Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn
Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

M. Griebel, C. Rieger

Reproducing kernel Hilbert spaces for parametric partial differential equations

INS Preprint No. 1511

June 2015

REPRODUCING KERNEL HILBERT SPACES FOR PARAMETRIC PARTIAL DIFFERENTIAL EQUATIONS

MICHAEL GRIEBEL* AND CHRISTIAN RIEGER†

Abstract. In this article, we present kernel methods for the approximation of quantities of interest which are derived from solutions of parametric partial differential equations. We explicitly construct a reproducing kernel Hilbert space containing the quantity of interest as a function of the parameters from a priori information on parameters in the differential equation. Based on the problem-adapted reproducing kernel, we suggest a regularized reconstruction technique from machine learning in order to approximate the quantity of interest from a finite number of point values. We present a deterministic a priori error analysis for this reconstruction process yielding a sub-exponential convergence order due to the smoothness of the quantity of interest as function of the parameters. The error estimates explicitly take into account the error of the numerical evaluation of the quantity of interest for fixed sets of parameters. This leads to a coupling condition between this evaluation error which contains the error of the numerical solution of the associated partial differential equation and the error due to the sampling approximation of the quantity of interest.

1. Introduction. In many practical applications, there is a growing interest in modeling physical processes with random input parameters. The random input usually accounts for inaccurate or incomplete measurements of some initial state or material property. We assume in the following that the physical process is modeled as a partial differential equation. As model example, we will restrict ourselves to a Dirichlet-Poisson problem with parametric diffusion coefficient. The uncertain input is given by the parametric diffusion coefficient $a^{(N_P)} : R_{N_P} \times \mathcal{D} \to \mathbb{R}$ where $R_{N_P} \subset \mathbb{R}^{N_P}$ is a finite dimensional parameter space. We will consider the most simple example for a parameter space: for numbers $r_j < 1$ and $1 \leq j \leq N_P$ we use $R_{N_P} := \prod_{j=1}^{N_P} (-r_j, r_j) \subset (-1,1)^{N_P}$ (see also (2.4)). The number N_P determines the effective dimension of the parameter space. The restriction to this high but finite dimensional parameter space, i.e., $N_P < \infty$, is referred to as *finite noise assumption* in the literature, see [7, 28]. The parametric partial differential equation is now given on a sufficiently nice domain $\mathcal{D} \subset \mathbb{R}^d$ and $G \in L^2(\mathcal{D})$

$$\begin{aligned} -\operatorname{div}\left(a^{(N_P)}(\boldsymbol{y},\boldsymbol{x})\,\nabla u^{(N_P)}(\boldsymbol{y},\boldsymbol{x})\right) &= G(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \mathcal{D} \text{ and all } \boldsymbol{y} \in R_{N_P} \\ u^{(N_P)}(\boldsymbol{y},\boldsymbol{x}) &= 0 \quad \text{for all } \boldsymbol{x} \in \partial\mathcal{D} \text{ and all } \boldsymbol{y} \in R_{N_P}. \end{aligned} \tag{1.1}$$

The solution of (1.1) is, by construction, a function $u^{(N_P)} : R_{N_P} \times \mathcal{D} \to \mathbb{R}$. Motivated by practical considerations, we do not consider $u^{(N_P)}$ but instead we work directly with some derived quantity of interest which will be modeled by a linear functional $q^{(N_P)}$ on the solution space, i.e.,

$$Q^{(N_P)}(\boldsymbol{y}) := q^{(N_P)}\left(u^{(N_P)}(\boldsymbol{y},\cdot)\right) \in \mathbb{R}. \tag{1.2}$$

Hence, $Q^{(N_P)} : R_{N_P} \to \mathbb{R}$ is a parameter-dependent function itself. The basic task is now to reconstruct the map $Q^{(N_P)}$ from sampled data $Q_k^{(N_P)} \cong Q^{(N_P)}(\boldsymbol{y}_k)$ at parameter values $\boldsymbol{y}_k \in \mathbb{Y}_{N_S} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N_S}\} \subset R_{N_P}$, $1 \leq k \leq N_S$, where here and in the following we shall denote the number of sampling points by $N_S \in \mathbb{N}$. The values $Q_k^{(N_P)}$ are generated by solving the parametric Poisson

*Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6-8, 53115 Bonn, Germany, and Fraunhofer-Institut für Algorithmen und Wissenschaftliches Rechnen SCAI, Schloss Birlinghoven, 53754 Sankt Augustin, Germany

†Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6-8, 53115 Bonn, Germany

problem (1.1) for $\boldsymbol{y}_k \in R_{N_P}$ and applying $q^{(N_P)}$. In general, this cannot be done analytically but only numerically. To this end, we have to introduce a discrete spacial trial space $\mathbb{V}_{N_\mathcal{D}} \subset \mathbb{V} := H_0^1(\mathcal{D})$. Then we can solve for a $u^{(N_P,N_\mathcal{D})}(\cdot,\boldsymbol{y}_k) \in \mathbb{V}_{N_\mathcal{D}}$ satisfying

$$-\operatorname{div}\left(a^{(N_P)}\left(\boldsymbol{y}_k,\cdot\right)\nabla u^{(N_P,N_\mathcal{D})}\left(\boldsymbol{y}_k,\cdot\right)\right) = G \quad \text{and} \quad u^{(N_P,N_\mathcal{D})}\left(\boldsymbol{y}_k,\cdot\right)|_{\partial\mathcal{D}} \equiv 0. \tag{1.3}$$

Since the numerical solution $u^{(N_P,N_\mathcal{D})}\left(\boldsymbol{y}_k,\cdot\right)$ will only approximate the true solution $u^{(N_P)}\left(\boldsymbol{y}_k,\cdot\right)$, we can compute only perturbed samples

$$Q_k^{(N_P,N_\mathcal{D})} := q^{(N_P)}\left(u^{(N_P,N_\mathcal{D})}\left(\boldsymbol{y}_k,\cdot\right)\right) \approx Q_k^{(N_P)} = q^{(N_P)}\left(u^{(N_P)}\left(\boldsymbol{y}_k,\cdot\right)\right). \tag{1.4}$$

We will use the quantities[1]

$$\epsilon_k(N_\mathcal{D}) := \left\|u^{(N_P,N_\mathcal{D})}\left(\boldsymbol{y}_k,\cdot\right) - u^{(N_P)}\left(\boldsymbol{y}_k,\cdot\right)\right\|_\mathbb{V} \quad \text{and} \quad \epsilon_{N_\mathcal{D}} := \max_{\boldsymbol{y}_k\in\mathbb{Y}_{N_S}} \epsilon_k(N_\mathcal{D}) \tag{1.5}$$

to denote the numerical error measured in the energy norm $\|\cdot\|_\mathbb{V}$ with $\mathbb{V} := H_0^1(\mathcal{D})$. Here, we do not discuss any specific method to solve equation (1.3). We only assume that it is possible to design a discretization space $\mathbb{V}_{N_\mathcal{D}}$ such that $\epsilon_{N_\mathcal{D}} \to 0$ for $N_\mathcal{D} \to \infty$. The maximum in the error term $\epsilon_k(N_\mathcal{D})$ in (1.5) for a fixed parameter value reflects the fact that we do not assume any difference in importance for the parameter values.

The final task is then to reconstruct a high-dimensional function from perturbed (possibly scattered) samples. But this is the classical setup for many machine learning algorithms. At this point, we stress the fact that it is necessary to use a regularized reconstruction method since we will never have exact data at our disposal. Many approaches in the literature are interpolatory by nature (see [5] and succeeding articles). In our opinion, this seems somewhat questionable in view of inexact data and the resulting ill-posedness of the associated interpolation problem. Thus, we will present an algorithm for semi–supervised learning by means of a regularization method. Other non-interpolatory methods, which are in particular based on polynomial projections, can be found in [22]. The basic structure is a variational problem over a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_K$ containing real-valued functions of the form

$$\tilde{Q}^{(N_S;\lambda_K,\lambda_\mathcal{M};\epsilon_{N_\mathcal{D}};N_P)} := \arg\min_{h\in\mathcal{H}_K} \sum_{k=1}^{N_S} V\left(Q_k^{(N_P,N_\mathcal{D})}, h\left(\boldsymbol{y}_k\right)\right) + \lambda_K \|h\|_{\mathcal{H}_K}^2 + \lambda_\mathcal{M} \|h\|_\mathcal{M}^2, \tag{1.6}$$

where $V : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ denotes a loss function, $\lambda_K, \lambda_\mathcal{M} > 0$ are fixed regularization parameters and $\|\cdot\|_\mathcal{M}$ encodes geometric information on the samples $Q_k^{(N_P,N_\mathcal{D})}$. The sum in (1.6) acts as a fidelity term which is weighted by the number of sampling points N_S. As loss function for the error analysis, we will choose Vapnik's ϵ-insensitive loss function [30]. We will present a choice for the geometric term in Section 5. The intuition behind this optimization problem is to implicitly make use of the fact that the samples $Q_k^{(N_P,N_\mathcal{D})}$ will lie on a low dimensional structure due to the covariance decay of the input field $a^{(N_P)}$. This decay will also lead to a natural choice for the RKHS which we will present in Section 2. For an approach to employ the regularity properties in a polynomial setting we refer to [8]. We would like, however, to make use of the superior properties

[1] Note here that we could allow also for a perturbed functional $\tilde{q}^{N_P}$ without changing our error model. Such a perturbation may stem from a further numerical discretization process for q^{N_P}.

of kernel-based reconstruction methods. Most of the additional features of kernel-based methods stem from the fact that the interpolant is also a best-approximand in the norm of the Hilbert space. Reproducing kernels in the context pf parametric partial differential equations also arise in [14, 19, 21] and in the context of kriging [15]. It is well known in the theory of uncertainty quantification that the dependence of the quantity of interest on the parameter, i.e., $\boldsymbol{y} \mapsto Q^{(N_P)}(\boldsymbol{y}) \in \mathbb{R}$ is an analytical map around the origin if $a^{(N_P)}$ is regular enough. Hence, one would expect to obtain *sub-exponential* convergence rates for the approximation $\tilde{Q}^{(N_S;\lambda_K,\lambda_{\mathcal{M}};\epsilon_{N_{\mathcal{D}}};N_P)} \to Q^{(N_P)}$ for $N_S \to \infty$ and $\epsilon_{N_{\mathcal{D}}}, \lambda_K, \lambda_{\mathcal{M}} \to 0$ with a reasonable loss function V. Note here that N_P is considered as fixed constant. For practical purposes, it is important to balance all error contributions. We will derive such a coupling by means of a deterministic a priori error analysis provided by a *sampling inequality* given in Theorem 2.2, which is in the spirit of recent work on such estimates, see, e.g., [1, 2, 3, 17, 20, 23, 24, 26, 27, 29, 31]. We will show (see Corollary 5.4) that there are (a priori known) parameters of (1.6) such that a sub-exponential bound of the form

$$\begin{aligned}\left\|\tilde{Q}^{(N_S;\lambda_K,\lambda_{\mathcal{M}};\epsilon_{N_{\mathcal{D}}:N_P})} - Q^{(N_P)}\right\|_{L^\infty(R)} &\leq \exp(-C_1/\sqrt{h_{\mathbb{Y}_{N_S},R_{N_P}}})\left(\left\|Q^{(N_P)}\right\|_{\mathcal{H}_K} + \left\|Q^{(N_P)}\right\|_{\mathcal{M}}\right) \\ &\quad + c\sqrt{\epsilon_{N_{\mathcal{D}}}},\end{aligned}$$

holds, where $\tilde{Q}^{(N_S;\lambda_K,\lambda_{\mathcal{M}};\epsilon_{N_{\mathcal{D}}}:N_P)}$ is given in (1.6), and $Q^{(N_P)} \in \mathcal{H}_K$ is the true solution of (1.1). Moreover,

$$h_{\mathbb{Y}_{N_S},R_{N_P}} := \sup_{\boldsymbol{x}\in R_{N_P}} \min_{\boldsymbol{x}_j\in\mathbb{Y}_{N_S}} \|\boldsymbol{x}-\boldsymbol{x}_j\|_2 \tag{1.7}$$

denotes the fill distance of the discrete set $\mathbb{Y}_{N_S} \subset R_{N_P}$ (see also [32] for more discussion on this quantity) and the regularization parameters satisfy

$$\sqrt{\lambda_K} \sim \exp(-C_1/\sqrt{h_{\mathbb{Y}_{N_S},R_{N_P}}}) \sim \sqrt{\lambda_{\mathcal{M}}}.$$

This finally allows us to couple[2] the numerical discretization error $\epsilon_{N_{\mathcal{D}}}$ to the number of sampling nodes via $\sqrt{\epsilon_{N_{\mathcal{D}}}} \sim \exp\left(-C_1/\sqrt{h_{\mathbb{Y}_{N_S},R_{N_P}}}\right)$.

Note at this point that we treat here only an easy model problem and implicitly assume a short correlation length by using the isotropic fill-distance $h_{\mathbb{Y}_{N_S},R_{N_P}}$.

The remainder of this paper is organized as follows: In section 2, we derive a sampling inequality for a specific class of multivariate Taylor kernels. This generalizes [35, Sec. 6] to our situation. In Section 3, we collect basic regularity results from uncertainty quantification. In particular, we demonstrate the relation to Taylor kernels. Furthermore, we identify a geometric substructure in the parameter space associated with the quantity of interest. In Section 5, we present a classical approach to semi–supervised learning which takes this geometric insight into account. Finally, we show how sampling inequalities can be used to derive an exponential convergence rate of the reconstruction scheme. Here, the rate depends on the covariance decay of the input field a^{N_P}.

[2]The choice presented here is optimized for good approximation rates. If the data is severely corrupted by noise, other statistically motivated choices for the regularization parameters could lead to slightly different results. We view the error in the data as numerical residual and hence as deterministic. Furthermore, the error in the data could at least theoretically be made arbitrarily small by allowing more numerical work in the solution of the partial differential equation for a fixed parameter.

2. Power series kernels. In this section, we follow mainly [26, 34, 35]. We will denote multi-indices by small bold Greek letters, e.g., $\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}$, and set $\boldsymbol{\nu}! = \prod_{j=1}^{N_P} \nu_j!$ and $\boldsymbol{\nu}^{\boldsymbol{\alpha}} = \prod_{j=1}^{N_P} \nu_j^{\alpha_j}$ for $\boldsymbol{\alpha} \geq \mathbf{0}$. We write $\boldsymbol{\alpha} \geq \mathbf{0}$ if $\alpha_j \geq 0$ for all $1 \leq j \leq N_P$. Furthermore, we use the notation $|\boldsymbol{\nu}|_1 := \nu_1 + \cdots + \nu_{N_P}$ for $\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}$ and $|\boldsymbol{\nu}|_\infty := \max_{j=1}^{N_P} |\nu_j|$.

Let $\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}$ be a multi-index and $w_{\boldsymbol{\nu}}$ be a sequence of positive numbers such that the summability condition $\sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \frac{w_{\boldsymbol{\nu}}}{\boldsymbol{\nu}!^2} < \infty$ holds. Then, a power series kernel $K : (-1,1)^{N_P} \times (-1,1)^{N_P} \to \mathbb{R}$

$$K(\boldsymbol{x}, \boldsymbol{y}) := \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \frac{w_{\boldsymbol{\nu}}}{\boldsymbol{\nu}!^2} \boldsymbol{x}^{\boldsymbol{\nu}} \boldsymbol{y}^{\boldsymbol{\nu}} \tag{2.1}$$

is well-defined. Such kernels are known to reproduce in Hilbert spaces $\mathcal{H}_K$ of functions

$$\mathcal{H}_K\left((-1,1)^{N_P}\right) := \left\{ f : (-1,1)^{N_P} \to \mathbb{R} \ : \ f(\boldsymbol{x}) = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \hat{f}_{\boldsymbol{\nu}} \boldsymbol{x}^{\boldsymbol{\nu}} \text{ s.t. } \|f\|_{\mathcal{H}_K\left((-1,1)^{N_P}\right)} < \infty \right\} \tag{2.2}$$

with the inner product given by

$$(f,g)_{\mathcal{H}_K\left((-1,1)^{N_P}\right)} := \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} w_{\boldsymbol{\nu}}^{-1} D^{\boldsymbol{\nu}} f(\mathbf{0}) D^{\boldsymbol{\nu}} g(\mathbf{0}). \tag{2.3}$$

Next, we define the anisotropic hypercube R_{N_P} for $\boldsymbol{r} = (r_1, \ldots, r_{N_P})^T$ with $r_j < 1$ as

$$R_{N_P} := \prod_{j=1}^{N_P} (-r_j, r_j) \subset (-1,1)^{N_P} \subset \mathbb{R}^{N_P}. \tag{2.4}$$

Following [26], we consider the embedding constants of the non-standard function spaces given by (2.2) into classical Sobolev spaces. To this end, we introduce for $s \in \{1, \infty\}$ the function spaces

$$W_{2;s}^k(R_{N_P}) := \left\{ f \in L^2(R_{N_P}) \ : \ \|f\|_{W_{2;s}^k(R_{N_P})}^2 := \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}_0^{N_P} \\ |\boldsymbol{\alpha}|_s = j}} \int_{R_{N_P}} \left(D^{\boldsymbol{\alpha}} f(\boldsymbol{x})\right)^2 d\boldsymbol{x} < \infty \right\}. \tag{2.5}$$

The choice $s = 1$ corresponds to the classical isotropic setting, whereas the choice $s = \infty$ corresponds to tensor product Sobolev spaces. The embedding constant is given as norm of the injection

$$\mathcal{W}_s(k) : \mathcal{H}_K\left((-1,1)^{N_P}\right) \hookrightarrow W_{2;s}^k(R_{N_P}). \tag{2.6}$$

LEMMA 2.1. *Suppose that there is a $\hat{c} < 1$ such that the weights $w_{\boldsymbol{\alpha}}$ satisfy $w_{\boldsymbol{\nu}} \leq \hat{c}^{|\boldsymbol{\nu}|_1} \boldsymbol{\nu}!^2$ for all $\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}$. Then, there is a constant $C > 0$ such that*

$$\|\mathcal{W}_s(k)\| \leq \exp(C/2k)\, k! \tag{2.7}$$

for every R_{N_P} as in (2.4) and for $s \in \{1, \infty\}$.

Proof. Let $f \in \mathcal{H}_K\left((-1,1)^{N_P}\right)$ be arbitrary but fixed. Then, we get

$$\begin{aligned}
\|f\|^2_{W^k_{2;s}(R_{N_P})} &= \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \left(D^{\boldsymbol{\alpha}} f(\boldsymbol{x})\right)^2 d\boldsymbol{x} = \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \left(D^{\boldsymbol{\alpha}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \hat{f}_{\boldsymbol{\nu}} \boldsymbol{x}^{\boldsymbol{\nu}}\right)^2 d\boldsymbol{x} \\
&= \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \Big(\sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \hat{f}_{\boldsymbol{\nu}} D^{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\nu}}\Big)^2 d\boldsymbol{x} \\
&= \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \left(\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ \boldsymbol{\nu}\geq\boldsymbol{\alpha}}} \hat{f}_{\boldsymbol{\nu}} \frac{\boldsymbol{\nu}!}{(\boldsymbol{\nu}-\boldsymbol{\alpha})!} \boldsymbol{x}^{\boldsymbol{\nu}-\boldsymbol{\alpha}}\right)^2 d\boldsymbol{x} \\
&= \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \left(\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ \boldsymbol{\nu}\geq\boldsymbol{\alpha}}} \hat{f}_{\boldsymbol{\nu}} \frac{\boldsymbol{\nu}!}{\sqrt{w_{\boldsymbol{\nu}}}} \frac{\sqrt{w_{\boldsymbol{\nu}}}}{(\boldsymbol{\nu}-\boldsymbol{\alpha})!} \boldsymbol{x}^{\boldsymbol{\nu}-\boldsymbol{\alpha}}\right)^2 d\boldsymbol{x}.
\end{aligned}$$

We can use the Cauchy Schwartz inequality to obtain with $\boldsymbol{e} = (1,\dots,1)^T \in \mathbb{R}^{N_P}$

$$\begin{aligned}
\|f\|^2_{W^k_{2;s}(R_{N_P})} &\leq \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \left(\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ \boldsymbol{\nu}\geq\boldsymbol{\alpha}}} \hat{f}^2_{\boldsymbol{\nu}} \frac{\boldsymbol{\nu}!^2}{w_{\boldsymbol{\nu}}}\right)\left(\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ \boldsymbol{\nu}\geq\boldsymbol{\alpha}}} \frac{w_{\boldsymbol{\nu}}}{(\boldsymbol{\nu}-\boldsymbol{\alpha})!^2} \boldsymbol{x}^{2\boldsymbol{\nu}-2\boldsymbol{\alpha}}\right) \\
&\leq \|f\|^2_{\mathcal{H}_K\left((-1,1)^{N_P}\right)} \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ \boldsymbol{\nu}\geq\boldsymbol{\alpha}}} \frac{w_{\boldsymbol{\nu}}}{(\boldsymbol{\nu}-\boldsymbol{\alpha})!^2} \boldsymbol{x}^{2\boldsymbol{\nu}-2\boldsymbol{\alpha}} \\
&= \|f\|^2_{\mathcal{H}_K\left((-1,1)^{N_P}\right)} \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \int_{R_{N_P}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{w_{\boldsymbol{\nu}+\boldsymbol{\alpha}}}{\boldsymbol{\nu}!^2} \boldsymbol{x}^{2\boldsymbol{\nu}} \\
&= \|f\|^2_{\mathcal{H}_K\left((-1,1)^{N_P}\right)} \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{w_{\boldsymbol{\nu}+\boldsymbol{\alpha}}}{\boldsymbol{\nu}!^2} 2\boldsymbol{r}^{2\boldsymbol{\nu}+\boldsymbol{e}} \left(\prod_{\ell=1}^{N_P}(2\nu_\ell+1)\right)^{-1} \\
&= 2\,\|f\|^2_{\mathcal{H}_K\left((-1,1)^{N_P}\right)} \sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \prod_{\ell=1}^{N_P} \sum_{\nu_\ell\in\mathbb{N}_0} \frac{w_{\nu_\ell+\alpha_\ell} r_\ell^{2\nu_\ell+1}}{\nu_\ell!^2\,(2\nu_\ell+1)}.
\end{aligned}$$

Hence, we are left with the term $\prod_{\ell=1}^{N_P} \sum_{\nu_\ell\in\mathbb{N}_0} \frac{w_{\nu_\ell+\alpha_\ell} r_\ell^{2\nu_\ell+1}}{\nu_\ell!^2(2\nu_\ell+1)}$. We can invoke [34, Lemma 5], where it is shown that, for $w_{\boldsymbol{\nu}} \leq \hat{c}^{|\boldsymbol{\nu}|_1}\boldsymbol{\nu}!^2$ for all $\boldsymbol{\nu} \in N_0^{N_P}$ and for a fixed $\hat{c} < 1$, there is a constant $C > 0$

such that

$$j^{N_P} \max_{\substack{\boldsymbol{\beta}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\beta}|_1=j}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{w_{\boldsymbol{\nu}+\boldsymbol{\beta}}}{\boldsymbol{\nu}!^2} \le \exp(Cj)\, j!^2 \tag{2.8}$$

holds for all $k \in \mathbb{N}$. We have the following two results from combinatorics

$$\begin{aligned}
\#\left\{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P} \;:\; |\boldsymbol{\alpha}|_1 = j\right\} &= \binom{j+N_P-1}{N_P-1}\\
\#\left\{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P} \;:\; |\boldsymbol{\alpha}|_\infty = j\right\} &= \#\left\{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P} \;:\; |\boldsymbol{\alpha}|_\infty \le j\right\} - \#\left\{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P} \;:\; |\boldsymbol{\alpha}|_\infty \le j-1\right\}\\
&= (j+1)^{N_P} - j^{N_P}.
\end{aligned}$$

For the case $s = 1$, Stirling's formula yields the estimate with a constant $c > 0$

$$\binom{j+N_P-1}{N_P-1} \le c j^{N_P}.$$

For the second case, i.e., $s = \infty$, we also have obviously $(j+1)^{N_P} - j^{N_P} \le c j^{N_P}$ with a possibly different constant $c > 0$. Hence, we get

$$\begin{aligned}
\sum_{j=0}^{k} \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\alpha}|_s=j}} \prod_{\ell=1}^{N_P} \sum_{\nu_\ell\in\mathbb{N}_0} \frac{w_{\nu_\ell+\alpha_\ell} r_\ell^{2\nu_\ell+1}}{\nu_\ell!^2\,(2\nu_\ell+1)} &\le \sum_{j=0}^{k} j^{N_P} \max_{\substack{\boldsymbol{\beta}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\beta}|_1=j}} \sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}} \prod_{\ell=1}^{N_P} \sum_{\nu_\ell\in\mathbb{N}_0} \frac{w_{\nu_\ell+\alpha_\ell} r_\ell^{2\nu_\ell}}{\nu_\ell!^2}\\
&\le \sum_{j=0}^{k} j^{N_P} \max_{\substack{\boldsymbol{\beta}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\beta}|_1=j}} \sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^{N_P}} \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{(\boldsymbol{\nu}+\boldsymbol{\alpha})!^2}{\boldsymbol{\nu}!^2} \hat{c}^{|\boldsymbol{\nu}|_1}\\
&\le \sum_{j=0}^{k} \exp(Cj)\, j!^2 \le \exp(Ck)\, k!^2
\end{aligned}$$

This results in the assertion for $s = 1$ and $s = \infty$. □

Now we are in the position to state a sampling inequality which includes derivatives on the left-hand side. For similar arguments see also [26, Thm. 3.5] and [35, Thms 6.1 & 6.2].

THEOREM 2.2. *Suppose that there is $\hat{c} < 1$ such that $w_{\boldsymbol{\nu}} \le \hat{c}^{|\boldsymbol{\nu}|_1} \boldsymbol{\nu}!^2$ for all $\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}$. Then, for R_{N_P} as in (2.4) and for all $1 \le q \le \infty$ and all $\boldsymbol{\alpha} \in \mathbb{N}_0^{N_P}$, there are constants C_1, C_2 and $h_0 > 0$ depending only on $N_P, q, R_{N_P}, \boldsymbol{\alpha}$ and $\hat{c}$ such that for all data sets $\mathbb{Y}_{N_S} \subset R_{N_P} \subset \mathbb{R}^{N_P}$ with fill distance $h_{\mathbb{Y}_{N_S}, R_{N_P}} \le h_0$ the inequality*

$$\|D^{\boldsymbol{\alpha}} f\|_{L^q(R_{N_P})} \le \exp\left(-\frac{C_1}{\sqrt{h_{\mathbb{Y}_{N_S}, R_{N_P}}}}\right) \|f\|_{\mathcal{H}_K((-1,1)^{N_P})} + C_2 \left\| f|_{\mathbb{Y}_{N_S}} \right\|_{\ell^\infty(\mathbb{Y}_{N_S})} \tag{2.9}$$

holds for all $f \in \mathcal{H}_K\left((-1,1)^{N_P}\right)$ and for[3] *$s \in \{1, \infty\}$.*

[3]The s dependence is only in the lower order terms of the embedding constant which does not affect the rate but may affect the constants in the sampling inequality.

Proof. Note that Lemma 2.1 yields

$$\|\mathcal{W}_s(k)\| \leq \exp\left(C/2k\right) k! \leq \exp\left(1 + C/2k - k + 1/2\ln(k)\right) k^k$$

for all $k \in \mathbb{N}$ using Stirling's formula. Hence, we can apply [26, Thm. 3.5] directly with $\sigma = 1$ which shows (2.9). ∎

3. Basic parametric partial differential equation. In this section, we mainly follow [13], see also [5] for an early result on collocation. Let $\mathcal{D} \subset \mathbb{R}^d$ be a physical domain which is assumed to be smoothly bounded. We consider the model problem (1.1) and we assume the right–hand side $G \in L^2(\mathcal{D})$ to be sufficiently regular in order not to spoil the usual elliptic regularity theory. The energy space is $\mathbb{V} := H_0^1(\mathcal{D})$. The weak form of (1.1) (see also [6]) reads

$$\int_{\mathcal{D}} a^{(N_P)}(\boldsymbol{y}, \boldsymbol{x}) \nabla u^{(N_P)}(\boldsymbol{y}, \boldsymbol{x}) \nabla v(\boldsymbol{x}) \, d\boldsymbol{x} = \int_{\mathcal{D}} G(\boldsymbol{x}) \, v(\boldsymbol{x}) \, d\boldsymbol{x} \quad \text{for all } v \in \mathbb{V} = H_0^1(\mathcal{D}) \tag{3.1}$$

where we used the truncated input

$$a^{(N_P)}(\boldsymbol{y}, \boldsymbol{x}) = a_0(\boldsymbol{x}) + \sum_{k=1}^{N_P} \phi_k(\boldsymbol{x}) \, y_k,$$

where $\boldsymbol{y} = (y_1, \ldots, y_{N_P})^T$. Note at this point that there is no need for the *finite noise assumption*, i.e., $N_P < \infty$, see [13] for the corresponding results in the infinite dimensional setting. We will state therefore the (general) results from [13] but restricted to the finite dimensional setting. We invoke the *uniform (complex) ellipticity assumption* (see [13, Eq. 1.26] for the finite noise case, i.e.,

$$0 < r \leq \Re a^{(N_P)}(\boldsymbol{y}, \boldsymbol{x}) \leq \left| a^{(N_P)}(\boldsymbol{y}, \boldsymbol{x}) \right| \leq R \quad \text{for all } \boldsymbol{x} \in \mathcal{D} \text{ and all } \boldsymbol{y} \in R_{N_P}. \tag{3.2}$$

Here, we denote by $\Re$ the real part of a complex number. Simple rearranging yields that (3.2) is satisfied if the bounds

$$\sum_{k=1}^{N_P} |\phi_k(\boldsymbol{x})| \leq \sum_{k=1}^{\infty} |\phi_k(\boldsymbol{x})| \leq \Re \min\left\{a_0(\boldsymbol{x}) - r, R - a_0(\boldsymbol{x})\right\},$$

hold, where we used the fact that $\boldsymbol{y} \in R_{N_P} \subset (-1,1)^{N_P}$. The crucial observation in [13] is to sharpen this conditions into δ–admissibility. A sequence $(\rho_k)_{k\in\mathbb{N}}$ with $\rho_k > 0$ is called δ–admissible if

$$\sum_{k\in\mathbb{N}} \rho_k \, |\phi_k(\boldsymbol{x})| \leq \Re a_0(\boldsymbol{x}) - \delta, \tag{3.3}$$

for some number $0 < \delta < 2R$. For $\delta < r$, one can achieve $\rho_k > 1$ for all $k \in \mathbb{N}$. The map $\boldsymbol{y} \mapsto u^{(N_p)}(\boldsymbol{y}, \cdot) \in \mathbb{V}$ is analytical (see [13, Thm 1.2 & Lemma 2.4] or [5, Lemma 3.2]). To be precise, we have the following proposition from [13, Lemma 2.4].

PROPOSITION 3.1. *Suppose that the uniform (complex) ellipticity assumtion* (3.2) *holds with parameters* $0 < r \leq R < \infty$ *and suppose that* $\|\phi_k\|_{L^\infty(\mathcal{D})}$ *is bounded for* $1 \leq k \leq N_P$. *Then, we get*

$$u^{(N_P)}(\boldsymbol{y}, \cdot) = \sum_{k\in\mathbb{N}} \sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}^{N_P} \\ |\boldsymbol{\nu}|_1 = k}} u_{\boldsymbol{\nu}}^{(N_P)}(\cdot) \, \boldsymbol{y}^{\boldsymbol{\nu}}, \tag{3.4}$$

for $u_{\boldsymbol{\nu}}^{(N_P)} \in \mathbb{V}$ *and convergence will be understood with respect to the* $\|\cdot\|_{\mathbb{V}}$*–norm. Furthermore, there is a bound on the coefficient functions* $u_{\boldsymbol{\nu}}^{(N_P)} \in \mathbb{V}$*, i.e., for a* δ*–admissible sequence* $(\rho_k)_k$ *with* $\delta < r$*, we have*

$$\left\| u_{\boldsymbol{\nu}}^{(N_P)} \right\|_{\mathbb{V}} \leq \frac{\|G\|_{\mathbb{V}^\star}}{\delta} \boldsymbol{\rho}^{-\boldsymbol{\nu}}. \tag{3.5}$$

We clearly see a product structure in (3.5). If the δ–admissibility assumption is skipped, (3.5) deteriorates to (see [13, Eq. 1.17])

$$\left\| u_{\boldsymbol{\nu}}^{(N_P)} \right\|_{\mathbb{V}} \leq \frac{\|G\|_{\mathbb{V}^\star}}{r} \frac{|\boldsymbol{\nu}|_1!}{\boldsymbol{\nu}!} \boldsymbol{b}^{\nu}, \tag{3.6}$$

with the sequence $(b_k)_k$ $k \in \{1, \ldots, N_P\}$ defined by

$$b_k := \left(\inf_{\boldsymbol{x} \in \mathcal{D}} \int_{R_{N_P}} a\left(\boldsymbol{y}, \boldsymbol{x}\right) d\boldsymbol{y} \right)^{-1} \|\phi_k\|_{L^\infty(\mathcal{D})}. \tag{3.7}$$

Then, the product structure in (3.6) is lost due to the term $|\boldsymbol{\nu}|_1!$. This can be repaired by slightly changing the sequence (b_k), see the appendix for the computations, but we will stick to (3.5). Therefore, we will not consider this case separately in the following. The assumptions in Proposition 3.1 are simpler than those in [13] since in the finite dimensional case the summability for the $\left(\|\phi_k\|_{L^\infty(\mathcal{D})}\right)_{k \in \mathbb{N}}$ is trivial. We denote by $\langle \cdot, \cdot \rangle$ the dual pairing between $\mathbb{V}$ and $\mathbb{V}^\star$. Furthermore, we denote the Riesz representer of a linear functional λ by $\mathcal{R}(\lambda)$. We are mainly interested in reconstructing the quantity of interest (1.2), here written in the parametric form

$$\begin{aligned} Q^{(N_P)}\left(\boldsymbol{y}\right) = q^{(N_P)}\left(u^{(N_P)}\left(\boldsymbol{y}, \cdot\right)\right) &= \left\langle \sum_{k \in \mathbb{N}} \sum_{\substack{\boldsymbol{\nu} \in \mathbb{N}^{N_P} \\ |\boldsymbol{\nu}|_1 = k}} u_{\boldsymbol{\nu}}^{(N_P)} \boldsymbol{y}^{\boldsymbol{\nu}}, \mathcal{R}(q^{(N_P)}) \right\rangle_{\mathbb{V}} \\ &= \sum_{k \in \mathbb{N}} \sum_{\substack{\boldsymbol{\nu} \in \mathbb{N}^{N_P} \\ |\boldsymbol{\nu}|_1 = k}} \left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q^{(N_P)}) \right\rangle_{\mathbb{V}} \boldsymbol{y}^{\boldsymbol{\nu}} \end{aligned}$$

where we used (3.4). This directly yields the following corollary:

Corollary 3.2. *Suppose that the uniform (complex) ellipticity assumption* (3.2) *holds with parameters* $0 < r \leq R < \infty$ *and that* $\|\phi_k\|_{L^\infty(\mathcal{D})}$ *is bounded for* $1 \leq k \leq N_P$*. Then, we get*

$$Q^{(N_P)}\left(\boldsymbol{y}\right) = \sum_{k \in \mathbb{N}} \sum_{\substack{\boldsymbol{\nu} \in \mathbb{N}^{N_P} \\ |\boldsymbol{\nu}|_1 = k}} \left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q^{(N_P)}) \right\rangle_{\mathbb{V}} \boldsymbol{y}^{\boldsymbol{\nu}}. \tag{3.8}$$

Furthermore, let $(\rho_k)_k$ *be a* δ*–admissible sequence with* $\delta < r$*. Then, the coefficients are bounded by*

$$\left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q^{(N_P)}) \right\rangle_{\mathbb{V}} \leq \left\| u_{\boldsymbol{\nu}}^{(N_P)} \right\|_{\mathbb{V}} \left\| \mathcal{R}(q^{(N_P)}) \right\|_{\mathbb{V}} \leq \frac{\|G\|_{\mathbb{V}^\star} \left\| q^{(N_P)} \right\|_{\mathbb{V}^\star}}{\delta} \boldsymbol{\rho}^{-\boldsymbol{\nu}}. \tag{3.9}$$

Without the δ–admissibility assumption, we get (3.6) *and hence the bound deteriorates to*

$$\left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q^{(N_P)}) \right\rangle_{\mathbb{V}} \leq \frac{\|G\|_{\mathbb{V}^\star} \left\|q^{(N_P)}\right\|_{\mathbb{V}^\star}}{r} \frac{|\boldsymbol{\nu}|_1!}{\boldsymbol{\nu}!} \boldsymbol{b}^{\nu} \tag{3.10}$$

with the sequence $(b_k)_k$ defined in (3.7).

At this point, we can characterize the reproducing kernel Hilbert space of the form (2.2) using the weights defined in (3.9).

LEMMA 3.3. *Suppose that the uniform (complex) ellipticity assumption* (3.2) *holds with parameters $0 < r \leq R < \infty$ and suppose that $\|\phi_k\|_{L^\infty(\mathcal{D})}$ is bounded for $1 \leq k \leq N_P$. Furthermore, let $(\rho_k)_k$ be a δ–admissible sequence with $\delta < r$. Then, we get*

$$\begin{aligned} &Q^{(N_P)} \in \mathcal{H}_{K_a}\left((-1,1)^{N_P}\right) \\ &\quad := \left\{ f : (-1,1)^{N_P} \to \mathbb{R} \ : \ f(\boldsymbol{x}) = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \hat{f}_{\boldsymbol{\nu}} \boldsymbol{x}^{\boldsymbol{\nu}}, \|f\|_{\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)} < \infty \right\} \end{aligned}$$

with inner product

$$(f,g)_{\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)} := \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \frac{D^{\boldsymbol{\nu}} f(\mathbf{0}) D^{\boldsymbol{\nu}} g(\mathbf{0})}{\boldsymbol{c}^{\boldsymbol{\nu}} \boldsymbol{\nu}!^2},$$

where the weights $\boldsymbol{c}$ satisfy $c_j \in (\rho_j^{-1}, 1)$ for all $1 \leq j \leq N_P$.

Proof. Under the assumptions of Corollary 3.2, we have according to (3.9), the bound

$$Q^{(N_P)}(\boldsymbol{y}) = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \hat{Q}_{\boldsymbol{\nu}}^{(N_P)} \boldsymbol{y}^{\boldsymbol{\nu}} \quad \text{with} \quad \left|\hat{Q}_{\boldsymbol{\nu}}^{(N_P)}\right| \leq \frac{\|G\|_{\mathbb{V}} \|q\|_{\mathbb{V}^\star}}{\delta} \boldsymbol{\rho}^{-\boldsymbol{\nu}}.$$

Comparing the functions in (2.2) with the representation from (3.8), we see that we need weights $w_{\boldsymbol{\nu}}^{(a)}$ such that

$$\sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \frac{\boldsymbol{\nu}!^2}{w_{\boldsymbol{\nu}}^{(a)}} \left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q^{(N_P)}) \right\rangle_{\mathbb{V}} < \infty.$$

For the bound on the coefficients given in (3.9). i.e. $\left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q) \right\rangle_{\mathbb{V}} \leq \frac{\|G\|_{\mathbb{V}^\star} \left\|q^{(N_P)}\right\|_{\mathbb{V}^\star}}{\delta} \boldsymbol{\rho}^{-\boldsymbol{\nu}}$, we get

$$\begin{aligned} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \frac{\boldsymbol{\nu}!^2}{w_{\boldsymbol{\nu}}^{(a)}} \left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q^{(N_P)}) \right\rangle_{\mathbb{V}} &\leq \frac{\|G\|_{\mathbb{V}^\star} \left\|q^{(N_P)}\right\|_{\mathbb{V}^\star}}{\delta} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \frac{\boldsymbol{\nu}!^2}{w_{\boldsymbol{\nu}}^{(a)}} \boldsymbol{\rho}^{-\boldsymbol{\nu}} \\ &\leq \frac{\|G\|_{\mathbb{V}^\star} \left\|q^{(N_P)}\right\|_{\mathbb{V}^\star}}{\delta} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{N_P}} \left(w_{\boldsymbol{\nu}}^{(a)}\right)^{-1} \prod_{j=1}^{N_P} \nu_j!^2 \rho_j^{-\nu_j}. \end{aligned}$$

We now make an ansatz with $\boldsymbol{c} < 1$ component wise for the weights

$$w_{\boldsymbol{\nu}}^{(a)} = \prod_{j=1}^{N_P} w_{\boldsymbol{\nu}_j}^{(a)} = \prod_{j=1}^{N_P} c_j \boldsymbol{\nu}_j!^2 = \boldsymbol{c}^{\boldsymbol{\nu}} \boldsymbol{\nu}!^2. \tag{3.11}$$

We then see that for the choice (3.11) we get

$$\sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{\boldsymbol{\nu}!^2}{w_{\boldsymbol{\nu}}^{(a)}} \left\langle u_{\boldsymbol{\nu}}^{(N_P)}, \mathcal{R}(q)\right\rangle_{\mathbb{V}} \leq \frac{\|G\|_{\mathbb{V}^\star}\left\|q^{(N_P)}\right\|_{\mathbb{V}^\star}}{\delta} \prod_{j=1}^{N_P}\sum_{\nu_j\in\mathbb{N}} \left(c_j\rho_j\right)^{-\nu_j} < \infty,$$

for $c_j\rho_j > 1$. Hence we have an admissibility condition for the constant given as

$$c_j \in (\rho_j^{-1}, 1) \quad \text{for all} \quad 1 \leq j \leq N_P. \tag{3.12}$$

This is possible since we have $\rho_j > 1$ for all $1 \leq j \leq N_P$. Hence, we can identify the reproducing kernel Hilbert space $\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)$ as a special case of (2.2) where the inner product has the weights $w_{\boldsymbol{\nu}}^{(a)} := \boldsymbol{c}^{\boldsymbol{\nu}}\boldsymbol{\nu}!^2$. □

The reproducing kernel for $\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)$ is explicitly given by

$$K_a\left(\boldsymbol{x},\boldsymbol{y}\right) := \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{\boldsymbol{c}^{\boldsymbol{\nu}}\boldsymbol{\nu}!^2}{\boldsymbol{\nu}!^2}\boldsymbol{x}^{\boldsymbol{\nu}}\boldsymbol{y}^{\boldsymbol{\nu}} = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \boldsymbol{c}^{\boldsymbol{\nu}}\boldsymbol{x}^{\boldsymbol{\nu}}\boldsymbol{y}^{\boldsymbol{\nu}} = \prod_{k=1}^{N_P}\sum_{\nu_k\in\mathbb{N}_0} c_k^{\nu_k}x_k^{\nu_k}y_k^{\nu_k} = \prod_{k=1}^{N_P}\frac{1}{1-c_kx_ky_k}.$$

Using the notation $S_c(x,y) := \frac{1}{1-cxy}$ for the univariate Szegö kernel with paramter $c < 1$, we get

$$K_a\left(\boldsymbol{x},\boldsymbol{y}\right) = \otimes_{k=1}^{N_P} S_{c_k}(x_k,y_k). \tag{3.13}$$

Hence the reproducing kernel is a tensor product of Szegö kernels with parameters c_k. These kernels can already be found in [18]. There it is also mentioned that for a suitable normalization, i.e., $\tilde{S}_c(x,y) = c^{-1}S_c(x,y)$, we can reproduce polynomials for $c \to \infty$. This might explain the fact that polynomial approximation methods work so well in these function spaces.

4. Geometric structure. In this section, we show that the support of the map $Q^{(N_P)}$ is approximately given by a geometric substructure of the parameter space. The important ingredient is a coupling of the dimension-wise decay and the numerical discretization process which is used to solve the partial differential equation for a given parameter $\tilde{\boldsymbol{y}} \in R_{N_P}$. The usual procedure in conforming methods is to fix a finite dimensional subspace $\mathbb{V}_{N_\mathcal{D}} = \operatorname{span}\{\psi_1,\ldots,\psi_{N_\mathcal{D}}\} \subset \mathbb{V}$ and to consider the spatially discretized version of (3.1), i.e., the equation (1.3). The weak form of this equation reads: Find $u^{(N_P,N_\mathcal{D})}\left(\tilde{\boldsymbol{y}},\cdot\right) \in \mathbb{V}_{N_\mathcal{D}} \subset \mathbb{V}$ such that

$$\int_\mathcal{D} a^{(N_P)}\left(\tilde{\boldsymbol{y}},\boldsymbol{x}\right)\nabla u^{(N_P,N_\mathcal{D})}\left(\tilde{\boldsymbol{y}},\boldsymbol{x}\right)\nabla v^{(N_\mathcal{D})}\left(\boldsymbol{x}\right)d\boldsymbol{x} = \int_\mathcal{D} G\left(\boldsymbol{x}\right)v^{(N_\mathcal{D})}\left(\boldsymbol{x}\right)d\boldsymbol{x} \tag{4.1}$$

for all $v^{(N_\mathcal{D})} \in \mathbb{V}_{N_\mathcal{D}} \subset \mathbb{V}$. Without loss of generality we can assume that the set $\{\psi_1,\ldots,\psi_{N_\mathcal{D}}\}$ is a basis in $\mathbb{V}_{N_\mathcal{D}}$ and use the usual isomorphism between finite dimensional spaces to get $\mathbb{V}_{N_\mathcal{D}} \cong \mathbb{R}^{N_\mathcal{D}}$ via

$$\mathcal{E} : \mathbb{V}_{N_\mathcal{D}} \to \mathbb{R}^{N_\mathcal{D}} \quad \text{with} \quad f^{(N_\mathcal{D})} = \sum_{j=1}^{N_\mathcal{D}} \hat{f}_j\psi_j \mapsto \left(\hat{f}^{(1)},\ldots,\hat{f}^{(N_\mathcal{D})}\right)^T \in \mathbb{R}^{N_\mathcal{D}}.$$

By the expansion (3.4), we obtain for the solution $u^{(N_P,N_\mathcal{D})}$ of (4.1) for fixed $\tilde{\boldsymbol{y}} \in R_{N_P}$ that $u^{(N_P,N_\mathcal{D})}(\tilde{\boldsymbol{y}},\cdot) \in \mathbb{V}_{N_\mathcal{D}}$, and hence

$$\begin{aligned}\mathcal{E}u^{(N_P,N_\mathcal{D})}(\tilde{\boldsymbol{y}},\cdot) &= \mathcal{E}\left(\sum_{k\in\mathbb{N}_0}\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\nu}|_1=k}} u_{\boldsymbol{\nu}}^{(N_P,N_\mathcal{D})}(\cdot)\,\tilde{\boldsymbol{y}}^{\boldsymbol{\nu}}\right) = \sum_{k\in\mathbb{N}_0}\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\nu}|_1=k}} \left(\hat{u}_{\boldsymbol{\nu}}^{(1)},\dots,\hat{u}_{\boldsymbol{\nu}}^{(N_\mathcal{D})}\right)^T \tilde{\boldsymbol{y}}^{\boldsymbol{\nu}} \\ &= \left(\sum_{k\in\mathbb{N}_0}\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\nu}|_1=k}} \hat{u}_{\boldsymbol{\nu}}^{(1)}\tilde{\boldsymbol{y}}^{\boldsymbol{\nu}},\dots,\sum_{k\in\mathbb{N}_0}\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\nu}|_1=k}} \hat{u}_{\boldsymbol{\nu}}^{(N_\mathcal{D})}\tilde{\boldsymbol{y}}^{\boldsymbol{\nu}}\right)^T. \end{aligned} \tag{4.2}$$

Due to the decay of the coefficients (3.9), it is natural to also truncate the polynomial series, i.e., to consider

$$\mathcal{E}u^{(N_P,N_\mathcal{D},N_T)}(\tilde{\boldsymbol{y}},\cdot) := \left(\sum_{k=0}^{N_T}\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\nu}|_1=k}} \hat{u}_{\boldsymbol{\nu}}^{(1)}\boldsymbol{y}^{\boldsymbol{\nu}},\dots,\sum_{k=0}^{N_T}\sum_{\substack{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}\\ |\boldsymbol{\nu}|_1=k}} \hat{u}_{\boldsymbol{\nu}}^{(N_\mathcal{D})}\boldsymbol{y}^{\boldsymbol{\nu}}\right)^T, \tag{4.3}$$

where N_T is the global maximal polynomial index. For the sake of simplicity, we do not consider individual polynomial degrees for each coordinate direction. We can consider $\tilde{\boldsymbol{y}} \mapsto \mathcal{E}u^{(N_P,N_\mathcal{D},N_T)}(\tilde{\boldsymbol{y}},\cdot) \in \mathbb{R}^{N_\mathcal{D}}$ as a vector of polynomials, since each component $\left(\mathcal{E}u^{(N_P,N_\mathcal{D},N_T)}(\tilde{\boldsymbol{y}},\cdot)\right)_t \in \mathbb{R}[y_1,\dots,y_{N_\mathcal{D}}]$, where $\mathbb{R}[x]$ denotes the formal ring of polynomials, for $1 \le t \le N_\mathcal{D}$, see also [16]. Now, following [33, p. 8], since a hypercube such as R_{N_P} is a simple polytope, we can invoke [4, Thm 1.2] to infer that R_{N_P} itself is a semi–algebraic set. Therefore, as a consequence of the Tarski–Seidenberg Theorem (cf. [12]) the image of a polynomial map (semi–algebraic map) of R_{N_P} is itself semi-algebraic [16, p. 242]. To be precise, we have

$$\mathcal{M} := \left(\left(\mathcal{E}u^{(N_P,N_\mathcal{D},N_T)}\right)_1(R_{N_P},\cdot),\dots,\left(\mathcal{E}u^{(N_P,N_\mathcal{D},N_T)}\right)_{N_\mathcal{D}}(R_{N_P},\cdot)\right)^T. \tag{4.4}$$

This semi-algebraic set approximates the support of the map $Q^{(N_P)}$. To make use of this information, we follow the approach of manifold learning. We propose to consider a proper discretization of the Laplace-Beltrami operator, treating the semi-algebraic set $\mathcal{M}$ as if it was a Riemannian manifold. Here, we describe a popular discretization of the Laplace-Beltami operator on a Riemannian manifold. We follow mainly [9, 11]. The basic assumption is that $\mathcal{M}$ is a compact $d_\mathcal{M}-$dimensional manifold which is isometrically embedded into the ambient Euclidean space $R_{N_P} \subset \mathbb{R}^{N_P}$ and the Riemannian metric $\rho_\mathcal{M}$ is assumed to be induced by the standard Euclidean metric of $\mathbb{R}^{N_P}$. In order to infer any information about the manifold $\mathcal{M}$ from discrete samples modeled as a point cloud[4]

$$\mathbb{Y}_{N_S} := \{\boldsymbol{y}_1,\dots,\boldsymbol{y}_N\} \subset R_{N_P} \subset \mathbb{R}^{N_P}, \tag{4.5}$$

[4]Note at this point that we could consider in principle a set $\mathbb{L} \neq \mathbb{Y}_{N_S}$. This is a common approach in semi-supervised learning applications. Here and in the following, we restrict ourselves for reasons of simplicity to use the sample points $\mathbb{Y}_{N_S}$ also as point cloud in the parameter space.

it is necessary to have some density property. This is formalized by the following quantity

$$h_{\mathbb{Y}_{N_S},\mathcal{M}} := \sup_{\boldsymbol{x}\in\mathcal{M}} \min_{\boldsymbol{z}\in\mathbb{Y}_{N_S}} \|\boldsymbol{x}-\boldsymbol{z}\|_2 .$$

If $h_{\mathbb{Y}_{N_S},\mathcal{M}} \to 0$, we say that the point cloud resolves the underlying manifold $\mathcal{M}$. A key observation from [9] is that the intrinsic distance is comparable to the ambient Euclidean distance, i.e.,

$$\rho_{\mathcal{M}}(\boldsymbol{x}_1,\boldsymbol{x}_2) = \|\boldsymbol{x}_1-\boldsymbol{x}_2\|_{2,\mathbb{R}^{N_P}} + \mathcal{O}\left(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^3\right)$$

for $\boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathcal{M}$. Following [9], the heuristic behind the discrete Laplace-Beltrami operator stems from the heat equation in the ambient Euclidean space. One has for an $f \in C^\infty(\mathbb{R}^{N_P})$

$$\begin{aligned}
&-\Delta_{\mathbb{R}^{N_P}} f(\boldsymbol{x}) = \partial_t \left(\int_{\mathbb{R}^{N_P}} f(\boldsymbol{z})(4\pi t)^{-\frac{N_P}{2}} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{4t}\right) d\boldsymbol{y}\right)\Big|_{t=0} \\
&= \lim_{t\to 0} t^{-1}(4\pi t)^{-\frac{N_P}{2}} \left(\int_{\mathbb{R}^{N_P}} f(\boldsymbol{y}) \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{4t}\right) d\boldsymbol{y} - f(\boldsymbol{x}) \int_{\mathbb{R}^{N_P}} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{4t}\right) d\boldsymbol{y}\right) \\
&\approx \frac{(4\pi\tilde{t})^{-\frac{N_P}{2}}}{\tilde{t}N_S} \left(\sum_{j=1}^{N_S} f(\boldsymbol{y}_j) \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}_j\|_2^2}{4t}\right) - f(\boldsymbol{x}) \sum_{j=1}^{N_S} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{y}_j\|_2^2}{4t}\right)\right) \qquad (4.6) \\
&=: -\frac{(4\pi\tilde{t})^{-\frac{N_P}{2}}}{\tilde{t}N_S} \mathcal{L}_{N_S}^{(\tilde{t})} f(\boldsymbol{x})
\end{aligned}$$

for a small enough time $\tilde{t} > 0$. In (4.6) a simple equal weight quadrature was employed. This heuristic argument can be made rigorous as outlined in [9, Thm. 1] in the sense of point-wise convergence. There, it is shown that if the points $\mathbb{Y}_{N_S} \sim \mathcal{U}(\mathcal{M})$ are drawn from a uniform distribution on $\mathcal{M}$ and the parameter $\tilde{t}$ is chosen as $\tilde{t} := \tilde{t}(N_S) = N_S^{-\frac{1}{d_{\mathcal{M}}+2+\epsilon}}$ for some $\epsilon > 0$, then

$$\Delta_{\mathcal{M}} f(\boldsymbol{x}) = \lim_{N_S\to\infty} \frac{(4\pi\tilde{t})^{-\frac{N_P}{2}}}{\tilde{t}N_S} \mathcal{L}_{N_S}^{(\tilde{t})} f(\boldsymbol{x}).$$

Here $d_{\mathcal{M}}$ denotes the intrinsic dimension of $\mathcal{M}$. By construction, we have $d_{\mathcal{M}} \le N_P$. The limit $N_S \to \infty$ for uniformly distributed points implies also $h_{\mathbb{Y}_{N_S},\mathcal{M}} \to 0$. It is important to see that the discrete Laplace operator $\mathcal{L}_{N_S}^{(\tilde{t})}$ depends only on the point values. We formalize this observation in the following way: We introduce a sampling operator

$$S_{\mathbb{Y}_{N_S}} : C(\mathbb{R}^{N_P}) \to \mathbb{R}^{N_S}, \quad f \mapsto (f(\boldsymbol{y}_1),\dots,f(\boldsymbol{y}_N))^T . \qquad (4.7)$$

Next, we can restate the action of the discrete Laplace operator by using the sampling operator. In particular, we use the identity

$$S_{\mathbb{Y}_{N_S}} \left(\mathcal{L}_{N_S}^{(\tilde{t})} f\right) = (\boldsymbol{D}-\boldsymbol{W})\, S_{\mathbb{Y}_{N_S}}(f)$$

with the matrices $\boldsymbol{D},\boldsymbol{W} \in \mathbb{R}^{N_S\times N_S}$ given by

$$(\boldsymbol{D})_{i,j} = \begin{cases} \sum_{k=1}^{N_S} \exp\left(-\frac{\|\boldsymbol{y}_i-\boldsymbol{y}_k\|_2^2}{4\tilde{t}}\right), & i=j \\ 0, & i\neq j \end{cases} \quad \text{and} \quad (\boldsymbol{W})_{i,j} = \exp\left(-\frac{\|\boldsymbol{y}_i-\boldsymbol{y}_j\|_2^2}{4\tilde{t}}\right).$$

Then,

$$S_{\mathbb{Y}_{N_S}}(f)\cdot S_{\mathbb{Y}_{N_S}}\left(\mathcal{L}_{N_S}^{(\tilde{t})}f\right)=S_{\mathbb{Y}_{N_S}}(f)\cdot(\boldsymbol{D}-\boldsymbol{W})\,S_{\mathbb{Y}_{N_S}}(f)$$
$$=\sum_{i,j=1}^{N_S}\exp\left(-\frac{\|\boldsymbol{y}_i-\boldsymbol{y}_j\|_2^2}{4\tilde{t}}\right)\left(f(\boldsymbol{y}_i)-f(\boldsymbol{y}_j)\right)^2.$$

We can use this expression to define the geometric regularization term in (1.6) as

$$\|f\|_{\mathcal{M}}^2:=\frac{(4\pi\tilde{t})^{-N_P}}{\tilde{t}^2N_S^2}S_{\mathbb{Y}_{N_S}}(f)\cdot(\boldsymbol{D}-\boldsymbol{W})\,S_{\mathbb{Y}_{N_S}}(f). \tag{4.8}$$

5. Manifold learning. Now we are in the position to consider the optimization problem (1.6). To this end, we briefly recall the setting for the reconstruction of the function $Q^{(N_P)}:R_{N_P}\to\mathbb{R}$. The input for the learning task are the perturbed samples

$$Q_k^{(N_P,N_{\mathcal{D}})}:=q^{(N_P)}\left(u_k^{(N_P,N_{\mathcal{D}})}\right)\approx Q_k^{(N_P)}=q^{(N_P)}\left(u^{(N_P)}\left(\boldsymbol{y}_k,\cdot\right)\right).$$

We note that the quantity

$$\epsilon_k:=Q_k^{(N_P,N_{\mathcal{D}})}-Q^{(N_P)}\left(\boldsymbol{y}_k\right)=q^{(N_P)}\left(u_k^{(N_P,N_{\mathcal{D}})}\right)-q^{(N_P)}\left(u^{(N_P)}(\boldsymbol{y}_k,\cdot)\right)$$
$$=q^{(N_P)}\left(u_k^{(N_P,N_{\mathcal{D}})}-u^{(N_P)}(\boldsymbol{y}_k,\cdot)\right)$$

is bounded by the numerical error which occurs in the solution of the equation (4.1). We have

$$|\epsilon_k|\le\left\|q^{(N_P)}\right\|_{\mathbb{V}^\star}\left\|u_k^{(N_P,N_{\mathcal{D}})}-u^{(N_P)}(\cdot,\boldsymbol{y}_k)\right\|_{\mathbb{V}}.$$

The basic variational problem over the reproducing kernel Hilbert space $\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)$ is given in (1.6) and we recall its definition here. We define

$$J_{Q_k^{(N_P,N_{\mathcal{D}})},\lambda_K,\lambda_{\mathcal{M}}}:\mathcal{H}_{K_a\left((-1,1)^{N_P}\right)}\to\mathbb{R}\quad\text{where} \tag{5.1}$$
$$J_{Q_k^{(N_P,N_{\mathcal{D}})},\lambda_K,\lambda_{\mathcal{M}}}(f):=\sum_{k=1}^{N_S}V\left(Q_k^{(N_P,N_{\mathcal{D}})},f\left(\boldsymbol{y}_k\right)\right)+\lambda_K\|f\|^2_{\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)}+\lambda_{\mathcal{M}}\|f\|^2_{\mathcal{M}}$$

and set

$$\tilde{Q}^{(N_S;\lambda_K,\lambda_{\mathcal{M}};\epsilon_{N_{\mathcal{D}}};N_P)}:=\arg\min_{f\in\mathcal{H}_{K_a}(-1,1)^{N_P}}J_{Q_k^{(N_P,N_{\mathcal{D}})},\lambda_K,\lambda_{\mathcal{M}}}(f). \tag{5.2}$$

In this situation, there is a numerically useful observation, namely a *representer theorem* (see [10, Thms 1&2]). It states that, although the minimization (5.2)

$$\tilde{Q}^{N_S;\lambda_K,\lambda_{\mathcal{M}};\epsilon_{N_{\mathcal{D}}};N_P}:=\arg\min_{f\in\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)}J_{Q_k^{(N_P,N_{\mathcal{D}})},\lambda_K,\lambda_{\mathcal{M}}}(f)$$
$$=\arg\min_{f\in\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)}\sum_{k=1}^{N_S}V\left(Q_k^{(N_P,N_{\mathcal{D}})},f\left(\boldsymbol{y}_k\right)\right)$$
$$+\lambda_K\|f\|^2_{\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)}+\lambda_{\mathcal{M}}\frac{(4\pi\tilde{t})^{-N_P}}{\tilde{t}^2N_S^2}S_{\mathbb{Y}_{N_S}}(f)\cdot(\boldsymbol{D}-\boldsymbol{W})\,S_{\mathbb{Y}_{N_S}}(f)$$

with the functional $J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}$ defined in (5.1) is done over the infinite dimensional Hilbert space $\mathcal{H}_{K_a}\left((-1,1)^{N_P}\right)$, the solution resides in a *finite dimensional* space. To be precise, [10, Thm. 2] yields that

$$\tilde{Q}^{(N_S;\lambda_K,\lambda_\mathcal{M};\epsilon_{N_\mathcal{D}};N_P)} \in \mathbb{V}_{\mathbb{Y}_{N_S}} := \operatorname{span}\left\{K(\cdot,\boldsymbol{y}_j) \ : \ \boldsymbol{y}_j \in \mathbb{Y}_{N_S}\right\} \subset \mathcal{H}_{K_a}\left((-1,1)^{N_P}\right). \tag{5.3}$$

Numerically, this has the advantage that the computation reduces to a finite dimensional optimization problem. In the following, we will restrict ourselves to a specific loss function, namely Vapnik's ϵ insensitive loss function (see [30])

$$V_{\mathrm{Va}}^{(\epsilon)}(x,z) := \begin{cases} |x-z|-\epsilon & \text{if } |x-z| \geq \epsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

Note that $\epsilon > 0$ plays the rôle of an accuracy level, i.e., the parameter ϵ limits the numerical accuracy for the solution of the equation (1.3) which produces the sampling values $Q_k^{(N_P,N_\mathcal{D})}$.

Basically following [31, 25], we derive from the optimization problem (5.2) two important estimates for the solution $\tilde{Q} := \tilde{Q}^{(N_S;\lambda_K,\lambda_\mathcal{M};\epsilon_{N_\mathcal{D}};N_P)}$ (c.f. (5.2)). First, recall that the data is corrupted by additive errors, i.e., $Q_k^{(N_P,N_\mathcal{D})} \approx Q_k^{(N_P)} := q^{(N_P)}(u^{(N_P)}(\boldsymbol{y}_k,\cdot))$. Then

$$\begin{aligned}
&\left|Q_k^{(N_P,N_\mathcal{D})} - \tilde{Q}(\boldsymbol{y}_k)\right| \leq V_{\mathrm{Va}}^{\epsilon}(Q_k^{(N_P,N_\mathcal{D})},\tilde{Q}(\boldsymbol{y}_k)) + \epsilon \leq J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(\tilde{Q}) + \epsilon \\
&\leq J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(Q^{(N_P)}) + \epsilon \\
&= \sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})}, Q^{(N_P)}(\boldsymbol{y}_k)\right) + \lambda_K \left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \lambda_\mathcal{M}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}} + \epsilon.
\end{aligned}$$

Here, we have again that the sum can be made small by increasing the spatial accuracy in the numerical solver. Following [25], we can further estimate the consistency error. We get for the point wise error at the sampling nodes the bound

$$\begin{aligned}
&\left|Q^{(N_P)}(\cdot,\boldsymbol{y}_k) - \tilde{Q}(\boldsymbol{y}_k)\right| \leq \left|Q^{(N_P,N_\mathcal{D})}(\cdot,\boldsymbol{y}_k) - \tilde{Q}(\boldsymbol{y}_k)\right| + \left|Q^{(N_P)}(\cdot,\boldsymbol{y}_k) - Q^{(N_P,N_\mathcal{D})}(\boldsymbol{y}_k)\right| \\
&\leq 2\epsilon + V_{\mathrm{Va}}^{(\epsilon)}\left(Q^{(N_P,N_\mathcal{D})}(\cdot,\boldsymbol{y}_k), Q^{(N_P)}(\boldsymbol{y}_k)\right) + V_{\mathrm{Va}}^{(\epsilon)}\left(Q^{(N_P,N_\mathcal{D})}(\cdot,\boldsymbol{y}_k), \tilde{Q}(\boldsymbol{y}_k)\right) \\
&\leq 2\epsilon + V_{\mathrm{Va}}^{(\epsilon)}\left(Q^{(N_P,N_\mathcal{D})}(\cdot,\boldsymbol{y}_k), Q^{(N_P)}(\boldsymbol{y}_k)\right) + \sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})}, Q^{(N_P)}(\boldsymbol{y}_k)\right) \\
&+ \lambda_K \left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \lambda_\mathcal{M}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}.
\end{aligned} \tag{5.5}$$

Furthermore, we directly obtain

$$\begin{aligned}
&\lambda_K \left\|\tilde{Q}_{N_S;\lambda_K,\lambda_\mathcal{M};\epsilon_{N_\mathcal{D}};N_P}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} \leq J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(\tilde{Q}) \leq J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(Q^{(N_P)}) \\
&= \sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})}, Q^{(N_P)}(\boldsymbol{y}_k)\right) + \lambda_K \left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \lambda_\mathcal{M}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}},
\end{aligned}$$

from which we infer

$$\left\|\tilde{Q}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} \leq \left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \frac{1}{\lambda_K}\sum_{k=1}^{N_S} V^{(\epsilon)}_{\mathrm{Va}}\left(Q_k^{(N_P,N_\mathcal{D})}, Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right) + \frac{\lambda_\mathcal{M}}{\lambda_K}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}. \tag{5.6}$$

We collect these results in the following lemma using the notation from (5.2).

LEMMA 5.1. *Denote by* $\tilde{Q} := \arg\min_{f\in\mathcal{H}_{K_a((-1,1)^{N_P})}} J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(f)$. *Then we have the following bounds*

$$\left|Q^{(N_P)}\left(\cdot,\boldsymbol{y}_k\right) - \tilde{Q}(\boldsymbol{y}_k)\right| \leq 2\epsilon + V^{(\epsilon)}_{Va}\left(Q^{(N_P,N_\mathcal{D})}(\cdot,\boldsymbol{y}_k), Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right) + J_{Q_k^{(N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(Q^{(N_P)})$$
$$\left\|\tilde{Q}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} \leq \left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(Q^{(N_P)}).$$

Proof. The central observation is that $Q^{(N_P)} \in \mathcal{H}_{K_a((-1,1)^{N_P})}$ due to Lemma 3.3. ∎

Note that both terms get simpler for large $N_\mathcal{D}$, i.e., if the discretization error is small.

COROLLARY 5.2. *Under the assumptions of Lemma 5.1 and if* $N_\mathcal{D} \geq N_\mathcal{D}(\epsilon)$ *is large enough to guarantee* $\left|Q_k^{(N_P,N_\mathcal{D})} - Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right| \leq \epsilon$ *for* $1 \leq k \leq N_S$, *then we have*

$$\left|Q^{(N_P)}\left(\cdot,\boldsymbol{y}_k\right) - \tilde{Q}(\boldsymbol{y}_k)\right| \leq 2\epsilon + \lambda_K\left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \lambda_\mathcal{M}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}$$
$$\left\|\tilde{Q}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} \leq \left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \frac{\lambda_\mathcal{M}}{\lambda_K}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}.$$

We are now in the position to derive an error estimate by applying (2.9) to the error $\tilde{Q} - Q^{(N_P)}$.

THEOREM 5.3. *There exist* $h_0(R_{N_P}) > 0$ *and* $a > 0$ *such that for all point sets* $\mathbb{Y}_{N_S} \subset R_{N_P}$ *with* $h_{\mathbb{Y}_{N_S},R_{N_P}} < h_0(R_{N_P})$ *and all* $u \in \mathcal{H}_{K_a}((-1,1)^{N_P})$ *with* $w_{\boldsymbol{\nu}}^{(a)} = \boldsymbol{c}^{\boldsymbol{\nu}}\boldsymbol{\nu}!^2$ *and* $\boldsymbol{c}$ *satisfying the admissibility condition* (3.12), *we have for* $\tilde{Q} := \arg\min_{f\in\mathcal{H}_{K_a}((-1,1)^{N_P})} J_{Q_k^{(N_P,N_\mathcal{D})},\lambda_K,\lambda_\mathcal{M}}(f)$ *the error estimate*

$$\left\|\tilde{Q} - Q^{(N_P)}\right\|_{L^\infty(R_{N_P})} \leq \exp\left(-\frac{C_1}{\sqrt{h_{\mathbb{Y}_S,R_{N_P}}}}\right) \cdot \tag{5.7}$$
$$\cdot\left(\left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \frac{1}{\lambda_K}\sum_{k=1}^{N_S} V^{(\epsilon)}_{Va}\left(Q_k^{(N_P,N_\mathcal{D})}, Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right) + \frac{\lambda_\mathcal{M}}{\lambda_K}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}\right)^{\frac{1}{2}}$$
$$+ c\left(2\epsilon + 2\sum_{k=1}^{N_S} V^{(\epsilon)}_{Va}\left(Q_k^{(N_P,N_\mathcal{D})}, Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right) + \lambda_K\left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})} + \lambda_\mathcal{M}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}\right)^{\frac{1}{2}},$$

where $c > 0$ *is a constant.*

Proof. We can use the bound (5.5), i.e.

$$\begin{aligned}
&\left\|S_{\mathbb{Y}_{N_S}}(\tilde{Q}-Q^{(N_P)})\right\|_{\ell^\infty\left(\mathbb{Y}_{N_S}\right)}\\
&\leq 2\epsilon+\max_{k=1,\dots,N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q^{(N_P,N_\mathcal{D})}(\cdot,\boldsymbol{y}_k),Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right)+\sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})},Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right)\\
&+\lambda_K\left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})}+\lambda_\mathcal{M}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}\\
&\leq 2\epsilon+2\sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})},Q\left(\boldsymbol{y}_k\right)\right)+\lambda_K\left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})}+\lambda_\mathcal{M}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}.
\end{aligned}$$

Furthermore, we have the bound (5.6) which results here in

$$\begin{aligned}
\left\|\tilde{Q}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})}&\leq\left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})}+\frac{1}{\lambda_K}\sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})},Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right)\\
&+\frac{\lambda_\mathcal{M}}{\lambda_K}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}.
\end{aligned}$$

From the last estimate, we directly infer

$$\begin{aligned}
\left\|\tilde{Q}-Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})}&\leq 2\left\|Q^{(N_P)}\right\|^2_{\mathcal{H}_{K_a}((-1,1)^{N_P})}+\frac{1}{\lambda_K}\sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})},Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right)\\
&+\frac{\lambda_\mathcal{M}}{\lambda_K}\left\|Q^{(N_P)}\right\|^2_{\mathcal{M}}.
\end{aligned}$$

Inserting these bounds into (5.7) concludes the argument. ∎

For large $N_\mathcal{D}\geq N_\mathcal{D}(\epsilon)$, i.e., for small discretization errors, the error bounds simplify dramatically.

COROLLARY 5.4. *If $N_\mathcal{D}\geq N_\mathcal{D}(\epsilon)$ is large enough to guarantee $\left|Q_k^{(N_P,N_\mathcal{D})}-Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right|\leq\epsilon$ for $1\leq k\leq N_S$ and under the assumptions of Theorem 5.3 we get for the choice*

$$\sqrt{\lambda_K}\sim\exp\left(-\frac{C_1}{\sqrt{h_{\mathbb{Y}_S,R_{N_P}}}}\right)\sim\sqrt{\lambda_\mathcal{M}} \tag{5.8}$$

the estimate

$$\left\|\tilde{Q}-Q^{(N_P)}\right\|_{L^\infty(R_{N_P})}\leq\exp\left(-\frac{C_1}{\sqrt{h_{\mathbb{Y}_S,R_{N_P}}}}\right)\left(\left\|Q^{(N_P)}\right\|_{\mathcal{H}_{K_a}((-1,1)^{N_P})}+\left\|Q^{(N_P)}\right\|_{\mathcal{M}}\right)+c\sqrt{\epsilon}. \tag{5.9}$$

Proof. From the estimate $\left|Q_k^{(N_P,N_\mathcal{D})}-Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right|\leq\epsilon$, we directly infer

$$\frac{1}{\lambda_K}\sum_{k=1}^{N_S} V_{\mathrm{Va}}^{(\epsilon)}\left(Q_k^{(N_P,N_\mathcal{D})},Q^{(N_P)}\left(\boldsymbol{y}_k\right)\right)=0$$

and from the choice for the regularization parameters $\lambda_K \sim \lambda_{\mathcal{M}}$, we get $\frac{\lambda_{\mathcal{M}}}{\lambda_K} \left\| Q^{(N_P)} \right\|_{\mathcal{M}}^2 \sim 1$. This concludes the proof. □

This result directly allows us to couple the numerical error ϵ and the number of sampling points N_S in order to achieve a prescribed accuracy.

6. Concluding remarks and future directions. We presented a link between reproducing kernel Hilbert spaces and the regularity theory of parametric elliptic partial differential equations. In general, one needs to consider non-standard power series kernels. These kernels contain by construction the dimension-decay properties of the problem under consideration. Such kernels and their corresponding Hilbert spaces open the door to any kernel based method from the rich field of machine learning. Here, regularized approaches are important because in a non-intrusive method the data we have at hand are necessarily corrupted by a (numerical) error. This error stems from the discretization procedure of the partial differential equation. Usually these errors are well-known and controllable by the user. Hence, the question of an optimal cost-benefit strategy arises naturally. For this purpose sampling inequalities are well-suited tools because they allow for deterministic a priori error estimates which contain the numerical error in an explicit way. Hence an a priori coupling between the evaluation error which contains the discretized solution of a partial differential equation and the number of parameters for which we evaluate the quantity of interest is possible in such a situation. The drawback of the sampling inequalities is the fact that they measure the convergence in terms of the isotropic fill-distance. This is however only a realistic discretization parameter if one considers the situation of a few almost equally important coordinates in the parameter space and a rapid decay of the importance of the later coordinates. This assumption is valid for problems with a short correlation length as they often appear in the literature, but this is not true in general applications. Our approach needs then to be further modified and adopted for such situations.
Finally, we also discussed intrinsic low-dimensional manifold-like structures. This is, in our opinion, an important topic because all numerically feasible methods in high dimensions need an intrinsic low-dimensional object to overcome the curse of dimensionality in the first place. Any information on the geometric nature of the underlying structure will enable us to develop individually tailored algorithms. Here, we expect the recent developments for data analysis in a non-Riemannian manifold setting to have great impact.

Acknowledgement. This work was supported in part by the Collaborative Research Centre 1060 of the Deutsche Forschungsgemeinschaft. We thank Barbara Zwicknagl for helpful discussions and various improvements of the manuscript. Furthermore, we thank Jens Oettershagen for proofreading.

Appendix. We want to estimate the quantity $|\boldsymbol{\nu}|!$ for $\boldsymbol{\nu} \in \mathbb{N}^{N_P}$ in terms of products. We make use of Euler's integral representation of the factorial and Hoelder's inequality with $\boldsymbol{\beta} \in \mathbb{N}^{N_P}$ satisfying $\sum_{j=1}^{N_P} \frac{1}{\boldsymbol{\beta}_j} = 1$:

$$
\begin{aligned}
|\boldsymbol{\nu}|_1! &= \int_0^1 \left(-\ln(s)\right)^{|\boldsymbol{\nu}|_1} ds = \int_0^1 |\ln(s)|^{|\boldsymbol{\nu}|_1} ds = \left\| \prod_{j=1}^{N_P} (\ln(\cdot))^{\nu_j} \right\|_{L^1([0,1])} \\
&\leq \prod_{j=1}^{N_P} \left\| (\ln(\cdot))^{\nu_j} \right\|_{L^{\beta_j}([0,1])} = \prod_{j=1}^{N_P} \left(\int_0^1 \left(-\ln(s)\right)^{\beta_j \nu_j} ds \right)^{\frac{1}{\beta_j}} = \prod_{j=1}^{N_P} \left((\beta_j \nu_j)! \right)^{\frac{1}{\beta_j}} .
\end{aligned}
$$

Making the special isotropic choice $\beta_j = N_P$ leads to $|\boldsymbol{\nu}|! \leq \prod_{j=1}^{N_P} ((N_P \boldsymbol{\nu}_j)!)^{\frac{1}{N_P}}$. For a single factor, we use Stirling's approximation formula (where we use the asymptotic expressions here in order to avoid more lengthy formulas) to get

$$\begin{aligned}
((N_P\nu_j)!)^{\frac{1}{N_P}} &\sim \left(\sqrt{2\pi N_P \nu_j}\left(\frac{N_P\nu_j}{e}\right)^{N_P\nu_j}\right)^{\frac{1}{N_P}} = (2\pi N_P\nu_j)^{\frac{1}{2N_P}}\left(\frac{N_P\nu_j}{e}\right)^{\nu_j} \\
&= (2\pi)^{\frac{1-N_P}{2N_P}} N_P^{\frac{1}{2N_P}} \nu_j^{\frac{1-N_P}{2N_P}} \left(\frac{N_P}{e}\right)^{\nu_j} \sqrt{2\pi\nu_j}\left(\frac{\nu_j}{e}\right)^{\nu_j} \\
&\sim (2\pi)^{\frac{1-N_P}{2N_P}} N_P^{\frac{1}{2N_P}} \nu_j^{\frac{1-N_P}{2N_P}} \left(\frac{N_P}{e}\right)^{\nu_j} \nu_j!
\end{aligned}$$

Finally, we obtain

$$\begin{aligned}
|\boldsymbol{\nu}|_1! &\sim \prod_{j=1}^{N_P}\left((2\pi)^{\frac{1-N_P}{2N_P}} N_P^{\frac{1}{2N_P}} \nu_j^{\frac{1-N_P}{2N_P}} \left(\frac{N_P}{e}\right)^{\nu_j} \nu_j!\right) \\
&\lesssim (2\pi)^{\frac{1-N_P}{2}} N_P^{\frac{1}{2}} \left(\frac{N_P}{e}\right)^{|\boldsymbol{\nu}|_1} \boldsymbol{\nu}!,
\end{aligned}$$

where we used the fact that $N_P \geq 1$ to get rid of the polynomial term. Hence, we can modify the $\boldsymbol{b}$ in (3.6) to get a product structure decay estimate. Precisely, we have

$$\begin{aligned}
\left\| u_{\boldsymbol{\nu}}^{(N_P)} \right\|_{\mathbb{V}} &\leq \frac{\|G\|_{\mathbb{V}^\star}}{r} \frac{|\boldsymbol{\nu}|_1!}{\boldsymbol{\nu}!} \boldsymbol{b}^{\boldsymbol{\nu}} \lesssim (2\pi)^{\frac{1-N_P}{2}} N_P^{\frac{1}{2}} \left(\frac{N_P}{e}\right)^{|\boldsymbol{\nu}|_1} \boldsymbol{b}^{\boldsymbol{\nu}} \\
&\leq \frac{\|G\|_{\mathbb{V}^\star}}{r} (2\pi)^{\frac{1-N_P}{2}} N_P^{\frac{1}{2}} \tilde{\boldsymbol{b}}^{\boldsymbol{\nu}}
\end{aligned}$$

with the sequence $\left(\tilde{b}_k\right)_k$ defined by $\tilde{b}_k = \frac{N_P}{e} b_k$ and thus, with (3.7), we obtain

$$\tilde{b}_k := \frac{N_P}{e}\left(\inf_{\boldsymbol{x}\in\mathcal{D}} \int_{R_{N_P}} a(\boldsymbol{y},\boldsymbol{x})\, d\boldsymbol{y}\right)^{-1} \|\phi_k\|_{L^\infty(\mathcal{D})}. \tag{6.1}$$

Altogether, we have seen that we can slightly change the sequence b_k to $\tilde{b}_k$ to get a product structure decay bound. Note here that we only work in the finite dimensional case ($N_P < \infty$). This allows us to have dimension dependent factors in the estimates which are hidden in the $\sim$ notation and in the $\frac{N_P}{e}$ factor. In order to move to the infinite dimensional case, one has additionally to assume an appropriate dimension dependent decay of the sequence b_k. A typical a priori way to ensure this is a δ-admissibility condition as (3.3), see [13, Lemma 2.4]. Moreover, an anisotropic dimension dependent way to choose β_k and (6.1) open the way for various a posteriori techniques to also obtain estimates in the limit $N_P \to \infty$ with still sub-exponential convergence rates and accordingly modified constants.

REFERENCES

[1] R. Arcangéli, M. López di Silanes, and 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.

[2] R. Arcangéli, M. López di Silanes, and J. Torrens. Estimates for functions in Sobolev spaces defined on unbounded domains. *Journal of Approximation Theory*, 161(1):198 – 212, 2009.

[3] R. Arcangéli, M. López di Silanes, and J. Torrens. Extension of sampling inequalities to Sobolev semi-norms of fractional order and derivative data. *Numerische Mathematik*, 121:587–608, 2012.

[4] G. Averkov and M. Henk. Representing simple d-dimensional polytopes by d polynomials. *Mathematical Programming*, 126(2):203–230, 2011.

[5] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

[6] I. Babuška, R. Tempone, and G. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.

[7] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

[8] J. Beck, R. Tempone, F. Nobile, and L. Tamellini. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Mathematical Models and Methods in Applied Sciences*, 22(09):1250023, 2012.

[9] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289 – 1308, 2008.

[10] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.*, 7:2399–2434, December 2006.

[11] M. Belkin, J. Sun, and Y. Wang. Constructing Laplace operator from point clouds in $\mathbb{R}^d$. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, pages 1031–1040, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[12] J. Bochnak, M. Coste, and M. Roy. *Real Algebraic Geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge, A Series of Modern Surveys in Mathematics. Springer, 1998.

[13] A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. *Analysis and Applications*, 9(1):11–47, 2010.

[14] I. Cialenco, G. Fasshauer, and Q. Ye. Approximation of stochastic partial differential equations by a kernel-based collocation method. *International Journal of Computer Mathematics*, 89(18):2543–2561, 2012.

[15] K. Dalbey Efficient and Robust Gradient Enhanced Kriging Emulators *SANDIA REPORT SAND2013-7022*, 2013

[16] J. Fernando and J. Gamboa. Polynomial images of $\mathbb{R}^n$. *Journal of Pure and Applied Algebra*, 179(3):241 – 254, 2003.

[17] M. Griebel, C. Rieger, and B. Zwicknagl. Multiscale approximation and reproducing kernel Hilbert space methods. *SIAM Journal on Numerical Analysis*, 53(2):852–873,2015.

[18] F. Larkin. Optimal approximation in Hilbert spaces with reproducing kernel functions. *Mathematics of Computation*, 24(8):911–921, 1970.

[19] A. Litvinenko, H. Matthies, and T. El-Moselhy. Sampling and low-rank tensor approximation of the response surface. In J. Dick, F. Kuo, G. Peters, and I. Sloan, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2012*, volume 65 of *Springer Proceedings in Mathematics & Statistics*, pages 535–551. Springer Berlin Heidelberg, 2013.

[20] W. Madych. An estimate for multivariate approximation II. *J. Approx. Theory*, 142:116–128, 2006.

[21] H. Matthies, A. Litvinenko, O. Pajonk, B. Rosić, and E. Zander. Parametric and uncertainty computations with tensor product representations. In A. Dienstfrey and R. Boisvert, editors, *Uncertainty Quantification in Scientific Computing*, volume 377 of *IFIP Advances in Information and Communication Technology*, pages 139–150. Springer Berlin Heidelberg, 2012.

[22] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Approximation of quantities of interest in stochastic PDEs by the random discrete L^2 projection on polynomial spaces. *SIAM Journal on Scientific Computing*, 35(3):A1440–A1460, 2013.

[23] F. Narcowich, J. Ward, and H. Wendland. Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. *Math. Comp.*, 74:743–763, 2005.

[24] C. Rieger, R. Schaback, and B. Zwicknagl. Sampling and stability. In *Mathematical Methods for Curves and Surfaces*, volume 5862 of *Lecture Notes in Computer Science*, pages 347–369. Springer, New York, 2010.

[25] C. Rieger and B. Zwicknagl. Deterministic error analysis of support vector regression and related regularized kernel methods. *Journal of Machine Learning Research*, 10:2115–2132, 2009.

[26] C. Rieger and B. Zwicknagl. Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. *Adv. Comput. Math.*, 32(1):103–129, 2010.
[27] C. Rieger and B. Zwicknagl. Improved exponential convergence rates by oversampling near the boundary. *Constr. Approx.*, 39(2):323–341, 2014.
[28] H. Tran, C. Webster, and G. Zhang. Analysis of quasi-optimal polynomial approximations for parameterized PDEs with deterministic and stochastic coefficients. 2015. ORNL TM-2014/468
[29] F. Utreras Convergence rates for multivariate smoothing spline functions, *Journal of Approximation Theory* 52(1),1–27, 1988
[30] V. Vapnik *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
[31] H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numerische Mathematik*, 101(4):729–748, 2005.
[32] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2004. Cambridge Books Online.
[33] G. Ziegler. *Lectures on polytopes*. Springer-Verlag, New York, 1995.
[34] B. Zwicknagl. Power series kernels. *Constructive Approximation*, 29(1):61–84, 2009.
[35] B. Zwicknagl and R. Schaback. Interpolation and approximation in Taylor spaces. *Journal of Approximation Theory*, 171(0):65 – 83, 2013.