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

R. Kempf, H. Wendland and C. Rieger

Kernel-based Reconstructions for Parametric PDEs

INS Preprint No. 1804

March 2018

Kernel-based Reconstructions for Parametric PDEs

Rüdiger Kempf, Holger Wendland and Christian Rieger

Abstract In uncertainty quantification, an unknown quantity has to be reconstructed which depends typically on the solution of a partial differential equation. This partial differential equation itself may depend on parameters, some of them may be deterministic and some are random. To approximate the unknown quantity one therefore has to solve the partial differential equation (usually numerically) for several instances of the parameters and then reconstruct the quantity from these simulations. As the number of parameters may be large, this becomes a high-dimensional reconstruction problem.

We will address the topic of reconstructing such unknown quantities using kernel-based reconstruction methods on sparse grids. First, we will introduce into the topic, then explain the reconstruction process and finally provide new error estimates.

1 Introduction

In modern applied sciences dynamic processes are often modeled by partial differential equations, whereby coefficient functions, representing certain material parameters, and forcing terms serve as input. Often, these are obtained by certain measurements or experiments and therefore are prone to being either inaccurate or incomplete and

---

Rüdiger Kempf
Applied and Numerical Analysis, Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany e-mail: `Ruediger.Kempf@uni-bayreuth.de`

Holger Wendland
Applied and Numerical Analysis, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany e-mail: `Holger.Wendland@uni-bayreuth.de`

Christian Rieger
Institut für Numerische Simulation, Wegelstraße 5, Germany e-mail: `rieger@ins.uni-bonn.de`

consequently introduce an uncertainty to the model. For a general overview on the topic, see, for example, the recent books [8, 9, 10, 11].

To illustrate the general approach to such problems including randomness, we follow [5, 2, 1] and hence restrict ourselves to a Dirichlet-Poisson problem where the parametric diffusion coefficient is given by a function $a : R_{N_P} \times \mathcal{D} \to \mathbb{R}$. The set $R_{N_P} \subset \mathbb{R}^{N_P}$ serves as a finite dimensional parameter space and is, for the sake of simplicity, the hyper-cube $R_{N_P} := \bigtimes_{j=1}^{N_P} \left(-r_j, r_j\right) \subset (-1,1)^{N_P}$. The number N_P determines the dimension of the parameter space and will be large but finite, i.e. $1 \ll N_P < \infty$, which is known in the literature as *finite noise assumption*.

The parametric partial differential equation is now given on a sufficiently regular domain $\mathcal{D} \subset \mathbb{R}^d$ and for $G \in L^2(\mathcal{D})$ by

$$\begin{aligned} -\nabla \cdot (a(\boldsymbol{y},\boldsymbol{x})\nabla u(\boldsymbol{y},\boldsymbol{x})) &= G(\boldsymbol{x}) \quad \text{in } R_{N_P} \times \mathcal{D}, \\ u(\boldsymbol{y},\boldsymbol{x}) &= 0 \qquad\quad \text{in } R_{N_P} \times \partial\mathcal{D}, \end{aligned} \tag{1}$$

giving rise to a solution $u : R_{N_P} \times \overline{\mathcal{D}} \to \mathbb{R}$. Obviously, the spatial derivatives are only taken with respect to the spatial variable $\boldsymbol{x}$.

Depending on the practical application, we are not interested in the solution u directly but rather in a derived quantity of interest, which will be modeled by a linear functional q acting on the solution space, i.e.

$$Q(\boldsymbol{y}) := q\left(u(\boldsymbol{y},\cdot)\right) \in \mathbb{R}, \quad \boldsymbol{y} \in R_{N_P}. \tag{2}$$

Hence, $Q : R_{N_P} \to \mathbb{R}$ is a function operating only on the parameter space. The main task is now to reconstruct the map Q from sampled data $\{Q(\boldsymbol{y}_k)\}$ at specific parameter values $\boldsymbol{y}_k \in \mathbb{Y}_{N_S} := \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_{N_S}\} \subset R_{N_P}$, $1 \le k \le N_S$, where from now on we denote the number of sampling points by $N_S \in \mathbb{N}$. To avoid any confusion, we note here that N_P and N_S are uncorrelated.

By inserting the sampling points $\boldsymbol{y}_k \in \mathbb{Y}_{N_S}$ into (1), solving the now deterministic Poisson-Dirichlet problem and applying the functional q, we obtain the values $\{Q(\boldsymbol{y}_k)\}$. Except for only very few cases, these steps cannot be done analytically but only numerically. Hence, we introduce a finite dimensional *finite element space* $\mathbb{V}_h \subset \mathbb{V} = H_0^1(\mathcal{D})$, over which we solve

$$\begin{aligned} -\nabla \cdot (a(\boldsymbol{y}_k,\boldsymbol{x})\nabla u_h(\boldsymbol{y}_k,\boldsymbol{x})) &= G(\boldsymbol{x}) \quad \text{in } \mathcal{D}, \\ u_h(\boldsymbol{y}_k,\boldsymbol{x}) &= 0 \qquad\quad \text{on } \partial\mathcal{D}, \end{aligned} \tag{3}$$

weakly, yielding an approximation $u_h(\boldsymbol{y}_k,\cdot) \in \mathbb{V}_h$ to the true solution $u(\boldsymbol{y}_k,\cdot) \in \mathbb{V}$ and consequently perturbed samples

$$Q_h(\boldsymbol{y}_k) = q\left(u_h(\boldsymbol{y}_k,\cdot)\right) \approx Q(\boldsymbol{y}_k) = q\left(u(\boldsymbol{y}_k,\cdot)\right). \tag{4}$$

Note, that we assume that we can compute q analytically. This assumption is made for simplicity and has no impact on the error estimates in Section 5.

Choosing a standard finite element method (FEM) for solving (3) weakly, yields well-known error estimates (see for example Brenner & Scott [3]) for the quantities

$$\varepsilon_k := \|u_h(\boldsymbol{y}_k,\cdot) - u(\boldsymbol{y}_k,\cdot)\|_{\mathbb{V}} \quad \text{and} \quad \varepsilon := \max_{\boldsymbol{y}_k \in \mathbb{Y}_{N_S}} \varepsilon_k. \tag{5}$$

As mentioned above, we have to choose a discrete sampling set $\mathbb{Y}_{N_S}$ in the high-dimensional space R_{N_P}. This set should on the one hand be dense enough to represent R_{N_P} well and to allow a good reproduction of Q, but on the other hand, since we need to solve a partial differential equation for each element of $\mathbb{Y}_{N_S}$, has to be sparse enough for our method to be applicable. Hence, a natural choice for our set $\mathbb{Y}_{N_S}$ is a *sparse grid* $\mathbb{Y}_{N_S} := H(\ell, N_P)$ of level ℓ in N_P dimensions.

The final task is then to reconstruct the high-dimensional function Q from the data $\{(\boldsymbol{y}_k, Q(\boldsymbol{y}_k))\}$ which carry an intrinsic error. Thus, we do not want to use an interpolatory approach but rather a process from standard spline theory, called *smoothing splines* or *penalised least-squares*, see for example [12] and the references therein. The basic structure is a variational problem of the form

$$\widetilde{Q}_\lambda = \operatorname*{arg\,min}_{s \in \mathcal{H}_K} \sum_{k=1}^{N_S} |Q_h(\boldsymbol{y}_k) - s(\boldsymbol{y}_k)|^2 + \lambda \|s\|^2_{\mathcal{H}_K}, \tag{6}$$

where $\mathcal{H}_K$ denotes a *reproducing kernel Hilbert space (RKHS)* of real-valued functions with kernel K and where $\lambda > 0$ denotes a penalising parameter.

With this set-up we are able to base our error analysis on a new sampling inequality developed by Wendland and Rieger [6]. The main contribution of this paper, after choosing a specific RKHS $\mathcal{H}_{K_c}$, is the error estimate

$$\begin{aligned} \left\| Q - \widetilde{Q}_\lambda \right\|_{L^\infty(R_{N_P})} &\le C \left(f_{N_P,k}(N_S) + \sqrt{\lambda} g_{N_P}(N_S) \right) \|Q\|_{\mathcal{H}_{K_c}} \\ &+ C \left(\frac{1}{\sqrt{\lambda}} f_{N_P,k}(N_S) + g_{N_P}(N_S) \right) N_S h^t \max_{1 \le k \le N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})} \end{aligned}$$

where $C > 0$ is a constant, $\widetilde{Q}_\lambda$ is given in (6), $Q \in \mathcal{H}_{K_c}$ is the function from (2) defined by applying the functional q to the exact solution u of (1), h is the discretisation parameter of the finite element mesh, $t \ge 1$ and $f_{N_P,k}$ and g_{N_P} are known functions which have known asymptotic behaviour for $N_S \to \infty$. Furthermore, we derive conditions for λ and h such that

$$\left\| Q - \widetilde{Q}_\lambda \right\|_{L^\infty(R_{N_P})} \to 0, \quad N_S \to \infty,$$

with the order of $f_{N_P,k}$ and by further sharpening the estimate we even get nearly spectral convergence of the error.

As mentioned above, the setup described above closely follows in particular Griebel and Rieger [5]. However, there are two significant differences to their approach. On the one hand, we use a sparse grid as the sampling set $\mathbb{Y}_{N_S} \subset \mathbb{R}^{N_P}$, instead of a quasi-uniform data set as it is done in [5]. In the given, particular setting this is of significance as we deal with a high-dimensional problem and choosing a sparse grid helps to reduce the effect of the curse of dimensionality. On the other hand, we use a penalised least-squares approach for reconstructing the function Q instead of a support vector machine with Vapnik's loss function, as it has been done in [5]. However, the analysis carried out here for the penalised least-squares problem can easily be replaced by a similar analysis for a support vector machine.

This paper is organised as follows. In Section 2 we will review the existence, uniqueness and regularity of the solution of parametric partial differential equations of type (1). In Section 3 we will briefly review the theory of reproducing kernel Hilbert spaces (RKHSs), introduce the kernel and associated Hilbert space which we will use throughout this paper and discuss the advantages of the penalised least-squares reconstruction process. A justification for choosing sparse grids as our sampling space $\mathbb{Y}_{N_S}$ will be given in Section 4. Finally, in Section 5, we will state our main result, the above mentioned error estimate on the reconstruction process, which will be based upon a recently introduced sampling inequality for sparse grids.

We will use the following notation. We denote multi-indices by small, bold Greek letters, e.g. $\boldsymbol{\nu} \in \mathbb{N}_0^d$, and set $\boldsymbol{\nu}! = \prod_{j=1}^d \nu_j!$ and $\boldsymbol{\nu}^{\boldsymbol{\alpha}} = \prod_{j=1}^d {\nu_j}^{\alpha_j}$ for $\boldsymbol{\alpha} \in \mathbb{N}_0^d$. Furthermore, we use the notation $\|\boldsymbol{\nu}\|_1 = \nu_1 + \cdots + \nu_d$ for $\boldsymbol{\nu} \in \mathbb{N}_0^d$.

Additionally, we will use two kinds of Sobolev spaces over domains $\Omega \subseteq \mathbb{R}^d$ of the form $\Omega = \Omega_1 \times \cdots \times \Omega_d$ with $\Omega_j = (-1, 1)$ or $\Omega_j = (-r_j, r_j) \subset (-1, 1)$. On the one hand we will employ the classical Sobolev space

$$W_1^{k,2}(\Omega) := \left\{ f \in L^2(\Omega) \;:\; D^{\boldsymbol{\alpha}} f \in L^2(\Omega),\ \|\boldsymbol{\alpha}\|_1 \leq k \right\}$$

equipped with norm

$$\|f\|^2_{W_1^{k,2}(\Omega)} := \sum_{\|\boldsymbol{\alpha}\|_1 \leq k} \|D^{\boldsymbol{\alpha}} f\|^2_{L^2(\Omega)}\,.$$

On the other hand, we will use the tensor product Sobolev space defined by

$$\begin{aligned} W_\infty^{k,2}(\Omega) &:= \bigotimes_{j=1}^d W^{k,2}(\Omega_j) \\ &= \left\{ f \in L^2(\Omega) \;:\; D^{\boldsymbol{\alpha}} f \in L^2(\Omega),\ \|\boldsymbol{\alpha}\|_\infty \leq k \right\} \end{aligned}$$

together with the norm

$$\|f\|^2_{W_\infty^{k,2}(\Omega)} := \sum_{\|\boldsymbol{\alpha}\|_\infty \leq k} \|D^{\boldsymbol{\alpha}} f\|^2_{L^2(\Omega)}\,.$$

2 Parametric Partial Differential Equations

In this section, we will give an introduction to the theory of parametric partial differential equations by looking at existence and uniqueness of the solutions of the model problem (1). To this end, we need two domains. On the one hand we require a high-dimensional parameter domain. In our case this will be the anisotropic hyper-cube

$$R_{N_P} := \bigtimes_{j=1}^{N_P} \left(-r_j, r_j\right) \subset (-1,1)^{N_P}, \tag{7}$$

where $1 \ll N_P < \infty$.

On the other hand we need a spatial domain $\mathcal{D} \subset \mathbb{R}^d$, where usually $d = 2, 3$. We will assume $\mathcal{D}$ to be a bounded, convex and polygonal domain. If $G \in L^2(\mathcal{D})$ then the usual elliptic regularity theory holds for the weak formulation of (1), which is, with the usual energy space $\mathbb{V} := H_0^1(\mathcal{D})$, given by

$$\int_{\mathcal{D}} a(\boldsymbol{y},\boldsymbol{x})\nabla u(\boldsymbol{y},\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathcal{D}} G(\boldsymbol{x})v(\boldsymbol{x})\, d\boldsymbol{x}, \quad v \in \mathbb{V}, \boldsymbol{y} \in R_{N_P}. \tag{8}$$

In this paper, we assume the coefficient function a to have the form

$$a(\boldsymbol{y},\boldsymbol{x}) = a_0(\boldsymbol{x}) + \sum_{k=1}^{N_P} \phi_k(\boldsymbol{x}) y_k \tag{9}$$

with given $\phi_k \in L^\infty(\mathcal{D})$, $k \in \mathbb{N}$. In general, the sum in (9) will not be finite, so that the restriction to the first N_P terms introduces an additional error, which we will ignore throughout this paper.

We now follow [4] and extend the usual Lax-Milgram theory to complex valued coefficient functions $\tilde{a} : R_{N_P} \times \mathcal{D} \to \mathbb{C}$. To this end, we introduce the so-called *uniform (complex) ellipticity assumption* which requires the existence of two constants $R \geq r > 0$ such that

$$0 < r \leq \Re\left(\tilde{a}(\boldsymbol{y},\boldsymbol{x})\right) \leq |\tilde{a}(\boldsymbol{y},\boldsymbol{x})| \leq R \quad \boldsymbol{x} \in \mathcal{D}, \boldsymbol{y} \in R_{N_P}. \tag{10}$$

Here, $\Re(\cdot)$ denotes the real part of a complex number. By rearranging, we see that (10) is satisfied for the function a from (9) if the bounds

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

hold. Here, we have used that $\boldsymbol{y} \in R_{N_P} \subset (-1,1)^{N_P}$. While this assumption already leads to solutions $u(\boldsymbol{y},\boldsymbol{x})$, which are analytic as functions of $\boldsymbol{y}$, we need one additional assumption to also bound the coefficients of a Taylor expansion of the function $\boldsymbol{y} \mapsto u(\boldsymbol{y},\boldsymbol{x})$. Following [4] again, we will call a sequence (ρ_k) of positive numbers *δ-admissible* for the sequence (ϕ_k) if

$$\sum_{k=1}^{\infty} \rho_k\, |\phi_k(\boldsymbol{x})| \leq \Re\left(a_0(\boldsymbol{x})\right) - \delta, \tag{11}$$

For $\delta \leq r$, one can even have $\rho_k \geq 1$ for all $k \in \mathbb{N}$, see [4]. With this, we have the following result from [4, Theorem 1.2, Lemma 2.4].

Proposition 1. *Suppose that the uniform (complex) ellipticity assumption* (10) *holds with parameters* $0 < r \leq R < \infty$. *Then, the solution of* (8) *has the form*

$$u(\boldsymbol{y}, \cdot) = \sum_{\boldsymbol{\nu} \in \mathbb{N}^{N_P}} u_{\boldsymbol{\nu}}(\cdot) \boldsymbol{y}^{\boldsymbol{\nu}}. \tag{12}$$

for $u_{\boldsymbol{\nu}} \in \mathbb{V}$, *where convergence of the infinite series is understood with respect to the* $\|\cdot\|_{\mathbb{V}}$*-norm. Furthermore, if* (ρ_k) *is a* δ*-admissible sequence, then*

$$\|u_{\boldsymbol{\nu}}(\boldsymbol{y}, \cdot)\|_{\mathbb{V}} \leq \frac{\|G\|_{\mathbb{V}^*}}{\delta} \prod_{k=1}^{N_P} \rho_k^{-\nu_k}, \qquad \boldsymbol{y} \in R_{N_P}.$$

We now want to derive a parametric representation of the quantity of interest (2). To this end, we introduce the notation $\mathcal{R}(\lambda) \in \mathbb{V}$ for the Riesz representer of a linear functional $\lambda \in \mathbb{V}^*$. Then we have, by (12),

$$Q(\boldsymbol{y}) = q\left(u(\boldsymbol{y}, \cdot)\right) = \left\langle \sum_{\boldsymbol{\nu} \in \mathbb{N}^{N_P}} u_{\boldsymbol{\nu}}(\cdot) \boldsymbol{y}^{\boldsymbol{\nu}}, \mathcal{R}(q) \right\rangle_{\mathbb{V}} = \sum_{\boldsymbol{\nu} \in \mathbb{N}^{N_P}} \langle u_{\boldsymbol{\nu}}, \mathcal{R}(q) \rangle_{\mathbb{V}}\, \boldsymbol{y}^{\boldsymbol{\nu}},$$

which shows that the function Q, under certain assumption on the functional q, is also analytic. Later on, this representation of Q will guarantee that Q is indeed an element of the reproducing kernel Hilbert space of our specific choice.

3 Reproducing Kernel Hilbert Spaces

The reproduction problem (6) is at first sight an optimisation problem over an infinite dimensional function space. However, basic linear algebra shows that the solution must be contained in the span of the Riesz representers of the point evaluation functionals $\delta_{\boldsymbol{y}_k}$. These Riesz representers become particularly simple if the underlying Hilbert space $\mathcal{H}_K$ is a *Hilbert space with a reproducing kernel*. The Hilbert space $\mathcal{H}_K$ is a reproducing kernel Hilbert space with kernel $K : \Omega \times \Omega \to \mathbb{R}$ if K satisfies $K(\cdot, \boldsymbol{y}) \in \mathcal{H}_K$ and $f(\boldsymbol{y}) = \langle f, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{H}_K}$ for all $\boldsymbol{y} \in \Omega$ and all $f \in \mathcal{H}_K$. Details on such spaces can be found, for example, in [13].

3.1 Taylor Spaces and Power Series Kernels

In this paper, we are interested in a particular reproducing kernel Hilbert space, which consists of analytic functions and which was introduced in [14] and further investigated in [15]. The results below are taken from [5].

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. Under these assumptions, a *power series kernel* $K : (-1,1)^{N_P} \times (-1,1)^{N_P} \to \mathbb{R}$, which is a kernel of the form

$$K(\boldsymbol{x},\boldsymbol{y}) := \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{w_{\boldsymbol{\nu}}}{\boldsymbol{\nu}!^2}\boldsymbol{x}^{\boldsymbol{\nu}}\boldsymbol{y}^{\boldsymbol{\nu}}, \qquad \boldsymbol{x},\boldsymbol{y}\in(-1,1)^{N_P}, \tag{13}$$

is well-defined and analytic in each variable. The so-defined kernel K is the reproducing kernel of the Hilbert space $\mathcal{H}_K$ of functions

$$\mathcal{H}_K := \left\{ f : (-1,1)^{N_P} \to \mathbb{R} \;:\; f(\boldsymbol{x}) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} f_{\boldsymbol{\nu}}\boldsymbol{x}^{\boldsymbol{\nu}} \text{ with } \|f\|_{\mathcal{H}_K} < \infty \right\}, \tag{14}$$

where the norm is defined by the inner product

$$\langle f,g\rangle_{\mathcal{H}_K} := \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{1}{w_{\boldsymbol{\nu}}} D^{\boldsymbol{\nu}} f(\boldsymbol{0}) D^{\boldsymbol{\nu}} g(\boldsymbol{0}) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{\boldsymbol{\nu}!^2}{w_{\boldsymbol{\nu}}} f_{\boldsymbol{\nu}} g_{\boldsymbol{\nu}}, \tag{15}$$

see [14]. The next two results are taken from [5] and illustrate the reason for using such Taylor spaces $\mathcal{H}_K$ in this context.

First, we consider the embedding constant of the embedding of $\mathcal{H}_K$ into either of both Sobolev spaces, the classical isotropic space $W_1^{k,2}$ and the tensor product space $W_\infty^{k,2}$, i.e. we investigate the norm of the injection

$$\mathcal{W}_s(k) : \mathcal{H}_K \hookrightarrow W_s^{k,2}(R_{N_P}), \tag{16}$$

for $s \in \{1,\infty\}$, which is given in the next lemma.

Lemma 1. *Let R_{N_P} be defined by* (7) *Let $s\in\{1,\infty\}$. Suppose that there is a constant $\widehat{c}\in(0,1)$ such that the weights $w_{\boldsymbol{\nu}}$ satisfy $w_{\boldsymbol{\nu}} \le \widehat{c}^{\|\boldsymbol{\nu}\|_1}\boldsymbol{\nu}!^2$ for all $\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}$. Then, there is a constant $C>0$ such that the norm of the embedding operator (16) can be bounded by*

$$\|\mathcal{W}_s(k)\| \le \exp\left(\frac{C}{2}k\right) k!.$$

The second result which we require from [5] states that the function Q which we want to reconstruct indeed belongs to a Taylor space $\mathcal{H}_K$ if the weights $w_{\boldsymbol{\nu}}$ are chosen appropriately.

Lemma 2. *Suppose that the uniform (complex) ellipticity assumption* (10) *holds with parameters $0<r\le R<\infty$. Furthermore, let $(\rho_k)_k$ be a δ-admissible sequence with $0<\delta<r$ and $\rho_k>1$ for all k. Let $\boldsymbol{c}\in\mathbb{R}^{N_P}$ have components c_j with $c_j\in(\rho_j^{-1},1)$, $1\le j\le N_P$. Let $K=K_{\boldsymbol{c}}$ be defined by (13) with weights $w_{\boldsymbol{\nu}}=\boldsymbol{c}^{\boldsymbol{\nu}}\boldsymbol{\nu}!^2$. Then we have $Q\in\mathcal{H}_{K_{\boldsymbol{c}}}$.*

The space $\mathcal{H}_{K_{\boldsymbol{c}}}$ is a special case of $\mathcal{H}_K$. With the given, specific weights, the inner product becomes

$$\langle f, g\rangle_{\mathcal{H}_{K_{\boldsymbol{c}}}} = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{1}{\boldsymbol{c}^{\boldsymbol{\nu}}\boldsymbol{\nu}!^2} D^{\boldsymbol{\nu}} f(\boldsymbol{0}) D^{\boldsymbol{\nu}} g(\boldsymbol{0}) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \frac{1}{\boldsymbol{c}^{\boldsymbol{\nu}}} f_{\boldsymbol{\nu}} g_{\boldsymbol{\nu}}.$$

Furthermore, it is easy to see that this specific choice of weights leads to an explicit, analytic form of the reproducing kernel $K_{\boldsymbol{c}}$ given by

$$K_{\boldsymbol{c}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^{N_P}} \boldsymbol{c}^{\boldsymbol{\nu}} \boldsymbol{x}^{\boldsymbol{\nu}} \boldsymbol{y}^{\boldsymbol{\nu}} = \prod_{k=1}^{N_P} \frac{1}{1 - c_k x_k y_k}. \tag{17}$$

3.2 Penalised Least Squares

A typical application of reproducing kernel Hilbert spaces $\mathcal{H}_K$ are reconstruction processes of the form

$$\min_{s\in\mathcal{H}_K} \left(\sum_{k=1}^{N} |f(\boldsymbol{x}_k) - s(\boldsymbol{x}_k)|^2 + \lambda \|s\|^2_{\mathcal{H}_K} \right), \tag{18}$$

where the data $\{(\boldsymbol{x}_k, f(\boldsymbol{x}_k))\}_{1\le k\le N}$, $f \in \mathcal{H}_K$ is given. The parameter $\lambda > 0$ serves as a moderator between the fit to the data and the smoothness of the reconstruction $\widetilde{s}_\lambda$. In the RKHS setting, we have, by the well-known representer theorem, that the solution of the minimisation $\widetilde{s}_\lambda$ lies in the finite-dimensional space spanned by $K(\cdot, \boldsymbol{x}_k)$, $1 \le k \le N$, i.e. we have the representation

$$\widetilde{s}_\lambda = \sum_{k=0}^{N} \alpha_k K(\cdot, \boldsymbol{x}_k).$$

Furthermore, the coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$ can be computed by solving the linear system

$$(\boldsymbol{K} + \lambda \boldsymbol{I})\, \boldsymbol{\alpha} = \boldsymbol{f},$$

where $\boldsymbol{K}_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{f}_i = f(\boldsymbol{x}_i)$ and $\boldsymbol{I}$ is the identity matrix. It is well-known that this system has a positive definite system matrix and hence a unique solution. This also means that the least-squares problem (18) has a unique solution. Hence, in our situation, when employing the kernel $K = K_{\boldsymbol{c}}$ in (6), these general considerations give us a unique approximation $\widetilde{Q}_\lambda$ to Q derived from the noisy data $Q_h(\boldsymbol{y}_k)$, $1 \le k \le N_S$.

4 Sparse Grids

In this section, we demonstrate how we construct the sparse grid $H(\ell, d)$ of level ℓ and dimension d. Here, we follow mainly Wendland and Rieger [6].

To obtain the d-dimensional grid, we start with univariate sets of Chebyshev points. To do so we define a sequence of numbers (m_i) by

$$\begin{aligned} m_1 &= 1, \\ m_i &= 2^{i-1} + 1, \quad i > 1. \end{aligned}$$

Then, we define the Chebyshev point sets $X^{(i)}$ to be

$$\begin{aligned} X^{(1)} &:= X_{m_1} = \{0\}, \\ X^{(i)} &:= X_{m_i} = \left\{ x_j^{(i)} = -\cos\left(\frac{\pi(j-1)}{m_i - 1}\right) \;:\; 1 \le j \le m_i \right\}, \qquad i > 1. \end{aligned}$$

With these univariate point sets, we now define the *sparse grid* $\widetilde{H}(\ell, d)$ of level ℓ and dimension d, $\ell \ge d$, by

$$\widetilde{H}(\ell, d) = \bigcup_{\substack{\boldsymbol{i} \in \mathbb{N}^d \\ \|\boldsymbol{i}\|_1 = \ell}} X^{(i_1)} \times \cdots \times X^{(i_d)}. \tag{19}$$

As mentioned in Section 1, we choose the sampling space $\mathbb{Y}_{N_S}$ to be a sparse grid. As, by construction, $\widetilde{H}(\ell, N_P) \subset [-1, 1]^{N_P}$ is not a subset of R_{N_P}, we simply scale its points with a component-wise factor $r_j(1 - \mu)$, $0 < \mu \ll 1$, $1 \le j \le N_P$ and receive the *scaled sparse grid*

$$H(\ell, N_P) := \left\{ \left(r_1(1-\mu)x_1, \ldots, r_{N_P}(1-\mu)x_{N_P}\right) \;:\; \boldsymbol{x} \in \widetilde{H}(\ell, N_P) \right\}. \tag{20}$$

Now, we choose $\mathbb{Y}_{N_S} := H(\ell, N_P) \subset R_{N_P}$, where ℓ is a degree of freedom. For statements on the error of the reconstruction process we need to know the number of sampling points N_S. Unfortunately, the exact number is a priori unknown and there exist only lower and upper bounds, provided in [6],

$$2^{\ell - 2N_P + 1} \le N_S \le 2^{\ell - N_P + 1} \frac{\ell^{N_P - 1}}{(N_P - 1)!}.$$

Fortunately, as soon as we have created the sparse grid, we know exactly how many points it contains. A selection is given in Table 1. Clearly, we can control the number N_S for a given dimension N_P by choosing the level ℓ appropriately.

5 Error Estimates

We use this section to state the main results of this paper concerning error estimates for the optimisation problem (6), whose definition we recall here. We use the reproducing kernel Hilbert space $\mathcal{H}_{K_{\boldsymbol{c}}}$ introduced in (14) with the power series kernel $K_{\boldsymbol{c}}(\boldsymbol{x}, \boldsymbol{y}) = \prod_{j=1}^{N_P} \frac{1}{1 - c_j x_j y_j}$ of (17). Next, we define $J_{Q_h, \lambda} : \mathcal{H}_{K_{\boldsymbol{c}}} \to \mathbb{R}$, where

Table 1 Number of points N_S of the grid $H(\ell, d)$ for various space dimensions d and $\ell \geq d$.

$\ell \| d$	2	3	4	5	6	7	8
2	1						
3	5	1					
4	13	7	1				
5	29	25	9	1			
6	65	69	41	11	1		
7	145	177	137	61	13	1	
8	321	441	401	241	85	15	1
9	705	1073	1105	801	389	113	17
10	1537	2561	2929	2433	1457	589	145
11	3329	6017	7537	6993	4865	2465	849
12	7169	13953	18945	19313	15121	9017	3937
13	15361	32001	46721	51713	44689	30241	15713
14	32769	72705	113409	135073	127105	95441	56737
15	69633	163841	271617	345665	350657	287745	190881

$$J_{Q_h,\lambda}(s) := \sum_{k=1}^{N_P} |Q_h(\boldsymbol{y}_k) - s(\boldsymbol{y}_k)|^2 + \lambda \|s\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}}, \qquad s \in \mathcal{H}_{K_{\boldsymbol{c}}},$$

and set

$$\widetilde{Q}_\lambda := \operatorname*{arg\,min}_{s\in\mathcal{H}_{K_{\boldsymbol{c}}}} J_{Q_h,\lambda}(s).$$

The main objective is now to reconstruct the function $Q : R_{N_P} \to \mathbb{R}$ from perturbed samples $Q_h(\boldsymbol{y}_k) = q(u_h(\boldsymbol{y}_k, \cdot))$, $1 \leq k \leq N_S$, where the $\boldsymbol{y}_k \in H(\ell, N_P)$ and $u_h(\boldsymbol{y}_k, \cdot) \in \mathbb{V}_h$ is a FEM approximation of the exact solution.

As the data we have are corrupted by numerical error, we cannot directly employ the classic error estimates for penalised least-squares used in [6] since we cannot assume the function Q_h to be in the Hilbert space $\mathcal{H}_{K_{\boldsymbol{c}}}$. Nonetheless, we can assess this error. To do so we use the quantity

$$\varepsilon_k = Q_h(\boldsymbol{y}_k) - Q(\boldsymbol{y}_k) = q(u_h(\boldsymbol{y}_k, \cdot)) - q(u(\boldsymbol{y}_k, \cdot)) = q\left(u_h(\boldsymbol{y}_k, \cdot) - u(\boldsymbol{y}_k, \cdot)\right).$$

Hence we have the estimate

$$|\varepsilon_k| \leq \|q\|_{\mathbb{V}^*} \|u_h(\boldsymbol{y}_k, \cdot) - u(\boldsymbol{y}_k, \cdot)\|_{\mathbb{V}}$$

which means that $|\varepsilon_k|$ is bounded by the numerical error, which occurs in the solution of equation (3). This error enjoys well-known bounds depending on the smootheness of the solution. For example, we have from Brenner & Scott [3] the following estimate.

Lemma 3. *Let the finite element space be made up of elements up to order t with mesh width h. Assume that $u(\boldsymbol{y}_k, \cdot) \in H^{s+1}(\mathcal{D})$, for an $1 \leq s \leq t$ and all $1 \leq k \leq N_P$. Then, there is a constant $c > 0$ such that*

$$\|u_h(\boldsymbol{y}_k, \cdot) - u(\boldsymbol{y}_k, \cdot)\|_{\mathbb{V}} \leq ch^s |u(\boldsymbol{y}_k, \cdot)|_{H^{s+1}(\mathcal{D})}, \quad 1 \leq k \leq N_P. \tag{21}$$

Another tool we require is a sampling inequality, which allows us to bound the L^∞-norm of a function by a weighted sum of a full Sobolev norm and an ℓ^∞-norm over the discrete sampling set. The particular inequality we use is a new approach tailored for sparse grids. It gives the weights in terms of the number of sampling points and not in terms of the mesh width of the discrete set as it is usually done in sampling inequalities. This is of particular importance when working with sparse grids and in higher dimensions. The version we use in this paper is a special case of the one presented in [6].

Theorem 1. *Let $\widetilde{H}(\ell, N_P)$, $\ell \geq N_P$, be the sparse grid of* (19) *with N_S points. Then, for every function $a \in W_\infty^{k,2}\left((-1,1)^{N_P}\right)$, $k \in \mathbb{N}$, we have*

$$\|a\|_{L^\infty((-1,1)^{N_P})} \leq c\left(f_{N_P,k}(N_S)\|a\|_{W_\infty^{k,2}((-1,1)^{N_P})} + g_{N_P}(N_S)\|a\|_{\ell^\infty\left(\widetilde{H}(\ell,N_P)\right)}\right). \tag{22}$$

Here, the functions $f_{N_P,k}$ and g_{N_P} have for $N_S \to \infty$ the asymptotic behaviour

$$f_{N_P,k}(N_S) = O\left(N_S^{-k+\frac{1}{2}} \log(N_S)^{N_P\left(k+\frac{5}{2}\right)-\left(k+\frac{1}{2}\right)}\right), \tag{23}$$

$$g_{N_P}(N_S) = O\left(\log(N_S)^{N_P}\right). \tag{24}$$

In [6], the weight-functions $f_{N_P,k}, g_{N_P} : \mathbb{N} \to \mathbb{R}$ are given explicitly, but for our purposes the asymptotic behaviour is sufficient. Obviously, the function $f_{N_P,k}$ goes to zero for $N_S \to \infty$ while the function g_{N_P} grows logarithmically.

As (22) holds for the unscaled sparse grid $\widetilde{H}(\ell, N_P) \subset [-1,1]^{N_P}$, we need to modify it to fit into our framework, namely the scaled sparse grid $H(\ell, N_P)$ of (20). We scale the occuring functions by a simple coordinate transform, i.e., we scale the arguments by the same factors we used in the construction of $H(\ell, N_P)$ in Section 4. In doing so, we arrive at

$$\|b\|_{L^\infty(R_{N_P})} \leq \widetilde{c}\left(f_{N_P,k}(N_S)\|b\|_{W_\infty^{k,2}(R_{N_P})} + g_{N_P}(N_S)\|b\|_{\ell^\infty(H(\ell,N_P))}\right), \tag{25}$$

where $b \in W_\infty^{k,2}(R_{N_P})$, $f_{N_P,k}$ and g_{N_P} are the functions of Theorem 1 and $\widetilde{c}$ is a modified constant, depending additionally on μ and $\boldsymbol{r}$.

As we can embed $\mathcal{H}_{K_{\boldsymbol{c}}}$ into $W_\infty^{k,2}(R_{N_P})$, see Lemma 1, (25) above holds particularly for functions $b \in \mathcal{H}_{K_{\boldsymbol{c}}}$.

With these tools we are now able to estimate the error $Q - \widetilde{Q}_\lambda$ in the L^∞-norm. We mainly follow the ideas employed in [5] with appropriate modifications. We start by deriving two estimates for $\widetilde{Q}_\lambda$. The first one is based upon the bound

$$\begin{aligned}
\left|Q_h(\boldsymbol{y}_k) - \widetilde{Q}_\lambda(\boldsymbol{y}_k)\right|^2 &\leq \sum_{i=1}^{N_S}\left|Q_h(\boldsymbol{y}_i) - \widetilde{Q}_\lambda(\boldsymbol{y}_i)\right|^2 + \lambda\left\|\widetilde{Q}_\lambda\right\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}} \\
&= J_{Q_h,\lambda}\left(\widetilde{Q}_\lambda\right) \\
&\leq J_{Q_h,\lambda}(Q),
\end{aligned}$$

where we introduced positive summands and used that $\widetilde{Q}_\lambda$ is the minimiser of the functional $J_{Q_h,\lambda}$. This leads to

$$\left|Q_h(\boldsymbol{y}_k) - \widetilde{Q}_\lambda(\boldsymbol{y}_k)\right| \leq \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)| + \sqrt{\lambda}\, \|Q\|_{\mathcal{H}_{K_{\boldsymbol{c}}}} . \tag{26}$$

Here, we used that for any $a, b \geq 0$ the estimate $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ holds.

Next, we can estimate the consistency error, i.e. the point-wise error at the sampling nodes. We have by applying the triangle inequality

$$\left|Q(\boldsymbol{y}_k) - \widetilde{Q}_\lambda(\boldsymbol{y}_k)\right| \leq |Q(\boldsymbol{y}_k) - Q_h(\boldsymbol{y}_k)| + \left|Q_h(\boldsymbol{y}_k) - \widetilde{Q}_\lambda(\boldsymbol{y}_k)\right|,$$

which, together with (26), leads to our first crucial estimate

$$\left|Q(\boldsymbol{y}_k) - \widetilde{Q}_\lambda(\boldsymbol{y}_k)\right| \leq |Q_h(\boldsymbol{y}_k) - Q(\boldsymbol{y}_k)| + \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)| + \sqrt{\lambda}\, \|Q\|_{\mathcal{H}_{K_{\boldsymbol{c}}}} .$$

The second estimate on $\widetilde{Q}_\lambda$ follows from

$$\lambda \left\|\widetilde{Q}_\lambda\right\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}} \leq J_{Q_h,\lambda}\left(\widetilde{Q}_\lambda\right) \leq J_{Q_h,\lambda}(Q) \leq \sum_{k=1}^{N_S} |Q_h(\boldsymbol{y}_k) - Q(\boldsymbol{y}_k)|^2 + \lambda\, \|Q\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}}$$

and leads to

$$\left\|\widetilde{Q}_\lambda\right\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}} \leq \frac{1}{\lambda} \sum_{k=1}^{N_S} |Q_h(\boldsymbol{y}_k) - Q(\boldsymbol{y}_k)|^2 + \|Q\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}} .$$

We collect these results in the following lemma.

Lemma 4. *The reconstruction $\widetilde{Q}_\lambda$ from (6) of the function Q satisfies the bounds*

$$\left|Q(\boldsymbol{y}_k) - \widetilde{Q}_\lambda(\boldsymbol{y}_k)\right| \leq |Q_h(\boldsymbol{y}_k) - Q(\boldsymbol{y}_k)| + \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)| + \sqrt{\lambda}\, \|Q\|_{\mathcal{H}_{K_{\boldsymbol{c}}}} ,$$

$$\left\|\widetilde{Q}_\lambda\right\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}} \leq \frac{1}{\lambda} \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)|^2 + \|Q\|^2_{\mathcal{H}_{K_{\boldsymbol{c}}}} .$$

With these results we arrive at the following error estimate.

Theorem 2. *Let $H(\ell, N_P)$ with $\ell \geq N_P$ be the scaled sparse grid from* (20) *with N_S points. Assume that $Q \in \mathcal{H}_{K_{\boldsymbol{c}}}$, where $\mathcal{H}_{K_{\boldsymbol{c}}}$ is as in* (14) *with $K = K_{\boldsymbol{c}}$ from (17) satisfying the assumptions of Lemma 2. Then, there is a constant $c > 0$ such that for $\widetilde{Q}_\lambda = \arg\min_{s \in \mathcal{H}_{K_{\boldsymbol{c}}}} J_{Q_h,\lambda}(s)$ the error estimate*

$$\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})} \leq c\left(f_{N_P,k}(N_S) + \sqrt{\lambda} g_{N_P}(N_S)\right) \|Q\|_{\mathcal{H}_{K_c}} + c\left(\frac{1}{\sqrt{\lambda}} f_{N_P,k}(N_S) + g_{N_P}(N_S)\right) \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)|$$

holds, where $f_{N_P,k}$ *and* g_{N_P} *are from* (23) *and* (24).

Proof. The modified sampling inequality (25) with $b = Q - \widetilde{Q}_\lambda$ and Lemma 1 show

$$\begin{aligned}\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})} &\leq c f_{N_P,k}(N_S) \left\|Q - \widetilde{Q}_\lambda\right\|_{W_\infty^{k,2}(R_{N_P})} \\ &\quad + c g_{N_P}(N_S) \left\|Q - \widetilde{Q}_\lambda\right\|_{\ell^\infty(H(\ell,N_P))} \\ &\leq c f_{N_P,k}(N_S) \left\|Q - \widetilde{Q}_\lambda\right\|_{\mathcal{H}_{K_c}} \\ &\quad + c g_{N_P}(N_S) \left\|Q - \widetilde{Q}_\lambda\right\|_{\ell^\infty(H(\ell,N_P))} .\end{aligned}$$

Next, Lemma 4 allows us to bound the terms $\left\|Q - \widetilde{Q}_\lambda\right\|_{\ell^\infty(H(\ell,N_P))}$ and $\left\|Q - \widetilde{Q}_\lambda\right\|_{\mathcal{H}_{K_c}}$ separately. We have

$$\begin{aligned}&\left\|Q - \widetilde{Q}_\lambda\right\|_{\ell^\infty(H(\ell,N_P))} \\ &\leq \max_{k=1,\dots,N_S} |Q_h(\boldsymbol{y}_k) - Q(\boldsymbol{y}_k)| + \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)| + \sqrt{\lambda}\, \|Q\|_{\mathcal{H}_{K_c}} \\ &\leq 2 \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)| + \sqrt{\lambda}\, \|Q\|_{\mathcal{H}_{K_c}}\end{aligned}$$

and

$$\begin{aligned}\left\|Q - \widetilde{Q}_\lambda\right\|_{\mathcal{H}_{K_c}} &\leq \|Q\|_{\mathcal{H}_{K_c}} + \left\|\widetilde{Q}_\lambda\right\|_{\mathcal{H}_{K_c}} \\ &\leq 2\,\|Q\|_{\mathcal{H}_{K_c}} + \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{N_S} |Q_h(\boldsymbol{y}_i) - Q(\boldsymbol{y}_i)| .\end{aligned}$$

Inserting these bounds into the above bound on $\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})}$ concludes the proof. □

Taking also the error bound (21) of the finite element approximation into account yields the following corollary.

Corollary 1. *Let the assumptions of Theorem 2 hold. Assume further that* $u(\boldsymbol{y}_k, \cdot) \in H^{t+1}(\mathcal{D})$, $t \in \mathbb{N}$ *, for every* $\boldsymbol{y}_k \in H(\ell, N_P)$, $1 \leq k \leq N_S$*. Then the error estimate*

$$\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})} \leq c\left(f_{N_P,k}(N_S) + \sqrt{\lambda} g_{N_P}(N_S)\right) \|Q\|_{\mathcal{H}_{K_c}} + c\left(\frac{1}{\sqrt{\lambda}} f_{N_P,k}(N_S) + g_{N_P}(N_S)\right) N_S h^t \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k, \cdot)|_{H^{t+1}(\mathcal{D})}$$

holds. The functions $f_{N_P,k}$ *and* g_{N_P} *are from* (23) *and* (24).

Next, we want to discuss the convergence behaviour of the estimate above. As it is, this result is problematic since for $N_S \to \infty$ the function g_{N_P} tends to infinity. Hence, to achieve convergence, we have to couple the penalisation parameter λ and the mesh width h of the finite element grid to the number of points N_S in our sparse grid appropriately.

We start with the first term on the right-hand side of the bound in Corollary 1 . Its behaviour is determined by

$$f_{N_P,k}(N_S) + \sqrt{\lambda} g_{N_P}(N_S),$$

To have this term to behave like $f_{N_P,k}(N_S)$, which converges to zero for $N_S \to \infty$, we must choose $\sqrt{\lambda}$ sufficiently small . However, as we have a $1/\sqrt{\lambda}$ in the second term of the bound of Corollary 1, we cannot choose it too small. Hence, we choose a proportional constant $c_p > 0$ and let

$$\lambda = c_p \left(\frac{f_{N_P,k}(N_S)}{g_{N_P}(N_S)} \right)^2 . \tag{27}$$

With this choice, the bound in Corollary 1 becomes

$$\begin{aligned} &\left\| Q - \widetilde{Q}_\lambda \right\|_{L^\infty(R_{N_P})} \\ &\le c \left(f_{N_P,k}(N_S) \|Q\|_{\mathcal{H}_{K_c}} + g_{N_P}(N_S) N_S h^t \max_{1 \le k \le N_S} |u(\boldsymbol{y}_k, \cdot)|_{H^{t+1}(\mathcal{D})} \right). \end{aligned}$$

Hence, in order to have convergence, we need to ensure that the factor $g_{N_P}(N_S)N_S h^t$ in the second summand also tends to zero. If we want to keep the convergence order of size $f_{N_P,k}(N_S)$ then we have to choose h to satisfy

$$h \le \left(\frac{f_{N_P,k}(N_S)}{N_S g_{N_P}(N_S)} \right)^{\frac{1}{t}} . \tag{28}$$

We summarise these results in the next corollary.

Corollary 2. *Under the assumptions of Corollary 1 and with the choices (27) for the smoothing parameter and (28) for the finite element mesh size, the reconstruction error satisfies*

$$\|Q - \widetilde{Q}_\lambda\|_{L^\infty(R_{N_P})} \le c f_{N_P,k}(N_S) = c N_S^{-k+\frac{1}{2}} \log(N_S)^{N_P\left(k+\frac{5}{2}\right)-\left(k+\frac{1}{2}\right)}$$

with a constant $c = c(Q, u, k)$ *depending only on* Q*,* u *and* k*.*

Comparing this result to the one obtained in [5] and by experience from classical RKHS results, see [13], one would, in light of the analycity of the kernel, expect spectral convergence of the reconstruction error, similar to [7]. And indeed, a more thorough analysis of the occuring constants leads to the following result.

Corollary 3. *Under the assumptions of Corollary 1 and with the choices (27) for the smoothing parameter and (28) for the finite element mesh size, the reconstruction error satisfies for sufficiently large N_S the bound*

$$\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})} \leq c_1 {N_S}^2 e^{-c_2 N_S} \left(\|Q\|_{\mathcal{H}_{K_c}} + \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})}\right),$$

where $c_1, c_2 > 0$ are constants.

Proof. Corollary 2, together with the embedding constant from Lemma 1, gives the estimate

$$\begin{aligned}&\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})}\\ &\leq ck!e^{\frac{C}{2}k} N_S^{-k+\frac{1}{2}} \log(N_S)^{N_P\left(k+\frac{5}{2}\right)-\left(k+\frac{1}{2}\right)} \left(\|Q\|_{\mathcal{H}_{K_c}} + \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})}\right)\end{aligned}$$

Using Stirling's estimate $k! \leq ck^{k+\frac{1}{2}}e^{-k}$, $k \geq 1$, and keeping in mind that the logarithmic term $\log(N_S)$ grows slower than any root of N_S, especially $N_S^{\frac{1}{N_P(k+5/2)-(k+1/2)}}$, we obtain

$$\begin{aligned}&\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})}\\ &\leq ck^{k+\frac{1}{2}}e^{-k}e^{\frac{C}{2}k}N_S^{-k+\frac{1}{2}}N_S \left(\|Q\|_{\mathcal{H}_{K_c}} + \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})}\right)\\ &= c\left(k{N_S}^3\right)^{\frac{1}{2}} \left(e^{1-\frac{C}{2}}\frac{N_S}{k}\right)^{-k} \left(\|Q\|_{\mathcal{H}_{K_c}} + \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})}\right). \qquad (29)\end{aligned}$$

Next, for sufficiently large N_S, we choose k as $k = \frac{N_S}{\nu}$, where ν is a fixed constant such that $k \in \mathbb{N}$ and $e^{\frac{C}{2}-1} < \nu \leq N_S$ holds. Inserting this particular choice of k into (29) yields

$$\begin{aligned}&\left\|Q - \widetilde{Q}_\lambda\right\|_{L^\infty(R_{N_P})}\\ &\leq c\nu^{-\frac{1}{2}}{N_S}^2\left(\nu e^{1-\frac{C}{2}}\right)^{-\frac{N_S}{\nu}} \left(\|Q\|_{\mathcal{H}_{K_c}} + \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})}\right)\\ &= c\nu^{-\frac{1}{2}}{N_S}^2\left(e^{\frac{1}{\nu}\left(1-\frac{C}{2}+\log\nu\right)}\right)^{-N_S} \left(\|Q\|_{\mathcal{H}_{K_c}} + \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})}\right)\\ &= c_1{N_S}^2e^{-c_2N_S} \left(\|Q\|_{\mathcal{H}_{K_c}} + \max_{1\leq k\leq N_S} |u(\boldsymbol{y}_k,\cdot)|_{H^{t+1}(\mathcal{D})}\right),\end{aligned}$$

with $c_1 = c\nu^{-\frac{1}{2}} > 0$ and $c_2 = \frac{1}{\nu}\left(1 - \frac{C}{2} + \log\nu\right) > 0$ for ν in the given range. □

6 Concluding Remarks and Future Work

We have recaptured the basics of the regularity theory of parametric elliptic partial differential equations. One important result was that the solution, as a function of the parameter, is analytic and hence so is the quantity of interest.

The analyticity of the function we wanted to reconstruct motivated the choice of the specific reproducing kernel Hilbert space, a Taylor space, whose kernel is a power series kernel and thus analytic itself. With these choices we employed a regularised reconstruction process for approximating the smooth function from data which are usually corrupted by a (numerical) error, which means that the data-giving function is not an element of the approximation space.

To alleviate the curse of dimensionality we employed sparse grids, and a new type of sampling inequality which is expressed in the number of points rather than the fill distance of the sampling set.

Finally, we used the two degrees of freedom at our disposal, namely the FEM mesh width and the penalty parameter of the reconstruction process, to derive an overall error estimate.

Unfortunately, numerical verification of the derived error estimate is difficult due to the lack of good examples and the high dimensionality of the parameter space. Nonetheless, this will be pursued in the future.

Additionally, the kernel we used is globally supported which will lead to dense system matrices which should be avoided in practical applications, especially if the number of sampling points, i.e. the dimension of the matrix, becomes large. Switching to compactly supported kernels is subject of ongoing research. However, due to the high-dimensional nature of the underlying domain, a compactly supported kernel might not "see" enough information unless various scales are employed.

References

1. I. Babuska, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal. **45** (2007), 1005–1034.
2. I. Babuska, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal. **42** (2004), 800–825.
3. S. Brenner and L.R. Scott, *The mathematical theory of finite element methods*, Texts in Applied Mathematics, Springer New York, 2002.
4. Albert Cohen, Ronald DeVore, and Christoph Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic pdes*, Analysis and Applications **9** (2010), no. 1, 11–47.
5. Michael Griebel and Christian Rieger, *Reproducing kernel Hilbert spaces for parametric partial differential equations*, SIAM/ASA J. Uncertainty Quantification **5** (2017), 111–137.
6. Christian Rieger and Holger Wendland, *Sampling inequalities for sparse grids*, Numer. Math. **136** (2017), 439 – 466.
7. Christian Rieger and Barbara Zwicknagl, *Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning*, Advances in Computational Mathematics **32** (2010), 103–129.

8. Ralph C. Smith, *Uncertainty quantification*, Computational Science & Engineering, vol. 12, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014, Theory, implementation, and applications. MR 3155184
9. Christian Soize, *Uncertainty quantification*, Interdisciplinary Applied Mathematics, vol. 47, Springer, Cham, 2017, An accelerated course with advanced applications in computational engineering, With a foreword by Charbel Farhat. MR 3618803
10. T. J. Sullivan, *Introduction to uncertainty quantification*, Texts in Applied Mathematics, vol. 63, Springer, Cham, 2015. MR 3364576
11. Luis Tenorio, *An introduction to data analysis and uncertainty quantification for inverse problems*, Mathematics in Industry (Philadelphia), Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2017. MR 3672154
12. G. Wahba, *Spline models for observational data*, CBMS-NSF, Regional Conference Series in Applied Mathematics, Siam, Philadelphia, 1990.
13. Holger Wendland, *Scattered data approximation*, Cambridge University Press, 2004.
14. Barbara Zwicknagl, *Power series kernels*, Constructive Approximation **29** (2009), no. 1, 61–84.
15. Barbara Zwicknagl and Robert Schaback, *Interpolation and approximation in taylor spaces*, J. Approx. Theory **171** (2013), 65–83.