{\tilde{x} \tilde{p}}\right| .
$$

Proposition 8.5.2 Assume that the solution $w_{g}$ of the adjoined problem (8.2.4) fulfills $w_{g} \in$ $W_{2}^{\sigma}(\Omega)$, where we have to impose $\sigma=\xi$. Let $\left\{\phi_{j}^{\mathcal{D}}\right\}_{j=1, \ldots, n}$ be an orthonormal basis of $\pi_{\tau}(\mathcal{D})$ with respect to the $W_{2}^{\tau}$-inner product. Then there is a constant $C>0$ such that for all $u \in W_{2}^{\tau}(\mathcal{D})$ we have

$$
\|u\|_{L_{2}(\mathcal{D})} \leq C\left(\rho_{\mathcal{D}}^{\sigma-\tau}\|u\|_{W_{2}^{\tau}(\mathcal{D})}+\gamma^{-1} \rho_{\mathcal{D}}^{\tau-\beta+d / 2}\|a(u, \phi)\|_{\ell_{2}(n)}\right)
$$

with $\sigma>\tau$.
Proof: We find for an arbitrary polynomial $p$ and any $u \in W_{2}^{\tau}(\mathcal{D})$

$$
\begin{align*}
\|u\|_{L_{2}(\mathcal{D})} & \leq\|u-p\|_{L_{2}(\mathcal{D})}+\|p\|_{L_{2}(\mathcal{D})} \\
& \leq\|u-p\|_{L_{2}(\mathcal{D})}+\operatorname{vol}(\mathcal{D})^{1 / 2}\left|\delta_{\tilde{x}} p\right| \\
& \leq\|u-p\|_{L_{2}(\mathcal{D})}+\rho_{\mathcal{D}}^{d / 2}\left|\sum_{j=1}^{n} b_{j}(\tilde{x}) a\left(p, \phi_{j}\right)\right| \\
& \leq\|u-p\|_{L_{2}(\mathcal{D})}+\rho_{\mathcal{D}}^{d / 2}\left|\sum_{j=1}^{n} b_{j}(\tilde{x})\left(a\left(u-p, \phi_{j}\right)+a\left(u, \phi_{j}\right)\right)\right| \\
& \leq\|u-p\|_{L_{2}(\mathcal{D})}+\rho_{\mathcal{D}}^{d / 2}\|b(\tilde{x})\|_{2}\|a(u-p, \phi)\|_{\ell_{2}(n)} \\
& +\rho_{\mathcal{D}}^{d / 2}\|b(\tilde{x})\|_{2}\|a(u, \phi)\|_{\ell_{2}(n)} . \tag{8.5.1}
\end{align*}
$$

Since $a(\cdot, \cdot)$ is coercive we can choose a polynomial $\tilde{p}_{\mathcal{D}} \in V=\pi_{\tau}(\mathcal{D})$ such that

$$
a\left(u, \phi_{j}\right)=a\left(\tilde{p}_{\mathcal{D}}, \phi_{j}\right) \quad \text { for all } 1 \leq j \leq n .
$$

We point out that the polynomial $\tilde{p}_{\mathcal{D}}$ depends on the domain $\mathcal{D}$. Furthermore, we can use Theorem 8.2.5 and Lemma 8.5.1 with $|\alpha|=0$ to note that

$$
\|b(\tilde{x})\|_{2} \leq \gamma^{-1}\left\|\delta_{\tilde{x}}\right\|_{V^{*}} \leq C \gamma^{-1} \rho_{\mathcal{D}}^{\tau-\beta}
$$

If the $\phi_{j}$ 's form an orthonormal basis for $\pi_{\tau}(\mathcal{D})$, the estimate (8.5.1) reduces to

$$
\|u\|_{L_{2}(\mathcal{D})} \leq\left\|u-p_{\mathcal{D}}\right\|_{L_{2}(\mathcal{D})}+C \gamma^{-1} \rho_{\mathcal{D}}^{\tau-\beta+d / 2}\|a(u, \phi)\|_{\ell_{2}(n)} .
$$

To bound the term $\left\|u-\tilde{p}_{\mathcal{D}}\right\|_{L_{2}(\mathcal{D})}$ we will use the Aubin-Nitsche construction from the introduction, i.e.,

$$
\left\|u-\tilde{p}_{\mathcal{D}}\right\|_{L_{2}(\mathcal{D})} \leq K^{a}\left\|u-\tilde{p}_{\mathcal{D}}\right\|_{W_{2}^{\tau}(\Omega)} \sup _{g \in L_{2}(\mathcal{D})} \frac{\inf _{\phi \in \pi_{\tau}(\mathcal{D})}\left\|w_{g}-\phi\right\|_{W_{2}^{\tau}(\Omega)}}{\|g\|_{L_{2}(\mathcal{D})}} .
$$

For the next step we assume a regularity assumption in the sense of(8.2.5). Then we can bound the second factor by the Bramble-Hilbert lemma, which gives

$$
\inf _{\phi \in \pi_{\tau}(\mathcal{D})}\left\|w_{g}-\phi\right\|_{W_{2}^{\tau}(\Omega)} \leq \rho_{\mathcal{D}}^{\sigma-\tau}\left|w_{g}\right|_{W_{2}^{\sigma}(\mathcal{D})} \leq C \rho_{\mathcal{D}}^{\sigma-\tau}\|g\|_{L_{2}(\mathcal{D})}
$$

For the first part we use the Cea Lemma [10] to get

$$
\left\|u-p_{\mathcal{D}}\right\|_{W_{2}^{\tau}(\mathcal{D})} \leq C\|u\|_{W_{2}^{\tau}(\mathcal{D})}
$$

Putting things together yields the claimed estimate.
This is a local sampling inequality. We shall now apply a covering argument to extend this result to the global domain. To derive a global sampling inequality we again cover the global domain $\Omega$ by nice small domains $\mathcal{D}_{t}$ as described in Theorem 3.3.10. From now on we assume $h \leq Q(k, \theta) R_{0}$, so that the construction is applicable. Then we can prove a global version of our sampling inequality.

Theorem 8.5.3 Under the assumption from Proposition 8.5 .2 and Theorem 3.3.10 there exists a constant $h_{0}>0$ with the following property. There is a constant $C>0$ such that for all functions $u \in W_{2}^{\tau}(\Omega)$ and all $h \leq h_{0}$ the sampling inequality

$$
\|u\|_{L_{2}(\Omega)} \leq C\left(h^{\sigma-\tau}\|u\|_{W_{2}^{\tau}(\Omega)}+h^{\tau-\beta+d / 2} \gamma^{-1}\|a(u, \phi)\|_{\ell_{2}(n \# T r)}\right)
$$

holds, where $\left\{\phi_{j}\right\}_{j=z \ell+1, \ldots,(z+1) \ell}$ is an orthonormal system in $\pi_{\tau}\left(\mathcal{D}_{z+1}\right)$ for all $z=0, \ldots, \# T_{r}-1$.

Proof: The decomposition of $\Omega$ together with Proposition 8.5 .2 and

$$
\frac{\sin \theta}{2(1+\sin \theta) Q(k, \theta)} h=2 r \leq \rho_{\mathcal{D}_{t}} \leq 2 R=\frac{2}{Q(k, \theta)} h
$$

shows that we have

$$
\begin{aligned}
\|u\|_{L_{2}(\Omega)}^{2} & =\int_{\Omega}|u(x)|^{2} d x \\
& \leq \sum_{t \in T_{r}} \int_{\mathcal{D}_{t}}|u(x)|^{2} d x=\sum_{t \in T_{r}}\|u\|_{L_{2}\left(\mathcal{D}_{t}\right)}^{2} \\
& \leq C \sum_{t \in T_{r}}\left(h^{\sigma-\tau}\|u\|_{W_{2}^{\tau}\left(\mathcal{D}_{t}\right)}+h^{\tau-\beta+d / 2} \gamma^{-1}\|a(u, \phi)\|_{\ell_{2}(n)}\right)^{2},
\end{aligned}
$$

where we have used the fact that we may choose the same constant $C>0$ for all regions $\mathcal{D}_{t}$. Further calculation yields

$$
\begin{aligned}
\|u\|_{L_{2}(\Omega)}^{2} & \leq C \sum_{t \in T_{r}}\left(h^{\sigma-\tau}\|u\|_{W_{2}^{\tau}\left(\mathcal{D}_{t}\right)}+h^{\tau-\beta+d / 2} \gamma^{-1}\|a(u, \phi)\|_{\ell_{2}(n)}\right)^{2} \\
& \leq C\left(h^{2(\sigma-\tau)} \sum_{t \in T_{r}}\|u\|_{W_{2}^{\tau}\left(\mathcal{D}_{t}\right)}^{2}+h^{2(\tau-\beta+d / 2)} \gamma^{-2} \sum_{t \in T_{r}}\|a(u, \phi)\|_{\ell_{2}(n)}^{2}\right) \\
& \leq C M_{1}\left(h^{2(\sigma-\tau)}\|u\|_{W_{2}^{\tau}(\Omega)}^{2}+h^{2(\tau-\beta+d / 2)} \gamma^{-2}\|a(u, \phi)\|_{\ell_{2}(n \# T r)}^{2}\right),
\end{aligned}
$$

where the last estimate follows from Theorem 3.3.10, since

$$
\sum_{t \in T_{r}}\|u\|_{W_{p}^{k+s}\left(\mathcal{D}_{t}\right)}^{p} \leq M_{1}\|u\|_{W_{p}^{k+s}(\Omega)}^{p} .
$$

This finishes the proof.
Please note the special meaning of the parameter $h$. It is not the usual fill distance of a discrete set $X \subset \Omega$. It rather mimics the meaning of $h$ in the finite element literature as the size of the local patches.
The error analysis of this section relies on good polynomial projections. This indicates some possible generalization in the context of generalized finite elements.

## Chapter 9

## Discussion and Outlook

In this thesis we have systematically generalized the concept of sampling inequalities to various situations, and we have illustrated applications. The first main part considered strong sampling inequalities. We derived sampling inequalities for infinitely smooth functions where the sampling order turned out to vary exponentially in the fill distance. As a special case our technique reproduces the well known error estimates for classical interpolation in the native spaces of Gaussian and inverse Multiquadric kernels. However, even in the special case of interpolation the optimal convergence rates are not known. Further research should address better sampling orders by avoiding the boundary effect. Although we did not pay much attention on this detail, a main drawback of the estimates lies in the constants involved, which depend exponentially on the space dimension. To avoid this spectral growth one should consider sparse grids.
We further presented a deterministic error analysis for support vector regression algorithms. We restricted ourselves to the $\epsilon$ - and the $\nu$-SVR, but the described procedure can be easily generalized to all learning algorithms with penalty terms induced by kernels whose native spaces are Sobolev spaces. The sampling orders we found are optimal and were confirmed numerically. The error analysis does not depend on any assumptions on the inaccuracy of the given data, so one should combine them with stochastical models on the noise to improve parameter choices.
As an auxiliary result we proved a Bernstein inequality, which provides equivalence constants between the Sobolev- and the $L_{2}$-norm on a finite dimensional space of translates of an RBF. For that, we need a technical condition on the distribution of centers, which seems to be artificial. Further research should overcome this.
The second main part addressed weak sampling inequalities. We considered stationary convolution-type data, which generalizes the usual finite volume methods. We derived a sampling inequality for this situation, which forms an important step towards an a priori error analysis of MLPG methods. Finally we derived a deterministic error analysis for the solution of regularized variational problems, which arise naturally as Ritz-Galerkin approximations of pde's in weak formulations. We presented sampling inequalities for both kernel based and polynomial ansatz spaces. For the analysis of MLPG methods it would be useful to prove sampling inequalities for other kinds of weak data that, e.g., covers derivatives.

## Bibliography

[1] A.N. Agadzhanov. Functional properties of Sobolev spaces of infinite order. Soviet. Math. Dokl., 38, No. 1:88-92, 1989.
[2] R. Arcangéli, M.C. López de Silanes, and J.J. Torrens. An extension of a bound for functions in Sobolev spaces, with applications to ( $m, s$ )-spline interpolation and smoothing. Numer. Math., 107(2):181-211, 2007.
[3] R. Arcangéli, M.C. López de Silanes, and J.J. Torrens. Estimates for functions in sobolev spaces defined on unbounded domains. J. Approx. Theory, 2008. doi:10.1016/j.jat.2008.09.001.
[4] S.N. Atluri. The meshless method (MLPG) for domain and BIE discretizations. Tech Science Press, Encino, CA, 2005.
[5] S.N. Atluri and S.P. Shen. The Meshless Local Petrov-Galerkin (MLPG) Method. Tech Science Press, Encino, CA, 2002.
[6] S.N. Atluri and T. Zhu. A new meshless local Petrov-Galerkin (MLPG) approach in computational mechanics. Computational Mechanics, 22:117-127, 1998.
[7] S.N. Atluri and T. Zhu. A new meshless local Petrov-Galerkin (MLPG) approach to nonlinear problems in computer modeling and simulation. Computer Modeling and Simulation in Engineering, 3:187-196, 1998.
[8] I. Babuska, U. Banerjee, and J.E. Osborn. Survey of meshless and generalized finite element methods: A unified approach. Acta Numerica, 12:1-125, 2003.
[9] P. Borwein and T. Erdelyi. Polynomials and Polynomial Inequalities. Springer, New York, 1995.
[10] D. Braess. Finite Elements. Theory, Fast Solvers and Applications in Solid Mechanics. Cambridge University Press, Cambridge, 2001.
[11] S. Brenner and L. Scott. The Mathematical Theory of Finite Element Methods. Springer, New York, 1994.
[12] C. C. Chang and C. L. Lin. Training $\nu$-Support Vector Regression: Theory and Algorithms. Neural Computation, 14(8):1959-1977, 2002.
[13] E. W. Cheney. An introduction to approximation theory. McGraw-Hill, New York, 1966.
[14] P. G. Ciarlet. The Finite Element Method For Elliptic Problems. Studies in Mathematics And Its Applications. North-Holland Publishing Company, Amsterdam, New York, Oxford, 1978.
[15] S. de Marchi and R. Schaback. Stability of Kernel-Based Interpolation. Advances in Computational Mathematics, 2008. DOI: 10.1007/s10444-008-9093-4.
[16] M. Dobrowolski. Angewandte Funktionalanalysis. Springer, Berlin Heidelberg, 2006.
[17] T. Evgeniou, M. Pontil, and T. Poggio. Regularization Networks and Support Vector Machines. Advances in Computational Mathematics, 13:1-50, 2000.
[18] G.E. Fasshauer. Solving partial differential equations by collocation with radial basis functions. In A.L. Mehaute, C. Rabut, and L.L. Schumaker, editors, Surface Fitting and Multiresolution Methods, pages 131-138. Vanderbilt University Press, Nashville, 1997.
[19] G.E. Fasshauer. On the numerical solution of differential equations with radial basis functions. In C.S. Chen, C.A. Brebbia, and D.W. Pepper, editors, Boundary Element Technology XIII, pages 291-300. WIT Press, Southampton, 1999.
[20] C. Franke and R. Schaback. Convergence oreder estimates of meshless collocation methods using radial basis functions. Advances in Computational Mathematics, 8:381-399, 1998.
[21] C. Franke and R. Schaback. Solving partial differential equations by collocation using radial basis functions. Advances in Computational Mathematics, 93:73-82, 1998.
[22] E. Freitag and R. Busam. Funktionentheorie 1. Springer Verlag, Berlin, 2000.
[23] F. Girosi. An Equivalence Between Sparse Approximation and Support Vector Machines. Neural Computation, 10 (8):1455-1480, 1998.
[24] W. Hackbusch. Theorie und Numerik elliptischer Differentialgleichungen. Teubner, Stuttgart, 1986.
[25] A. Iske and T. Sonar. On the structure of function spaces in optimal recovery of point functionals for eno-schemes by radial basis functions. Numerische Mathematik, 74:177-201, 1996.
[26] K. Jetter, J. Stöckler, and J.D. Ward. Norming sets and scattered data approximation on spheres. In Approximation Theory IX, Vol. II: Computational Aspects, pages 137 144. Vanderbilt University Press, 1998.
[27] L. Ling, R. Opfer, and R. Schaback. Results on meshless collocation techniques. Engeneering Analysis with Boundary Elements, 30:247-253, 2006.
[28] L. Ling and R. Schaback. Stable and Convergent Unsymmetric Meshless Collocation Methods. to appear in SIAM J. Numer. Anal., 2007.
[29] F. Lu and H. Sun. Positive definite dot product kernels in learning theory. Advances in Computational Mathematics, 22:181-198, 2005.
[30] G. Lube. Theorie und Numerik elliptischer Randwertprobleme. Lecture note, available at:
http://www.num.math.uni-goettingen.de/lube/FEM1_06_akt.pdf.
[31] W. R. Madych. An estimate for multivariate interpolation II. J. Approx. Theory, 142:116-128, 2006.
[32] W.R. Madych and S.A. Nelson. Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation. J. Approx. Theory, 70:94-114, 1992.
[33] J.M. Melenk. On Approximation in Meshless Methods, pages 65-141. Universitext. Springer Berlin / Heidelberg, 2005.
[34] C. A. Micchelli and M. Pontil. Learning the Kernel Function via Regularization. Journal of Machine Learning Research, 6:1099-1125, 2005.
[35] St. Müller and R. Schaback. A Newton Basis for Kernel Spaces. J. Approx. Theory, 2008. DOI: 10.1016/j.jat.2008.10.014.
[36] F.J. Narcowich and J.D Ward. Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. Math. Comput., 63:661-687, 1994.
[37] F.J. Narcowich and J.D Ward. Scattered-data interpolation on $\mathcal{R}^{n}$ : Error estimates for radial basis and band-limited functions. SIAM J. Math. Anal., 36:284-300, 2004.
[38] F.J. Narcowich, J.D. Ward, and H. Wendland. Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. Mathematics of Computation, 74:743-763, 2005.
[39] F.J. Narcowich, J.D. Ward, and H. Wendland. Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. Constructive Approximation, 24:175-186, 2006.
[40] K. Pelckmans, I. Goethals, J.D. Brabanter, J.A.K. Suykens, and B.D. Moor. Componentwise Least Squares Support Vector Machines, volume 177/2005 of Studies in Fuzziness and Soft Computing, pages 77-98. Springer Berlin / Heidelberg, 2005.
[41] C. Rieger. Approximative Interpolation. Master's thesis, NAM, Göttingen, 2005.
[42] C. Rieger and B. Zwicknagl. Deterministic Error Analysis of Support Vector Machines and Related Regularized Kernel Methods. to appear in: J. Machine Learning Research.
[43] C. Rieger and B. Zwicknagl. Sampling Inequalities for Infinitely Smooth Functions, with Applications to Interpolation and Machine Learning. Adv. Comp. Math., 2008. DOI: 10.1007/s1.10444-008-9089-0.
[44] R. Schaback. Kernel-Based Meshless Methods. Lecture note, available at: http://www.num.math.uni-goettingen.de/schaback/ teaching/07SS/vorl/kernel.pdf.
[45] R. Schaback. Unsymmetric Meshless Methods for Operator Equations. preprint, 2006.
[46] R. Schaback. Convergence of Unsymmetric Kernel-Based Meshless Collocation Methods. SIAM Journal of Numerical Analysis, 45/1:333-351, 2007.
[47] R. Schaback. Recovery of Functions from Weak Data Using Unsymmetric Meshless Kernel-Based Methods. Applied Numerical Mathematics, 58 (5):726-741, 2008.
[48] R. Schaback and H. Wendland. Inverse and saturation theorems for radial basis function interpolation. Mathematics of Computation, 71:669-681, 2002.
[49] R. Schaback and H. Wendland. Kernel Techniques: From Machine Learning to Meshless Methods. Acta Numerica, 15:543-639, 2006.
[50] B. Schölkopf, C. Burges, and V.Vapnik. Extracting support data for a given task. In Proceedings, First International Conference on Knowledge Discovery and Data Mining. CA:AAAI Press., Menlo Park, 1995.
[51] B. Schölkopf and A. J. Smola. Learning with kernels - Support Vector Machines, Regularisation, and Beyond. MIT Press, Cambridge, Massachusetts, 2002.
[52] B. Schölkopf, R. C. Williamson, and P. L. Bartlett. New Support Vector Algorithms. Neural Computation, 12:1207-1245, 2000.
[53] T. Sonar. Optimal recovery using thin plate splines in finite volume methods for the numerical solution of hyperbolic conservation laws. IMA J. Numer. Anal., 16:549581, 1996.
[54] T. Sonar. On the construction of essentially non-oscillatory finfite volume approximations to hyperbolic conservation laws on general triangulationss: Polynomial recovery, accuracy, and stencil selection. Comput. Methods in Appl. Mechanics and Engineering, 140:157-181, 1997.
[55] T. Sonar. On families of pointwise optimal finite volume ENO approximations. SIAM J. Numer. Anal., 35:2350-2369, 1998.
[56] Elias M. Stein. Singular integrals and differentialbility properties offunctions. Princeton University Press, Princeton, N.J., 1970.
[57] V. Vapnik. The nature of statistical learning theory. Springer-Verlag, New York, 1995.
[58] G. Wahba. Spline Models for Observational Data, CBMS-NSF. Regional Conference Series in Applied Mathematics. Siam, Philadelphia, 1990.
[59] H. Wendland. Ein Beitrag zur Interpolation mit radialen Basisfunktionen. Master's thesis, NAM, Göttingen, 1994.
[60] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial basis functions of minimal degree. Adv. Comput. Math., 4:389-396, 1995.
[61] H. Wendland. Error estimates for interpolation by compactly supported radial basis functions of minimal degree. J. Approx. Th., 93:258-272, 1998.
[62] H. Wendland. Meshless Galerkin methods using radial basis functions. Math. Comput., 68:1521-1531, 1999.
[63] H. Wendland. Local polynomial reproduction and moving least squares approximation. IMA J. Numer. Anal., 21:285-300, 2001.
[64] H. Wendland. On the convergence of a general class of finite volume methods. SIAM Journal of Numerical Analysis, 43:987-1002, 2005.
[65] H. Wendland. Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2005.
[66] H. Wendland. On the stability of meshless symmetric collocation for boundary value problems. BIT Numerical Mathematics, 47:455-468, 2007.
[67] H. Wendland and C. Rieger. Approximate Interpolation. Numerische Mathematik, 101:643-662, 2005.
[68] J. Wloka. Partielle Differentialgleichungen: Sobolevräume und Randwertaufgaben. Mathematische Leitfäden. Teubner, Stuttgart, 1982.
[69] Z. Wu. Hermite-Birkhoff interpolation of scattered data by radial basis functions. Approximation Theory Appl., 8:1-10, 1992.
[70] B. Zwicknagl. Power Series Kernels. Constructive Approximation, 29 (2):61-84, 2009.

## Acknowledgement

It is a great pleasure to thank my supervisor Professor Robert Schaback for giving me the opportunity to work on this topic and for many stimulating discussions. I am grateful for all his valuable advice and support during the time I spent on this thesis.
I would like to thank Professor Gerd Lube for his willingness to act as referee and his inspiring suggestions.
I am grateful to Professor Holger Wendland for introducing me to the field of sampling inequalities and his support.
I would like to thank Barbara Zwicknagl for the inspiring collaboration on which some chapters of this thesis are based, for numerous fruitful and clarifying discussions and for thoroughly proof-reading this thesis uncountably many times. Beside this, I would like to thank her for the great time with her, for providing inestimable support during the last years and for always being there for me.
The financial support by the Deutsche Forschungsgemeinschaft through the Graduiertenkolleg 1023 "Identifikation in mathematischen Modellen: Synergie stochastischer und numerischer Methoden" is gratefully acknowledged. Furthermore, I would like to thank the Institut für Numerische und Angewandte Mathematik der Universität Göttingen for the financial support and the chance to work in an inspiring environment.
Finally, I would like to thank my parents for their permanent support and their understanding.

## Curriculum Vitae

Christian Rieger
Schlözerweg 18
Date of birth: August 20, 1981
37085 Göttingen
Email: crieger@math.uni-goettingen.de

Place of birth: Göttingen
Citizenship: German

## EDUCATION

1987 to 1991 Höltyschule Göttingen
1991 to 1993 Lutherschule Göttingen
1993 to 2000 Theodor-Heuss-Gymnasium Göttingen
1997 Bradfield College, Great Britain
2000 to 2008 Georg-August University, Göttingen
Vordiplom University of Göttingen, April 2002
Major: Mathematics, Minor: Computer Science

Diplom University of Göttingen, January 2005
Major: Mathematics, Minor: Computer Science
Thesis: 'Approximative Interpolation mit radialen Basisfunktionen'.
Supervisor: Professor Holger Wendland

