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

C. Rieger

Iterated Landweber Method for Radial Basis Functions Interpolation with Finite Accuracy

INS Preprint No. 1806

May 2018

ITERATED LANDWEBER METHOD FOR RADIAL BASIS FUNCTIONS INTERPOLATION WITH FINITE ACCURACY

CHRISTIAN RIEGER

Abstract. We consider the reconstruction of a function stemming from a reproducing kernel Hilbert space using data which is perturbed by a deterministic error of maximal size ϵ_∞. The accuracy $\epsilon_\infty \geq 0$ provides an upper bound for reconstruction error estimates. Therefore, the main emphasis of this work is an a priori coupling of the data error, the error stemming from discretization and the numerical linear algebra. The coupling should provide an optimized cost-benefit ratio, i.e., we try to spend no numerical work to solving linear systems of equations if we cannot increase the overall accuracy of the reconstruction. Following [4], we focus here on the iterated Landweber method which serves both as numerical solver of the linear system of equations and as a regularization technique accounting for the inexact data. This method introduces a regularization parameter which should be chosen small from an error estimate perspective. On the other hand this parameter stabilizes the numerical computation. We outline this balance with the example of in-exact Cholesky decompositions. Here, we also take the finite precision of number representations in the computer into account.

1. Introduction

Let $\Omega \subset \mathbb{R}^d$ be a subset and let $X_N := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ be a discrete set. A Hilbert space $\left(\mathcal{H}_K(\Omega), (\cdot,\cdot)_{\mathcal{H}_K(\Omega)}\right)$ of functions

$$\mathcal{H}_K(\Omega) \subset \mathbb{R}^\Omega := \{f : \Omega \to \mathbb{R}\}$$

such that there is a function $K : \Omega \times \Omega \to \mathbb{R}$ satisfying

$$\begin{aligned} &K(\cdot, \boldsymbol{x}) \in \mathcal{H}_K(\Omega) \quad \text{for all } \boldsymbol{x} \in \Omega \\ &(f, K(\cdot, \boldsymbol{x}))_{\mathcal{H}_K(\Omega)} = f(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \Omega \text{ and all } f \in \mathcal{H}_K(\Omega) \end{aligned}$$

is called reproducing kernel Hilbert space (RKHS) with kernel K. In the sequel we will restrict ourselves to positive definite kernels. We consider the reconstruction of a function $f \in \mathcal{H}_K(\Omega)$ using data

$$(\boldsymbol{x}_j, y_j) = (\boldsymbol{x}_j, f(\boldsymbol{x}_j) + \epsilon_j) \in \mathbb{R}^d \times \mathbb{R}, \quad \text{for } j = 1, \ldots, N$$

where $\|\boldsymbol{\epsilon}\|_{\ell^p(\mathbb{R}^N)} \leq \epsilon_p$ for a $1 \leq p \leq \infty$.

We define a sampling map

$$S_{K;X_N} : \mathcal{H}_K(\Omega) \to \mathbb{R}^N, \quad f \mapsto (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N))^\top, \tag{1}$$

Date: May 8, 2018.

where we silently endow $\mathbb{R}^N$ with the usual Euclidean inner product. With respect to this inner product and the inner product on $\mathcal{H}_K(\Omega)$, we can define the adjoint map by

$$\left(S_{K;X_N}(f), \boldsymbol{v}\right)_{\ell^2(\mathbb{R}^N)} = \left(f, S^\star_{K;X_N}(\boldsymbol{v})\right)_{\mathcal{H}_K(\Omega)} \quad \text{for all } f \in \mathcal{H}_K(\Omega) \text{ and all } \boldsymbol{v} \in \mathbb{R}^N.$$

We can explicitly compute the adjoint to be

$$S^\star_{K;X_N} : \mathbb{R}^N \to \mathcal{H}_K(\Omega), \quad \boldsymbol{v} \mapsto \sum_{\boldsymbol{x}_j \in X_N} (\boldsymbol{v} \cdot \boldsymbol{e}_j) K(\cdot, \boldsymbol{x}_j) = \sum_{\boldsymbol{x}_j \in X_N} v_j K(\cdot, \boldsymbol{x}_j) \in V_{K;X_N}, \tag{2}$$

where we denote

$$V_{K;X_N} := \operatorname{span}\{K(\cdot, \boldsymbol{x}_j) \; : \; \boldsymbol{x}_j \in X_N\} \subset \mathcal{H}_K(\Omega). \tag{3}$$

We observe that

$$S_{K;X_N} S^\star_{K;X_N} = \boldsymbol{K}_{X_N,X_N} = \begin{pmatrix} K(\boldsymbol{x}_1, \boldsymbol{x}_1) & \dots & K(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ K(\boldsymbol{x}_N, \boldsymbol{x}_1) & \dots & K(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix} \in \mathbb{R}^{N \times N}, \tag{4}$$

where we use the standard basis in $\mathbb{R}^N$ for the matrix representation. We consider the orthogonal projection

$$\Pi_{V_{K;X_N}} : \mathcal{H}_K(\Omega) \to V_{K;X_N}, \quad f \mapsto \Pi_{V_{K;X_N}}(f) = S^\star_{K;X_N} \left(S_{K;X_N} S^\star_{K;X_N}\right)^{-1} S_{K;X_N}(f), \tag{5}$$

which, by construction, yields an interpolation to the function f, i.e., we have that the residual $S_{K;X_N}\left(\Pi_{V_{K;X_N}}(f) - f\right) = \boldsymbol{0}$ vanishes. The fact that we use perturbed data prohibit to use this best approximation property in the approximation scheme that we will employ. Nevertheless, we still can use perturbation arguments if we assume that the data error ϵ_∞ is small.

We have to assume, however, that $\boldsymbol{K}_{X_N,X_N} \in \mathbb{R}^{N \times N}$ is ill-conditioned. This makes the perturbation analysis more delicate. Moreover, ill-conditioned matrices are in particular problematic in numerical computations due to the limited accuracy of floating point operations. Let $\mathbb{F}$ be the floating point numbers as specified in an IEEE standard. As in [15], we focus on the IEEE 754 standard with a relative rounding error unit

$$\alpha := 2^{-54} \approx 1.1 * 10^{-16} \tag{6}$$

in double precision. Furthermore, we need sums of floating point numbers

$$\mathbb{F} = \mathbb{F}_{[1]} \subset \mathbb{F}_{[m]} := \left\{ x \in \mathbb{R} \; : \; x = \sum_{k=1}^{m} x_k \quad \text{with } x_k \in \mathbb{F} \right\} \subset \mathbb{R} \tag{7}$$

for a natural number $m \in \mathbb{N}$. Ill-conditioned means in the context of this manuscript

$$\operatorname{cond}_2(\boldsymbol{K}_{X_N,X_N}) = \|\boldsymbol{K}_{X_N,X_N}\|_{\ell_2 \to \ell_2} \left\|\boldsymbol{K}^{-1}_{X_N,X_N}\right\|_{\ell_2 \to \ell_2} > \alpha^{-1} \approx 10^{16}. \tag{8}$$

It is known that a Cholesky decomposition of a matrix $\boldsymbol{A} \in \mathbb{F}^{N \times N}$ might break down due the rounding errors which in turn might cause complex numbers to show up, see also [12].

The remainder of the manuscript is organized as follows: In Section 2 we present the basic algorithm for the iterated Landweber method and apply it to Gramian matrices arising in kernel methods. Moreover, we derive first discrete error estimates, that is, error estimates for the solutions of the linear systems. In Section 3 we use *sampling inequalities* to transfer those discrete error estimates to error estimates for function reconstruction problems. In order to motivate a non-vanishing choice for the regularization parameter, we study the influence of finite precision arithmetic on the iterated Landweber method in Section 4. It turns out that a worst case error analysis motivates rather large choices for the regularization parameter. We discuss this in a Conclusion, see Section 5.

2. Iterated Landweber method as preconditioner

Here, we follow [4]. Since the matrix $\boldsymbol{K}_{X_N,X_N} \in \mathbb{R}^{N\times N}$ defined in (4) is symmetric and positive definite due to the restriction to positive definite kernels, we can write

$$\boldsymbol{K}_{X_N,X_N} = \sum_{n=1}^{N} \lambda_{n;K;X_N} \boldsymbol{v}_{n;K;X_N} \otimes \boldsymbol{v}_{n;K;X_N}, \tag{9}$$

where $\lambda_{1;K;X_N} \geq \cdots \geq \lambda_{N;k;X_N} > 0$ denote the not necessarily distinct eigenvalues of the matrix $\boldsymbol{K}_{X_N,X_N} \in \mathbb{R}^{N\times N}$ and the $\boldsymbol{v}_{j;K;X_N} \in \mathbb{R}^N$ denote the corresponding orthonormal basis of eigenvectors. We can define for $\mu \geq 0$

(10)

$$\boldsymbol{K}_{\mu;X_N,X_N} := \boldsymbol{K}_{X_N,X_N} + \mu \,\mathrm{Id}_{N\times N} = \sum_{n=1}^{N} (\lambda_{n;K;X_N} + \mu)\, \boldsymbol{v}_{n;K;X_N} \otimes \boldsymbol{v}_{n;K;X_N} \in \mathbb{R}^{N\times N}$$

which is by construction also symmetric and positive definite. We start with the observation that

$$\mu \boldsymbol{K}^{-1}_{\mu;X_N,X_N} = \sum_{n=1}^{N} \overbrace{\frac{\mu}{\lambda_{n;K;X_N} + \mu}}^{<1} \boldsymbol{v}_{n;K;X_N} \otimes \boldsymbol{v}_{n;K;X_N} \in \mathbb{R}^{N\times N}. \tag{11}$$

Furthermore, direct calculations yield for $1 \leq n \leq N$

$$\begin{aligned}
\mu^{-1} \sum_{\ell=1}^{\infty} \left(\frac{\mu}{\lambda_{n;K;X_N} + \mu} \right)^{\ell} &= \mu^{-1} \left(\sum_{\ell=0}^{\infty} \left(\frac{\mu}{\lambda_{n;K;X_N} + \mu} \right)^{\ell} - 1 \right) \\
= \mu^{-1} \left(\frac{1}{1 - \frac{\mu}{\lambda_{n;K;X_N} + \mu}} - 1 \right) &= \mu^{-1} \left(\frac{\lambda_{n;K;X_N} + \mu}{\lambda_{n;K;X_N}} - 1 \right) = \lambda^{-1}_{n;K;X_N}.
\end{aligned}$$

Hence, we get, see [4]

$$\boldsymbol{K}^{-1}_{X_N,X_N} = \mu^{-1} \sum_{\ell=1}^{\infty} \left(\mu \boldsymbol{K}^{-1}_{\mu;X_N,X_N} \right)^{\ell}. \tag{12}$$

This can be rewritten as an iterative method for the solution of the linear system of equations. We get for the linear system

$$\boldsymbol{K}_{X_N,X_N} \boldsymbol{x}_{\boldsymbol{b}} = \boldsymbol{b}$$

the following fixed point formulation, see [4]

$$\begin{aligned}\boldsymbol{x}_{\boldsymbol{b}} &= \boldsymbol{K}_{X_N,X_N}^{-1}\boldsymbol{b} = \mu^{-1}\sum_{\ell=1}^{\infty}\left(\mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^{\ell}\boldsymbol{b} = \sum_{\ell=1}^{\infty}\left(\mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^{\ell-1}\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b}\\ &= \sum_{\ell=0}^{\infty}\left(\mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^{\ell}\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} = \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} + \sum_{\ell=1}^{\infty}\left(\mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^{\ell}\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b}\\ &= \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} + \mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\mu^{-1}\sum_{\ell=1}^{\infty}\left(\mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^{\ell}\boldsymbol{b} = \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} + \mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{x}_{\boldsymbol{b}}.\end{aligned}$$

Define

$$\Phi_{\boldsymbol{b}} : \mathbb{R}^N \to \mathbb{R}^N, \quad \boldsymbol{x} \mapsto \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} + \mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{x}.$$

Then, we observed that $\boldsymbol{x}_{\boldsymbol{b}} = \Phi_{\boldsymbol{b}}(\boldsymbol{x}_{\boldsymbol{b}})$. The usual fix point iteration now reads,

(13)

$$\boldsymbol{x}_{\boldsymbol{b}}^{(0)} := \boldsymbol{0} \quad \text{and} \quad \boldsymbol{x}_{\boldsymbol{b}}^{(n+1)} := \Phi_{\boldsymbol{b}}\left(\boldsymbol{x}_{\boldsymbol{b}}^{(n)}\right) = \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} + \mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{x}_{\boldsymbol{b}}^{(n)}, \quad \text{for } n > 0.$$

An equivalent formulation (see [5] for more details) is provided by the *iterative improvement*, i.e.,

$$\boldsymbol{x}_{\boldsymbol{b}}^{(0)} := \boldsymbol{0} \quad \text{and} \quad \begin{pmatrix}\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{e}^{(n)}\\ \boldsymbol{x}_{\boldsymbol{b}}^{(n+1)}\end{pmatrix} = \begin{pmatrix}-\boldsymbol{K}_{X_N,X_N} & \boldsymbol{0}\\ \mathbf{Id}_{N\times N} & \mathbf{Id}_{N\times N}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_{\boldsymbol{b}}^{(n+1)}\\ \boldsymbol{e}^{(n)}\end{pmatrix} \tag{14}$$

where $\boldsymbol{e}^{(n)} = \boldsymbol{x}_{\boldsymbol{b}}^{(n+1)} - \boldsymbol{x}_{\boldsymbol{b}}^{(n)}$. From equation (12), we can also derive an approximation, i.e.,

$$\boldsymbol{C}_{X_N,X_N}^{(L)} = \mu^{-1}\sum_{\ell=1}^{L}\left(\mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^{\ell} \approx \boldsymbol{K}_{X_N,X_N}^{-1} \tag{15}$$

for $L \in \mathbb{N}$ large.

Lemma 1. *For the matrix* $\boldsymbol{C}_{X_N,X_N}^{(L)}$ *defined in* (15) *we have the following identity*

$$\boldsymbol{C}_{X_N,X_N}^{(L)}\boldsymbol{b} = \boldsymbol{x}_{\boldsymbol{b}}^{(L)}.$$

Moreover, we have the bound

$$\left\|\boldsymbol{C}_{X_N,X_N}^{(L)}\right\|_{\ell_2\to\ell_2} \leq \lambda_{\min}^{-1}(\boldsymbol{K}_{X_N,X_N})\left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^{L}\right). \tag{16}$$

Proof. We use proof by induction. For $L = 0$, we get $\boldsymbol{C}_{X_N,X_N}^{(0)} = \boldsymbol{0}$ and hence $\boldsymbol{C}_{X_N,X_N}^{(0)}\boldsymbol{b} = \boldsymbol{x}_{\boldsymbol{b}}^{(0)} = \boldsymbol{0}$. Furthermore, for $L = 1$, we have

$$\boldsymbol{C}_{X_N,X_N}^{(1)}\boldsymbol{b} = \left(\mu^{-1}\sum_{\ell=1}^{1}\left(\mu\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^{\ell}\right)\boldsymbol{b} = \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} = \boldsymbol{x}_{\boldsymbol{b}}^{(1)}.$$

For the induction step, we assume $\boldsymbol{C}^{(\ell)}_{X_N,X_N}\boldsymbol{b} = \boldsymbol{x}^{(\ell)}_{\boldsymbol{b}}$ for all $1 \leq \ell \leq L$. Now, we have

$$\begin{aligned}
\mu^{-1}\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{C}^{(L+1)}_{X_N,X_N}\boldsymbol{b} &= \mu^{-1}\sum_{\ell=1}^{L+1}\left(\mu\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\right)^{\ell}\boldsymbol{b} = \mu^{-1}\sum_{\ell=1}^{L+1}\left(\mu\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\right)^{\ell-1}\boldsymbol{b}\\
&= \mu^{-1}\sum_{\ell=0}^{L}\left(\mu\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\right)^{\ell}\boldsymbol{b} = \mu^{-1}\boldsymbol{b} + \mu^{-1}\sum_{\ell=1}^{L}\left(\mu\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\right)^{\ell}\boldsymbol{b}\\
&= \mu^{-1}\boldsymbol{b} + \boldsymbol{C}^{(L)}_{X_N,X_N}\boldsymbol{b} = \mu^{-1}\boldsymbol{b} + \boldsymbol{x}^{(L)}_{\boldsymbol{b}}.
\end{aligned}$$

Hence, we obtain

$$\boldsymbol{C}^{(L+1)}_{X_N,X_N}\boldsymbol{b} = \boldsymbol{K}^{-1}_{\mu;X_N,X_N}\boldsymbol{b} + \mu\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\boldsymbol{x}^{(L)}_{\boldsymbol{b}} = \boldsymbol{x}^{(L+1)}_{\boldsymbol{b}},$$

which concludes the first part of the proof. We recall that we have for $0 < q < 1$

$$\begin{aligned}
\sum_{\ell=1}^{\infty} q^{\ell} &= \frac{1}{1-q} - 1 = \frac{q}{1-q},\\
\sum_{\ell=1}^{L} q^{\ell} &= \frac{1-q^{L+1}}{1-q} - 1 = \frac{q-q^{L+1}}{1-q} = q\frac{1-q^{L}}{1-q}, \quad \text{and}\\
\sum_{\ell=L+1}^{\infty} q^{\ell} &= \frac{q}{1-q} - \frac{q-q^{L+1}}{1-q} = \frac{q^{L+1}}{1-q}.
\end{aligned}$$

We apply those results to $q = \frac{\mu}{\lambda_{n;K;X_N}+\mu} \in (0,1)$ to get

$$\begin{aligned}
\sum_{\ell=1}^{L}\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{\ell} &= \frac{\mu}{\lambda_{n;K;X_N}+\mu}\frac{1}{1-\frac{\mu}{\lambda_{n;K;X_N}+\mu}}\left(1-\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{L}\right)\\
&= \frac{\mu}{\lambda_{n;K;X_N}}\left(1-\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{L}\right).
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\boldsymbol{C}^{(L)}_{X_N,X_N} &= \mu^{-1}\sum_{n=1}^{N}\sum_{\ell=1}^{L}\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{\ell}\boldsymbol{v}_{n;K;X_N}\otimes\boldsymbol{v}_{n;K;X_N}\\
&= \sum_{n=1}^{N}\frac{1}{\lambda_{n;K;X_N}}\left(1-\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{L}\right)\boldsymbol{v}_{n;K;X_N}\otimes\boldsymbol{v}_{n;K;X_N}. \qquad (17)
\end{aligned}$$

We consider for $\mu > 0$ and $L \in \mathbb{N}$

$$f_{\mu;L} : (\lambda_{\min}(\boldsymbol{K}_{X_N,X_N}), \lambda_{\max}(\boldsymbol{K}_{X_N,X_N})) \to \mathbb{R}, \quad x \mapsto f_{\mu;L}(x) := x^{-1}\left(\frac{\mu}{x+\mu}\right)^{L}. \qquad (18)$$

Direct calculations yield

$$\begin{aligned} \frac{d}{dx} f_{\mu;L}(x) &= (-1)x^{-2}\left(\frac{\mu}{x+\mu}\right)^{L} + x^{-1}L\left(\frac{\mu}{x+\mu}\right)^{L-1}\frac{-\mu}{(x+\mu)^2} \\ &= (-1)x^{-1}\left(\frac{\mu}{x+\mu}\right)^{L}\left(\frac{1}{x}+\frac{L}{(x+\mu)}\right) < 0. \end{aligned} \tag{19}$$

In order to find the maximal value in the sum (17), we consider

$$h_{\mu;L} : (\lambda_{\min}(\boldsymbol{K}_{X_N,X_N}), \lambda_{\max}(\boldsymbol{K}_{X_N,X_N})) \to \mathbb{R}, \quad x \mapsto h_{\mu;L}(x) := x^{-1}\left(1 - \left(\frac{\mu}{x+\mu}\right)^{L}\right).$$

We obtain

$$\frac{d}{dx} h_{\mu;L}(x) = -x^{-2} + x^{-1}\left(\frac{\mu}{x+\mu}\right)^{L}\left(\frac{1}{x}+\frac{L}{(x+\mu)}\right).$$

We claim that $\frac{d}{dx} h_{\mu;L}(x) \le 0$, i.e.,

$$x^{-2} \ge x^{-1}\left(\frac{\mu}{x+\mu}\right)^{L}\left(\frac{1}{x}+\frac{L}{(x+\mu)}\right) \Leftrightarrow 1 \ge \left(\frac{\mu}{x+\mu}\right)^{L}\left(1+\frac{Lx}{(x+\mu)}\right), \tag{20}$$

where we used that $x > 0$. To show (20), we define

$$g_{\mu;L} : (\lambda_{\min}(\boldsymbol{K}_{X_N,X_N}), \lambda_{\max}(\boldsymbol{K}_{X_N,X_N})) \to \mathbb{R}, \quad x \mapsto g_{\mu;L}(x) := \left(\frac{\mu}{x+\mu}\right)^{L}\left(1+\frac{Lx}{(x+\mu)}\right).$$

We observe $g_{\mu;L}(0) = 1$ and

$$\begin{aligned} \frac{d}{dx} g_{\mu;L}(x) &= -\frac{L}{x+\mu}\left(\frac{\mu}{x+\mu}\right)^{L}\left(1+\frac{Lx}{(x+\mu)}\right) + \left(\frac{\mu}{x+\mu}\right)^{L}\frac{x+\mu-x}{(x+\mu)^2}L \\ &= \left(\frac{\mu}{x+\mu}\right)^{L}\frac{1}{(x+\mu)^2}\left(-L(x+\mu) - L^2x + L\mu\right) = -\left(\frac{\mu}{x+\mu}\right)^{L}\frac{L(x+x^2)}{(x+\mu)^2} < 0. \end{aligned}$$

This shows the inequality (20) and hence proves that $h_{\mu;L}$ is decreasing in x. This shows that we can estimate

$$\begin{aligned} \left\|\boldsymbol{C}^{(L)}_{X_N,X_N}\right\|_{\ell_2\to\ell_2} &\le \max_{x\in(\lambda_{\min}(\boldsymbol{K}_{X_N,X_N}),\lambda_{\max}(\boldsymbol{K}_{X_N,X_N}))} |h_{\mu;L}(x)| \le h_{\mu;L}(\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})) \\ &= \lambda_{\min}^{-1}(\boldsymbol{K}_{X_N,X_N})\left(1-\left(\frac{\mu}{\mu+\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^{L}\right), \end{aligned}$$

which finishes the proof. □

2.1. **Error Estimates for exact Landweber iteration.** In this Section we assume to be given data $\boldsymbol{y} = S_{X_N}(f) + \boldsymbol{\epsilon}$ where $f \in \mathcal{H}_K(\Omega)$ is a fixed but unknown function. We study the influence of the approximation if we replace the matrix $\boldsymbol{K}^{-1}_{X_N,X_N}$ by $\boldsymbol{C}^{(L)}_{X_N,X_N}$ for the interpolation process, c.f. (5).

Lemma 2. *Let the data $\boldsymbol{y} = S_{X_N}(f) + \boldsymbol{\epsilon}$ with some $f \in \mathcal{H}_K(\Omega)$ be given. We consider the error in the reproducing kernel Hilbert space norm by*

$$\begin{aligned}\mathcal{E}_{\mathcal{H}}^2(\boldsymbol{y}) &:= \left\| S_{K;X_N}^\star \boldsymbol{K}_{X_N,X_N}^{-1}\boldsymbol{y} - S_{K;X_N}^\star \boldsymbol{C}_{X_N,X_N}^{(L)}\boldsymbol{y}\right\|_{\mathcal{H}_K(\Omega)}^2 \\ &\leq \lambda_{\min}^{-1}(\boldsymbol{K}_{X_N,X_N})\left(\frac{\mu}{\mu+\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^{2L}\|\boldsymbol{y}\|_{\ell^2(\mathbb{R}^N)}^2 \end{aligned} \tag{21}$$

Moreover, we have for the error in the discrete norm on the sampling nodes

(22)

$$\mathcal{E}_{\ell^2}^2(\boldsymbol{y}) := \left\|\left(\operatorname{Id} - \boldsymbol{K}_{X_N,X_N}\boldsymbol{C}_{X_N,X_N}^{(L)}\right)\boldsymbol{y}\right\|_{\ell^2(\mathbb{R}^N)}^2 \leq \left(\frac{\mu}{\mu+\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^{2L}\|\boldsymbol{y}\|_{\ell^2(\mathbb{R}^N)}^2 .$$

Proof. We have

$$\begin{aligned}\mathcal{E}_{\mathcal{H}}^2(\boldsymbol{y}) &:= \left\| S_{K;X_N}^\star \boldsymbol{K}_{X_N,X_N}^{-1}\boldsymbol{y} - S_{K;X_N}^\star \boldsymbol{C}_{X_N,X_N}^{(L)}\boldsymbol{y}\right\|_{\mathcal{H}_K(\Omega)}^2 \\ &= \left\| S_{K;X_N}^\star\left(\boldsymbol{K}_{X_N,X_N}^{-1} - \boldsymbol{C}_{X_N,X_N}^{(L)}\right)\boldsymbol{y}\right\|_{\mathcal{H}_K(\Omega)}^2 = \left\| S_{K;X_N}^\star \boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{y}\right\|_{\mathcal{H}_K(\Omega)}^2 \\ &= \left(\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{y}, \boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{y}\right)_{\ell^2(\mathbb{R}^N)} \\ &= \left(\boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{y}\right)^\top \boldsymbol{K}_{X_N,X_N}\boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{y} \\ &= \boldsymbol{y}^\top \boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{y},\end{aligned}$$

where we define the symmetric matrix

$$\boldsymbol{E}_{X_N,X_N}^{(L)} := \mu^{-1}\sum_{\ell=L+1}^{\infty}\left(\mu \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\right)^\ell = \boldsymbol{K}_{X_N,X_N}^{-1} - \boldsymbol{C}_{X_N,X_N}^{(L)} \in \mathbb{R}^{N\times N}. \tag{23}$$

We observe using again the geometric series with $q = \frac{\mu}{\lambda_{j;k;X_N}+\mu} \in (0,1)$ that

$$\begin{aligned}\boldsymbol{E}_{X_N,X_N}^{(L)} &= \mu^{-1}\sum_{n=1}^{N}\sum_{\ell=L+1}^{\infty}\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^\ell \boldsymbol{v}_{n;K;X_N}\otimes \boldsymbol{v}_{n;K;X_N} \\ &= \sum_{n=1}^{N}\mu^{-1}\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{L+1}\frac{\lambda_{n;k;X_N}+\mu}{\lambda_{n;K;X_N}}\boldsymbol{v}_{j;K;X_N}\otimes \boldsymbol{v}_{j;K;X_N} \\ &= \sum_{n=1}^{N}\lambda_{n;K;X_N}^{-1}\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{L}\boldsymbol{v}_{n;K;X_N}\otimes \boldsymbol{v}_{n;K;X_N}\end{aligned} \tag{24}$$

Combining (9) and (24), yields

$$\boldsymbol{E}_{X_N,X_N}^{(L)}\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}_{X_N,X_N}^{(L)} = \sum_{n=1}^{N}\lambda_{n;K;X_N}^{-1}\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^{2L}\boldsymbol{v}_{n;K;X_N}\otimes \boldsymbol{v}_{n;K;X_N}. \tag{25}$$

To see which is the largest eigenvalue of the matrix we recall the function (18) and get

$$f_{\mu;2L}(x) = x^{-1}\left(\frac{\mu}{x+\mu}\right)^{2L}.$$

Since $f_{\mu,L}$ is decreasing (see (19)), we obtain

$$\begin{aligned}\lambda_{\max}\left(\boldsymbol{E}^{(L)}_{X_N,X_N}\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}^{(L)}_{X_N,X_N}\right) &= f_{\mu;2L}(\lambda_{\min}(\boldsymbol{K}_{X_N,X_N}))\\ &= \lambda^{-1}_{\min}(\boldsymbol{K}_{X_N,X_N})\left(\frac{\mu}{\mu+\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^{2L}.\end{aligned}$$

This implies

$$\begin{aligned}\mathcal{E}^2_{\mathcal{H}}(\boldsymbol{y}) &= \boldsymbol{y}^\top\boldsymbol{E}^{(L)}_{X_N,X_N}\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}^{(L)}_{X_N,X_N}\boldsymbol{y}\\ &\le \lambda^{-1}_{\min}(\boldsymbol{K}_{X_N,X_N})\left(\frac{\mu}{\mu+\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^{2L}\|\boldsymbol{y}\|^2_{\ell^2(\mathbb{R}^N)}.\end{aligned} \tag{26}$$

Moreover, we have for a given $f\in\mathcal{H}_K(\Omega)$

$$\mathcal{E}^2_{\ell^2}(\boldsymbol{y}) := \left\|\left(\mathrm{Id}-\boldsymbol{K}_{X_N,X_N}\boldsymbol{C}^{(L)}_{X_N,X_N}\right)\boldsymbol{y}\right\|^2_{\ell^2(\mathbb{R}^N)} = \left\|\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}^{(L)}_{X_N,X_N}\boldsymbol{y}\right\|^2_{\ell^2(\mathbb{R}^N)}.$$

We get

$$\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}^{(L)}_{X_N,X_N} = \sum_{n=1}^N\left(\frac{\mu}{\lambda_{n;K;X_N}+\mu}\right)^L\boldsymbol{v}_{n;K;X_N}\otimes\boldsymbol{v}_{n;K;X_N}.$$

We consider for $\mu>0$ and $L\in\mathbb{N}$

$$\tilde{g}_{\mu;L}:(\lambda_{\min}(\boldsymbol{K}_{X_N,X_N}),\lambda_{\max}(\boldsymbol{K}_{X_N,X_N}))\to\mathbb{R},\quad x\mapsto g_{\mu;L}(x) := \left(\frac{\mu}{x+\mu}\right)^L.$$

Direct calculations yield

$$\frac{d}{dx}\tilde{g}_{\mu;L}(x) = -\frac{L}{(x+\mu)}\left(\frac{\mu}{x+\mu}\right)^L < 0$$

and hence

$$\begin{aligned}\lambda_{\max}\left(\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}^{(L)}_{X_N,X_N}\right) &= g_{\mu;L}(\lambda_{\min}(\boldsymbol{K}_{X_N,X_N}))\\ &= \left(\frac{\mu}{\mu+\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^L.\end{aligned}$$

This implies

$$\begin{aligned}\mathcal{E}^2_{\ell^2}(\boldsymbol{y}) &= \boldsymbol{y}^\top\left(\boldsymbol{E}^{(L)}_{X_N,X_N}\boldsymbol{K}_{X_N,X_N}\right)^\top\boldsymbol{K}_{X_N,X_N}\boldsymbol{E}^{(L)}_{X_N,X_N}\boldsymbol{y}\\ &\le \left(\frac{\mu}{\mu+\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^{2L}\|\boldsymbol{y}\|^2_{\ell^2(\mathbb{R}^N)}.\end{aligned} \tag{27}$$

□

3. Error Estimates for the function approximation

In this section, we provide an estimate of the approximation error based on Lemma 2. In order to describe the spatial distribution of the point set X_N, we use the fill-distance

$$h_{X_N;\Omega} := \sup_{x\in\Omega} \min_{x_n\in X_N} \|x - x_n\|_2 . \tag{28}$$

To transfer the discrete error from Lemma 2 to the function setting, we employ so-called sampling inequalities, see e.g., [1, 2, 3, 6, 7, 9, 11, 13, 14]. Those estimates are all of the form

$$\|u\|_{L^\infty(\Omega)} \le \rho_1(h_{X_N,\Omega}) \|u\|_{\mathcal{H}_K(\Omega)} + \rho_2(h_{X_N,\Omega}) \|S_{K,X_N} u\|_{\ell^p(\mathbb{R}^N)} \quad \text{for all } u \in \mathcal{H}_K(\Omega),$$

where $\rho_1(h_{X_N,\Omega}) \to 0$ as $h_{X_N;\Omega} \to 0$ and $\rho_2(h_{X_N,\Omega}) \to C_{N,p}$. There are also sampling inequalities expressed in the number of sampling points N, if for instance sparse grids are used as discrete sets X_N, see [10]. The following analysis does not depend on the specific form of those sampling inequalities and hence, we work with a generic sampling inequality.

Theorem 1. *We assume a generic* sampling inequality *of the form*

$$\|u\|_{L^\infty(\Omega)} \le \rho_1(h_{X_N,\Omega}) \|u\|_{\mathcal{H}_K(\Omega)} + \rho_2(h_{X_N,\Omega}) \|S_{X_N} u\|_{\ell^2(\mathbb{R}^N)} \tag{29}$$

for all $u \in \mathcal{H}_K(\Omega)$. Let $\boldsymbol{x}_{\boldsymbol{y}}^{(L)} \in \mathbb{R}^N$ be the L-th iterate stemming from (13)

$$\boldsymbol{x}_{\boldsymbol{y}}^{(0)} := \boldsymbol{0} \quad \text{and} \quad \boldsymbol{x}_{\boldsymbol{y}}^{(n+1)} := \boldsymbol{K}^{-1}_{\mu;X_N,X_N} \boldsymbol{y} + \mu \boldsymbol{K}^{-1}_{\mu;X_N,X_N} \boldsymbol{x}_{\boldsymbol{y}}^{(n)}, \quad \text{for } n > 0$$

where $\boldsymbol{y} = S_{K,X_N}(f) + \boldsymbol{\epsilon}$. Then, we get the following error estimate

$$\begin{aligned} \left\| S^\star_{X_N} \boldsymbol{x}_{\boldsymbol{y}}^{(L)} - f \right\|_{L^\infty(\Omega)} &\le \left(q^L_{\mu,K,X_N} C(\rho_{1,2}, X_N, K) + 2\rho_1(h_{X_N,\Omega})\right) \|f\|_{\mathcal{H}_K(\Omega)} \\ &\quad + \left(\left(q^L_{\mu,K,X_N} + 1\right) C(\rho_{1,2}, X_N, K) + \rho_2(h_{X_N,\Omega})\right) \epsilon_\infty, \end{aligned} \tag{30}$$

where we used the abbreviations

$$q_{\mu,K,X_N} := \frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} \in (0,1) \tag{31}$$

and

$$C(\rho_{1,2}, X_N, K) := \sqrt{N} \left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega}) \right). \tag{32}$$

Proof. We denote by $\boldsymbol{x}_{\boldsymbol{y}}, \boldsymbol{x}^\star_f \in \mathbb{R}^N$ the solutions to the linear systems

$$\boldsymbol{K}_{X_N,X_N} \boldsymbol{x}_{\boldsymbol{y}} = \boldsymbol{y} \quad \text{and} \quad \boldsymbol{K}_{X_N,X_N} \boldsymbol{x}^\star_f = S_{X_N} f. \tag{33}$$

We note that these vectors are only for theoretical purposes and not numerically not available. We also consider the associated approximation corresponding to the vectors from (33), i.e., we consider

$$S^\star_{X_N} \boldsymbol{x}_{\boldsymbol{y}} \quad \text{and} \quad S^\star_{X_N} \boldsymbol{x}^\star_f, \tag{34}$$

which are the interpolant to the inexact and the exact data. Please note that the latter is the usual kernel based interpolant to the function f. We split the overall error

$$\begin{aligned}\left\| S^\star_{X_N}\boldsymbol{x}^{(L)}_{\boldsymbol{y}} - f\right\|_{L^\infty(\Omega)} \leq & \overbrace{\left\| S^\star_{X_N}\boldsymbol{x}^{(L)}_{\boldsymbol{y}} - S^\star_{X_N}\boldsymbol{x}_{\boldsymbol{y}}\right\|_{L^\infty(\Omega)}}^{E_{\text{iter}}} + \overbrace{\left\| S^\star_{X_N}\boldsymbol{x}_{\boldsymbol{y}} - S^\star_{X_N}\boldsymbol{x}^\star_f\right\|_{L^\infty(\Omega)}}^{E_{\text{data}}} \\ & + \underbrace{\left\| S^\star_{X_N}\boldsymbol{x}^\star_f - f\right\|_{L^\infty(\Omega)}}_{E_{\text{interpol}}}. \end{aligned} \tag{35}$$

We consider the terms on the right hand side separately. First, we get for E_{iter} from (35) inserting (26) and (27) into (29)

$$\begin{aligned} E_{\text{iter}} &= \left\| S^\star_{K;X_N}\boldsymbol{x}^{(L)}_{\boldsymbol{y}} - S^\star_{K;X_N}\boldsymbol{x}_{\boldsymbol{y}}\right\|_{L^\infty(\Omega)} \leq \rho_1(h_{X_N,\Omega}) \left\| S^\star_{K;X_N}\boldsymbol{x}^{(L)}_{\boldsymbol{y}} - S^\star_{K;X_N}\boldsymbol{x}_{\boldsymbol{y}}\right\|_{\mathcal{H}_K(\Omega)} \\ &\quad + \rho_2(h_{X_N,\Omega}) \left\| \boldsymbol{K}_{X_N,X_N}\boldsymbol{x}^{(L)}_{\boldsymbol{y}} - S_{X_N} f\right\|_{\ell^2(\mathbb{R}^N)} \\ &\leq \rho_1(h_{X_N,\Omega})\mathcal{E}_{\mathcal{H}}(\boldsymbol{y}) + \rho_2(h_{X_N,\Omega})\mathcal{E}_{\ell^2}(\boldsymbol{y}) \\ &\leq \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^L \left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega})\right) \|\boldsymbol{y}\|_{\ell^2(\mathbb{R}^N)} \\ &\leq \sqrt{N}\left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^L \left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega})\right) \left(\|f\|_{\mathcal{H}_K(\Omega)} + \epsilon_\infty\right). \end{aligned}$$

For the second error term E_{data} from (35), we observe

$$\begin{aligned} \left\| S^\star_{K;X_N}\boldsymbol{x}_{\boldsymbol{y}} - S^\star_{K;X_N}\boldsymbol{x}^\star_f\right\|^2_{\mathcal{H}_K(\Omega)} &= (\boldsymbol{y} - S_{X_N} f)^\top \boldsymbol{K}^{-1}_{X_N,X_N} (\boldsymbol{y} - S_{X_N} f) = \boldsymbol{\epsilon}^\top \boldsymbol{K}^{-1}_{X_N,X_N}\boldsymbol{\epsilon} \\ &\leq \lambda^{-1}_{\min}(\boldsymbol{K}_{X_N,X_N}) \|\boldsymbol{\epsilon}\|^2_{\ell^2(\mathbb{R}^N)} \leq \frac{N\epsilon^2_\infty}{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} \end{aligned}$$

and by construction

$$\left\| S_{K;X_N}\left(S^\star_{K;X_N}\boldsymbol{x}_{\boldsymbol{y}} - S^\star_{K;X_N}\boldsymbol{x}^\star_f\right)\right\|^2_{\ell^2(\mathbb{R}^N)} = \|\boldsymbol{\epsilon}\|^2_{\ell^2(\mathbb{R}^N)} \leq N\epsilon^2_\infty.$$

Hence, using again (29), we obtain

$$E_{\text{data}} = \left\| S^\star_{K;X_N}\boldsymbol{x}_{\boldsymbol{y}} - S^\star_{K;X_N}\boldsymbol{x}^\star_f\right\|_{L^\infty(\Omega)} \leq \epsilon_\infty\sqrt{N}\left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega})\right).$$

The third error contribution E_{iter} from (35) is the most straight forward to analyze and we get

$$E_{\text{interpol}} = \left\| S^\star_{K;X_N}\boldsymbol{x}^\star_f - f\right\|_{L^\infty(\Omega)} \leq 2\rho_1(h_{X_N,\Omega}) \|f\|_{\mathcal{H}_K(\Omega)}.$$

Finally, we obtain that for all $L \in \mathbb{N}$ the total error as in (35) is bounded as

$$\left\| S^{\star}_{K;X_N} \boldsymbol{x}_{\boldsymbol{y}}^{(L)} - f \right\|_{L^\infty(\Omega)} \leq E_{\text{iter}} + E_{\text{data}} + E_{\text{interpol}} = E_{\text{upper}} =$$

$$\left(\left(q^L_{\mu,K,X_N} \sqrt{N} \left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega}) \right) \right) + 2\rho_1(h_{X_N,\Omega}) \right) \|f\|_{\mathcal{H}_K(\Omega)}$$

$$+ \left(\left(q^L_{\mu,K,X_N} + 1 \right) \sqrt{N} \left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega}) \right) \right) \epsilon_\infty. \tag{36}$$

Using the abbreviations (31) and (32) the error estimate (36) reduces to (30). This finishes the proof. □

We would like to choose a number of iterations to obtain a quasi-optimal error estimate. Such a choice is provided by the following corollary.

Corollary 1. *Let the constant* $C(\rho_{1,2}, X_N, K)$ *from* (32) *satisfy*

$$C(\rho_{1,2}, X_N, K) = \sqrt{N} \left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega}) \right) > 1. \tag{37}$$

If we choose

$$\tilde{L}_\mu = \left\lceil \frac{\ln\left(\rho_1(h_{X_N,\Omega})\right) - \ln\left(\sqrt{N} \left(\frac{\rho_1(h_{X_N,\Omega})}{\lambda^{\frac{1}{2}}_{\min}(\boldsymbol{K}_{X_N,X_N})} + \rho_2(h_{X_N,\Omega}) \right) \right)}{\ln\left(q_{\mu,K,X_N}\right)} \right\rceil, \tag{38}$$

then, we obtain an error estimate of the form

$$\left\| S^{\star}_{X_N} \boldsymbol{x}_{\boldsymbol{y}}^{(\tilde{L}_\mu)} - f \right\|_{L^\infty(\Omega)} \leq 2C(\rho_{1,2}, X_N, K)\epsilon_\infty + 4\rho_1(h_{X_N,\Omega}) \|f\|_{\mathcal{H}_K(\Omega)}. \tag{39}$$

Note that the assumption (37) is trivially satisfied if $C_{N,p} \sim 1$.

Proof. We observe that the choice (38), i.e., $L = \tilde{L}_\mu$ yields

$$q^{\tilde{L}}_{\mu,K,X_N} C(\rho_{1,2}, X_N, K) \leq \rho_1(h_{X_N,\Omega}).$$

Observing $\rho_1(h_{X_N,\Omega}) < 1 \sim C(\rho_{1,2}, X_N, K)$ due to (37) finishes the proof. □

In order to take the numerical costs into account, we have to bound the number of iterations. We observe that due to assumption (37) and the fact that $0 < q_{\mu,K,X_N} < 1$, the expression for $\tilde{L}$ is positive. Hence, we choose for $L^\star = \lceil \tilde{L} \rceil \in \mathbb{N}$. We also observe that

$$f : (0,1) \to (0,1), \quad \mu \mapsto q_{\mu,K,X_N} := \frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}$$

satisfies

$$f'(\mu) = \frac{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}{(\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N}))^2} \geq 0$$

and hence, we obtain that $0 = \inf_{\mu\in(0,1)} f(\mu) = \lim_{\mu\to 0} f(\mu)$. Hence, considering only the number of iterations would suggest to choose a very small μ. But, then we loose the stabilizing effect that $\boldsymbol{K}_{\mu;X_N,X_N} := \boldsymbol{K}_{X_N,X_N} + \mu\,\mathrm{Id}_{N\times N}$ is better conditioned than $\boldsymbol{K}_{X_N,X_N}$. At this point, there are several choices for μ possible. The practical approach is to choose μ large enough to make $\boldsymbol{K}_{\mu;X_N,X_N}$ numerically positive definite which means that the smallest eigenvalue is far enough away from zero. Then, one can use an ordinary Cholesky decomposition of $\boldsymbol{K}_{\mu;X_N,X_N}$ in the iteration (13). The problem with this approach is that the errors due to the floating point arithmetic can accumulate during the iteration. Furthermore, the Cholesky factors inherit the conditioning issues from the original matrix and hence the solution of the linear system in each step is still ill-conditioned. We therefore propose to use the inverse Cholesky decomposition which is developed for ill conditioned linear systems in a finite floating point arithmetic.

4. Numerical Linear Algebra in the presence of finite precision

The main observation is the fact that in the iteration (13)

$$\boldsymbol{x}_{\boldsymbol{b}}^{(0)} := \boldsymbol{0} \quad \text{and} \quad \boldsymbol{x}_{\boldsymbol{b}}^{(n+1)} := \Phi_{\boldsymbol{b}}\left(\boldsymbol{x}_{\boldsymbol{b}}^{(n)}\right) = \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b} + \mu \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{x}_{\boldsymbol{b}}^{(n)}, \quad \text{for } n > 0$$

we have to solve a linear system with matrix $\boldsymbol{K}_{\mu;X_N,X_N}$ in every iteration step. This motivates to precompute a decomposition of the matrix $\boldsymbol{K}_{\mu;X_N,X_N}$ or $\boldsymbol{K}_{\mu;X_N,X_N}^{-1}$ and to use this decomposition to speed up the calculation in every iteration step. Since we need to invert the matrix $\boldsymbol{K}_{\mu;X_N,X_N}$ in every iteration step and since this matrix is symmetric positive definite for all $\mu \geq 0$, we can use a Cholesky decomposition. We face, however, the problem of the high condition number in (8) and hence the direct decompositions of $\boldsymbol{K}_{\mu;X_N,X_N}$ might break down for too small values of μ. In particular, we have with (8)

$$\mathrm{cond}(\boldsymbol{K}_{\mu;X_N,X_N}) \approx \mathrm{cond}(\boldsymbol{K}_{X_N,X_N}) \geq \alpha^{-1}$$

for $\mu < \alpha$. For the numerical computations, we fix an error level $E_{\mathrm{tol}} > 0$. We first discuss the Cholesky decomposition. We note that $\boldsymbol{K}_{\mu;X_N,X_N} \in \mathbb{R}^{N\times N}$ is symmetric and that we have $\boldsymbol{e}_i^\top \boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{e}_i = K(x_i,x_i) + \mu > 0$. Having these properties we can apply a result from [15, Algorithm 1] to obtain for an accuracy

$$N^2\alpha < E_{\mathrm{tol}} < 1 \tag{40}$$

an upper triangular matrix $\boldsymbol{U} \in \mathbb{F}_{[m]}^{N\times N}$ with $m \in \mathbb{N}$ such that

$$\left\|\boldsymbol{U}^\top \boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U} - \mathbf{Id}_{N\times N}\right\|_{\ell_2\to\ell_2} < E_{\mathrm{tol}}. \tag{41}$$

Note that (40) limits the number of points N if we fix an error tolerance E_{tol} since α is the constant from (6). Having such a decomposition, we can use this to solve a linear system

$$\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{x} = \boldsymbol{\beta}. \tag{42}$$

We get

$$\begin{aligned}\boldsymbol{U}^\top\boldsymbol{\beta} &= \boldsymbol{U}^\top \boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{x} = \boldsymbol{U}^\top \boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U}\boldsymbol{U}^{-1}\boldsymbol{x}\\ &= \left(\boldsymbol{U}^\top \boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U} - \mathbf{Id}_{N\times N}\right)\boldsymbol{U}^{-1}\boldsymbol{x} + \boldsymbol{U}^{-1}\boldsymbol{x}\end{aligned}$$

which yields

$$\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\beta} = \boldsymbol{x} + \boldsymbol{U}\left(\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U} - \mathbf{Id}_{N\times N}\right)\boldsymbol{U}^{-1}\boldsymbol{x} \approx \boldsymbol{x}.$$

We hence get for the relative error

$$\frac{\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\beta} - \boldsymbol{K}^{-1}_{\mu;X_N,X_N}\boldsymbol{\beta}\right\|_{\ell_2}}{\|\boldsymbol{x}\|_{\ell_2}} = \frac{\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\beta} - \boldsymbol{x}\right\|_{\ell_2}}{\|\boldsymbol{x}\|_{\ell_2}} \leq \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N} - \mathbf{Id}_{N\times N}\right\|_{\ell_2\to\ell_2}.$$

We compute

$$\begin{aligned}
&\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N} - \mathbf{Id}_{N\times N}\right\|_{\ell_2\to\ell_2} = \left\|\boldsymbol{U}\left(\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U} - \mathbf{Id}_{N\times N}\right)\boldsymbol{U}^{-1}\right\|_{\ell_2\to\ell_2}\\
&= \left\|\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U}\left(\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U} - \mathbf{Id}_{N\times N}\right)\boldsymbol{U}^{-1}\right\|_{\ell_2\to\ell_2}\\
&\leq \left\|\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\right\|_{\ell_2\to\ell_2}\left\|\boldsymbol{U}^{-\top}\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U}\left(\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U} - \mathbf{Id}_{N\times N}\right)\boldsymbol{U}^{-1}\right\|_{\ell_2\to\ell_2}\\
&\leq \left\|\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\right\|_{\ell_2\to\ell_2}\left\|\boldsymbol{U}^{-1}\right\|^2_{\ell_2\to\ell_2}\left\|\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U}\right\|_{\ell_2\to\ell_2} E_{\mathrm{tol}}.
\end{aligned}$$

Using the arguments from [15, Analysis 2], we get

$$\left\|\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N}\boldsymbol{U}\right\|_{\ell_2\to\ell_2} \approx 1 \quad \text{and} \quad \left\|\boldsymbol{U}^{-1}\right\|^2_{\ell_2\to\ell_2} \approx \left\|\boldsymbol{K}_{\mu;X_N,X_N}\right\|_{\ell_2\to\ell_2}.$$

This leaves us with

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{K}_{\mu;X_N,X_N} - \mathbf{Id}_{N\times N}\right\|_{\ell_2\to\ell_2} \lesssim \mathrm{cond}_2\left(\boldsymbol{K}_{\mu;X_N,X_N}\right)E_{\mathrm{tol}} \lesssim \left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)E_{\mathrm{tol}}.$$

Hence, we get

(43)

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\beta} - \boldsymbol{K}^{-1}_{\mu;X_N,X_N}\boldsymbol{\beta}\right\|_{\ell_2} \leq \left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)E_{\mathrm{tol}}\left\|\boldsymbol{K}^{-1}_{\mu;X_N,X_N}\boldsymbol{\beta}\right\|_{\ell_2} \quad \text{for all } \beta \in \mathbb{R}^N.$$

Similar arguments apply for the approximate inverse QR decomposition as in [8].

4.1. **Error analysis for iteration.** In this section we bound the error influence of the inexact iteration. Again, we start with the errors measured in discrete norms.

Lemma 3. *We assume*

$$\mu \in (0,1) \quad \textit{and} \quad N \geq 3. \tag{44}$$

Then, we get for the error $\boldsymbol{e}^{(m)}_b := \boldsymbol{x}^{(m)}_{\boldsymbol{b}} - \boldsymbol{u}^{(m)}_{\boldsymbol{b}}$

$$\mu\left\|\boldsymbol{e}^{(m+1)}_{\boldsymbol{b}}\right\|_{\ell_2} \leq \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}E_{tol}\right)^m\left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)E_{tol}\|\boldsymbol{b}\|_{\ell_2} +$$

(45)

$$+\frac{(\mu+1)E_{tol}}{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^L\right)\left(1 - \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}E_{tol}\right)^m\right),$$

for $m \geq 1$ and with $\boldsymbol{x}_{\boldsymbol{b}}^{(m)}$ given by the iteration (13) *and $\boldsymbol{u}_{\boldsymbol{b}}^{(m)}$ through the modified iteration*

$$\boldsymbol{u}_{\boldsymbol{b}}^{(0)} := \boldsymbol{0} \quad \text{and} \quad \boldsymbol{u}_{\boldsymbol{b}}^{(m+1)} := \Phi_{\boldsymbol{U};\boldsymbol{b}}\left(\boldsymbol{u}_{\boldsymbol{b}}^{(m)}\right) = \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{b} + \mu \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{u}_{\boldsymbol{b}}^{(m)}, \quad \text{for } m > 0. \tag{46}$$

Proof. For the error $\boldsymbol{e}_b^{(n)}$, we observe $\boldsymbol{e}_{\boldsymbol{b}}^{(0)} = \boldsymbol{0}$ and

$$\begin{aligned}\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(1)}\right\|_{\ell_2} &= \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{b} - \boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b}\right\|_{\ell_2} \leq \left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right) E_{\mathrm{tol}} \left\|\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{b}\right\|_{\ell_2} \\ &\leq \mu^{-1}\left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right) E_{\mathrm{tol}} \left\|\boldsymbol{b}\right\|_{\ell_2}, \end{aligned} \tag{47}$$

where we use (43) and (47). Next, we obtain for any $m \geq 1$

$$\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m+1)}\right\|_{\ell_2} = \left\|\boldsymbol{x}_{\boldsymbol{b}}^{(m+1)} - \boldsymbol{u}_{\boldsymbol{b}}^{(m+1)}\right\|_{\ell_2} \leq \left\|\boldsymbol{e}_{\boldsymbol{b}}^{(1)}\right\|_{\ell_2} + \mu\left\|\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{x}_{\boldsymbol{b}}^{(m)} - \boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{u}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2}. \tag{48}$$

For the second term, we get for any $m \geq 1$ using (43)

$$\begin{aligned}&\left\|\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\boldsymbol{x}_{\boldsymbol{b}}^{(m)} - \boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{u}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} \leq \\ &\leq \left\|\left(\boldsymbol{K}_{\mu;X_N,X_N}^{-1} - \boldsymbol{U}\boldsymbol{U}^\top\right)\boldsymbol{u}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} + \left\|\boldsymbol{K}_{\mu;X_N,X_N}^{-1}\left(\boldsymbol{x}_{\boldsymbol{b}}^{(m)} - \boldsymbol{u}_{\boldsymbol{b}}^{(m)}\right)\right\|_{\ell_2} \\ &\leq \mu^{-1}\left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right) E_{\mathrm{tol}}\left\|\boldsymbol{u}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} + \mu^{-1}\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} \\ &\leq \mu^{-1}\left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}E_{\mathrm{tol}}\right)\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} + \mu^{-1}\left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right) E_{\mathrm{tol}}\left\|\boldsymbol{x}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2}.\end{aligned}$$

In order the get a bound on $\left\|\boldsymbol{x}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2}$ we use Lemma 1 to get

$$\begin{aligned}\left\|\boldsymbol{x}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} &= \left\|\boldsymbol{C}_{X_N,X_N}^{(m)}\boldsymbol{b}\right\|_{\ell_2} \leq \left\|\boldsymbol{C}_{X_N,X_N}^{(m)}\right\|_{\ell_2\to\ell_2}\left\|\boldsymbol{b}\right\|_{\ell_2} \\ &\leq \lambda_{\min}^{-1}(\boldsymbol{K}_{X_N,X_N})\left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^m\right)\left\|\boldsymbol{b}\right\|_{\ell_2}.\end{aligned}$$

Inserting the bounds into (48) implies that

$$\begin{aligned}&\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m+1)}\right\|_{\ell_2} \leq \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}E_{\mathrm{tol}}\right)\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} \\ &+ \left(1 + \frac{1}{\mu}\right)\left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)\frac{E_{\mathrm{tol}}}{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^m\right)\left\|\boldsymbol{b}\right\|_{\ell_2}.\end{aligned}$$

Schematically, we have the following situation:

$$\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m+1)}\right\|_{\ell_2} \leq A\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2} + B,$$

with the choices

$$A := \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} E_{\mathrm{tol}}\right) > 1$$

$$B := \left(1 + \frac{1}{\mu}\right)\left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right) \frac{E_{\mathrm{tol}}}{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} \left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^m\right).$$

Iterating this inequality yields

$$\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m+1)}\right\|_{\ell_2} \leq A^m \left\|\boldsymbol{e}_{\boldsymbol{b}}^{(1)}\right\|_{\ell_2} + B \sum_{k=0}^{m-1} A^k = A^m \left\|\boldsymbol{e}_{\boldsymbol{b}}^{(1)}\right\|_{\ell_2} + B \frac{1-A^m}{1-A},$$

where we used the formulas for the geometric series. Hence, we obtain

$$\mu \left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m+1)}\right\|_{\ell_2} \leq \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} E_{\mathrm{tol}}\right)^m \left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right) E_{\mathrm{tol}} \|\boldsymbol{b}\|_{\ell_2} +$$
$$+ \frac{(\mu+1)}{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} \left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^m\right) \left(1 - \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} E_{\mathrm{tol}}\right)^m\right),$$

which concludes the proof. □

4.2. **Error analysis for numerical linear algebra.** As in Section 3, we can bound the influence of the discrete (numerical linear algebra) error on the continuous function approximation. Using (29), we get

Corollary 2. *Under the assumption of Corollary 1 and Lemma 3, we get for $m \geq \tilde{L}_\mu$ from* (38) *the bound*

$$\left\|S_{X_N}^{\star} \boldsymbol{u}_{\boldsymbol{y}}^{(m)} - f\right\|_{L^\infty(\Omega)} \leq 2C(\rho_{1,2}, X_N, K)\epsilon_\infty + 4\rho_1(h_{X_N,\Omega}) \|f\|_{\mathcal{H}_K(\Omega)}$$
$$+ \mu^{-1} \left(\sqrt{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}\rho_1(h_{X_N,\Omega}) + \lambda_{\max}(\boldsymbol{K}_{X_N,X_N})\rho_2(h_{X_N,\Omega})\right) \times$$
$$\times \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} E_{tol}\right)^m \left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right) E_{tol} \|S_{X_N} f\|_{\ell_2} +$$
$$+ \frac{(\mu+1)}{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} \left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^m\right) \left(1 - \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})} E_{tol}\right)^m\right)$$

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

Proof. We use the abbreviation $\boldsymbol{b} = S_{X_N} f$. We bound the terms separately to get

$$\left\|S_{X_N}^{\star} \boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\mathcal{H}_K(\Omega)}^2 = \left(\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right)^\top \boldsymbol{K}_{X_N,X_N} \boldsymbol{e}_{\boldsymbol{b}}^{(m)} \leq \lambda_{\max}\left(\boldsymbol{K}_{X_N,X_N}\right) \left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2}^2.$$

Moreover, we get

$$\left\|S_{X_N} S_{X_N}^{\star} \boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell^2(\mathbb{R}^N)} \leq \lambda_{\max}\left(\boldsymbol{K}_{X_N,X_N}\right) \left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m+}\right\|_{\ell_2}.$$

We have the ingredients to apply the generic sampling inequality to get

$$\begin{aligned}\left\|S^\star_{X_N}\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{L^\infty(\Omega)} &= \left\|S^\star_{X_N}\boldsymbol{x}_{\boldsymbol{b}}^{(m)} - S^\star_{X_N}\boldsymbol{u}_{\boldsymbol{b}}^{(m)}\right\|_{L^\infty(\Omega)}\\ &\leq \rho_1(h_{X_N,\Omega})\left\|S^\star_{X_N}\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\mathcal{H}_K(\Omega)} + \rho_2(h_{X_N,\Omega})\left\|S_{X_N}S^\star_{X_N}\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell^2(\mathbb{R}^N)}.\end{aligned}$$

Hence, we end up with

$$\left\|S^\star_{X_N}\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{L^\infty(\Omega)} \leq \left(\sqrt{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}\rho_1(h_{X_N,\Omega}) + \lambda_{\max}(\boldsymbol{K}_{X_N,X_N})\rho_2(h_{X_N,\Omega})\right)\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2}$$

Since we choose $m \geq \tilde{L}_\mu$ from (38), we obtain by triangle inequality

$$\begin{aligned}&\left\|S^\star_{X_N}\boldsymbol{u}_{\boldsymbol{y}}^{(m)} - f\right\|_{L^\infty(\Omega)} \leq \left\|S^\star_{X_N}\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{L^\infty(\Omega)} + \left\|S^\star_{X_N}\boldsymbol{x}_{\boldsymbol{y}}^{(m)} - f\right\|_{L^\infty(\Omega)}\\ &\leq 2C(\rho_{1,2}, X_N, K)\epsilon_\infty + 4\rho_1(h_{X_N,\Omega})\|f\|_{\mathcal{H}_K(\Omega)}\\ &+ \left(\sqrt{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}\rho_1(h_{X_N,\Omega}) + \lambda_{\max}(\boldsymbol{K}_{X_N,X_N})\rho_2(h_{X_N,\Omega})\right)\left\|\boldsymbol{e}_{\boldsymbol{b}}^{(m)}\right\|_{\ell_2}\\ &\leq 2C(\rho_{1,2}, X_N, K)\epsilon_\infty + 4\rho_1(h_{X_N,\Omega})\|f\|_{\mathcal{H}_K(\Omega)}\\ &+ \mu^{-1}\left(\sqrt{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}\rho_1(h_{X_N,\Omega}) + \lambda_{\max}(\boldsymbol{K}_{X_N,X_N})\rho_2(h_{X_N,\Omega})\right)\times\\ &\times\left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}E_{\text{tol}}\right)^m\left(\frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)E_{\text{tol}}\|\boldsymbol{b}\|_{\ell_2} +\\ &+ \frac{(\mu+1)}{\lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\left(1 - \left(\frac{\mu}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}\right)^m\right)\left(1 - \left(1 + \frac{\lambda_{\max}(\boldsymbol{K}_{X_N,X_N})}{\mu + \lambda_{\min}(\boldsymbol{K}_{X_N,X_N})}E_{\text{tol}}\right)^m\right),\end{aligned}$$

where we used (39) in the last step. □

The last result shows that we have to choose large values for μ to get a reasonable error. Such error estimates are of course far from being sharp since they assume in each step the worst possible case.

5. Conclusion

We analyzed the iterative Landweber regularization technique to solve kernel-based reconstruction problems with inexact data. We derived deterministic a priori error bounds which are explicit in the various parameters. Since it turned out that from a pure error analysis perspective small regularization appears to be favorable we demonstrated that a finite precision arithmetic might produce small errors in each step of the Landweber iteration which, however, might in a worst case scenario blow up. Those accumulating errors make large regularization parameters beneficial. We consider this worst case scenario as a first step towards a more realistic averaged case analysis is still work in progress. Nevertheless we believe that such deterministic error estimates will gain more and more importance as computations can be made accurate enough to make also the machine precision an important quantity.

Acknowledgement

The author acknowledges partial support of the Deutsche Forschungsgemeinschaft (DFG) through the Sonderforschungsbereich 1060: The Mathematics of Emergent Effects.

References

[1] R. Arcangéli, M. C. López di Silanes, and J. J. Torrens. An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing. *Numer. Math.*, 107(2):181–211, 2007.

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

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

[4] G. Fasshauer. Dealing with Ill-Conditioned RBF Systems. *Dolomites Research Notes on Approximation*, 1, 2008. published slides.

[5] G. Golub. Numerical methods for solving linear least squares problems. *Numerische Mathematik*, 7(3):206–216, 1965.

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

[7] F. J. Narcowich, J. D. 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.

[8] T. Ogita. Accurate Matrix Factorization: Inverse LU and Inverse QR Factorizations. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2477–2497, 2010.

[9] 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.

[10] C. Rieger and H. Wendland. Sampling inequalities for sparse grids. *Numerische Mathematik*, 136(2):439–466, Jun 2017.

[11] C. Rieger and B. Zwicknagl. Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning. *Adv. Comp. Math.*, 32(1):103–129, 2010.

[12] S. M. Rump. Inversion of Extremely Ill-Conditioned Matrices in Floating-Point. *Japan J. Indust. Appl. Math.*, 26(2-3):249–277, 10 2009.

[13] F. Utreras. Convergence rates for multivariate smoothing spline functions. *J. Approx. Theory*, 52:1–27, 1988.

[14] H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.*, 101:729–748, 2005.

[15] Y. Yanagisawa, T. Ogita, and S. Oishi. Convergence analysis of an algorithm for accurate inverse Cholesky factorization. *Japan Journal of Industrial and Applied Mathematics*, 31(3):461–482, 2014.