

Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

vorgelegt von
Stefan Müller
aus Gießen

Göttingen 2008

D7

Referent: Prof. Dr. Robert Schaback

Koreferent: Prof. Dr. Gert Lube

Tag der mündlichen Prüfung: 21.01.2009

Danksagungen

An dieser Stelle möchte ich verschiedenen Personen danken, die mich besonders beim Schreiben der Doktorarbeit unterstützt haben.

Vor allem möchte ich meinem Betreuer Prof. Dr. Schaback für die hervorragende Betreuung danken. Durch seine vielen wertvollen Ideen war es erst möglich, dass die Arbeit in dieser Form entstanden ist. Er hat stets mit sehr viel Geduld und Ausdauer meine Fragen beantwortet und es immer verstanden, die Mathematik sehr lebendig zu vermitteln.

Danken möchte ich auch Koreferent Prof. Dr. Lube für die Begutachtung der Arbeit.

Barbara Zwicknagl und Dr. Christian Rieger danke ich für die sorgfältige Korrekturlesung meiner Arbeit und die zahlreichen Verbesserungsvorschläge.

Ebenfalls danken möchte ich den anderen Doktoranden, ohne die es im Institut sehr einsam gewesen wäre.

Der Deutschen Forschungsgemeinschaft möchte ich für ihre Unterstützung im Rahmen des Graduiertenkolleg 1023 'Identifikation in mathematischen Modellen' danken.

Prof. Dr. Schöbel und Michael Schachtebeck waren mir eine grosse Hilfe bei der Verwendung von Xpress[©] zur Lösung des Optimierungsproblems in Abschnitt [6.4](#).

Schließlich danke ich meinen Eltern, dass sie mir dieses Studium ermöglicht haben.

Inhaltsverzeichnis

1	Einleitung	1
2	Interpolation mit Kernfunktionen	3
2.1	Grundlegende Begriffe	3
2.2	Resultate aus der Native Space Theorie	6
3	Konvergenzeigenschaften des f-Greedy-Verfahrens	13
3.1	Greedy-Verfahren	13
3.2	f -Greedy bei Kernfunktionen	15
3.3	Konvergenzaussagen für f -Greedy	17
3.4	Höhere Konvergenzordnungen bei f -Greedy in \mathbb{R}	26
4	Eine Newton-Basis des Native Space	30
4.1	Einführung	30
4.2	Die Newton-Basis	32
4.3	Eigenschaften der Newton-Basis	36
4.4	Reihenentwicklungen	39
4.5	Reihenentwicklung des Kerns mit der Powerfunktion	40
5	Die Newton-Basis mit Greedy-Punktwahl	43
5.1	Beziehungen zwischen der Powerfunktion und der Newton-Basis	43
5.2	Eigenschaften der Newton-Basis bei P -Greedy	44
5.3	Beziehungen zwischen der Determinante und der Powerfunktion	45
5.4	Fehlerabsch. für Greedy-Verfahren mit Hilfe der Newton-Basis	46
6	Tests verschiedener Greedy-Verfahren	48
6.1	Untersuchung der Punkteverteilungen	49
6.2	Fehlerkurven der Greedy-Methoden	51
6.3	Vergleich von Greedy-Verfahren zu Gitterpunkten	55
6.4	Vergleich von f -Greedy mit Optimallösung	61
6.5	Stabilität der Greedy-Verfahren	62

7 Tests zur Newton-Basis	64
7.1 Stabilitätsverhalten	64
7.2 Die Newton-Basis kombiniert mit <i>P</i> -Greedy	69
8 Ergebnisse und Ausblick	72
A Xpress[©]-Programm	74
Symbolverzeichnis	77
Literaturverzeichnis	79

Kapitel 1

Einleitung

Möchte man zu gegebenen Daten $Y = \{f(x_1), \dots, f(x_n)\} \subset \mathbb{R}$ an Punkten $X = \{x_1, \dots, x_n\} \subset \Omega \subset \mathbb{R}^d$ die Funktion f rekonstruieren, muss man sich zunächst überlegen, welche Eigenschaften die Rekonstruktion erfüllen soll. Als Minimaleigenschaft fordert man in der Regel, dass zu beliebigen Daten genau eine Lösung existiert. Für Raumdimension $d \geq 2$ ist das nach einem Satz von Mairhuber-Curtis (siehe Kapitel 2 von [22]) nur möglich, wenn der Ansatzraum, also der Raum der möglichen Rekonstruktionen, von X abhängt. Eine einfache Möglichkeit, dies zu erfüllen, besteht in der Verwendung sogenannter Kernfunktionen $K : \Omega \times \Omega \rightarrow \mathbb{R}$. Diese Funktionen haben den Namen daher, dass sie gleichzeitig reproduzierende Kerne eines Hilbertraumes, des sogenannten Native Space, sind. Zur Erzeugung der Interpolante benutzt man Translate $K(\cdot, x)$, $x \in X$, der Kernfunktion K . Darüber hinaus lässt sich zeigen, dass die so gewonnenen Interpolanten unter allen möglichen Interpolanten in dem dazugehörigen Hilbertraum die Daten mit minimaler Norm rekonstruieren.

Ist der Kern K positiv definit, lassen sich zu beliebigen Daten eindeutig Rekonstruktionen finden. Stammen diese Daten von einer Funktion des Native Space, bekommt man zusätzlich Abschätzungen für den dabei im schlimmsten Fall entstehenden Approximationsfehler (siehe [22]). Ein weiterer Vorteil dieser Methoden besteht darin, dass die Punkte beliebig im Raum verteilt sein dürfen und keine speziellen Punktgeometrien benötigen.

Als Motivation sollen zunächst ein paar Anwendungen kernbasierter Methoden genannt werden. Die Rekonstruktion mehrdimensionaler Daten wird zum Beispiel in der Oberflächenmodellierung benötigt (siehe [13, 10]), zur Berechnung von Strömungen und zur Modellierung geophysikalischer Daten (z.B. Wetterdaten). Auf der anderen Seite werden Kernmethoden aber auch für maschinelles Lernen (siehe [19, 9, 1]) verwendet, um zum Beispiel Klassifikationsprobleme zu lösen.

Allerdings gibt es auch Schwierigkeiten bei der Verwendung von Kernfunktionen. Um das Interpolationsproblem mit n Datenpunkten zu lösen, muss man nämlich ein Gleichungssystem mit n Unbekannten und n Bedingungen lösen. Damit liegt die Komple-

Kapitel 1: Einleitung

xität bei $O(n^3)$, falls die Kernfunktion keinen beschränkten Träger hat und direkte Methoden angewendet werden. Außerdem ist bei diesen Methoden in der Regel das Gleichungssystem sehr schlecht konditioniert. Besonders im Mehrdimensionalen, wo für eine hinreichende Genauigkeit viele Punkte benötigt werden, führt dies zu Komplexitäts- und Stabilitätsproblemen.

Je größer die Anzahl der Punkte ist, desto mehr Speicher und Zeit wird zunächst benötigt, um das zugehörige Gleichungssystem aufzustellen und zu lösen. Dies bezeichnet man als Ausrechnungskomplexität, weil dabei die Koeffizienten der Interpolante ausgerechnet werden. Bei steigender Punktzahl steigt aber auch der Aufwand, um anschließend den Wert der Interpolante an anderen Stellen zu bestimmen. Dabei spricht man von der Auswertungskomplexität. In dieser Arbeit untersuchen wir anhand von sogenannten Greedy-Verfahren, wie durch geeignete Punktwahl die Anzahl der Interpolationspunkte reduziert werden kann, ohne dass die Reproduktion zu stark darunter leidet. In der Literatur bezeichnet man dies als Sparse Approximation, und dadurch werden sowohl die Ausrechnungs- als auch die Auswertungskomplexität verringert. Zunächst werden Greedy-Verfahren in Kapitel 3 theoretisch untersucht und in Kapitel 6 werden dazu numerische Tests durchgeführt.

Genauso wie bei der Komplexität unterscheidet man auch bei der Stabilität zwischen der Ausrechnungs- und der Auswertungsstabilität. Bei steigender Punktzahl, besonders wenn die Punkte dicht zusammenliegen, steigt die Konditionszahl der Interpolationsmatrix. Deshalb sind Prädiktionierungstechniken notwendig, um das Gleichungssystem trotzdem stabil zu lösen. Kleine Fehler in den Messdaten führen sonst zu sehr großen Fehlern im Endergebnis. Auf der anderen Seite sollen kleine Fehler in den Koeffizienten der Interpolante bei der Auswertung das Ergebnis auch nur minimal beeinflussen. Bei Verwendung der Standardbasis kann es dabei zu großen Problemen kommen. Multipliziert man das Gleichungssystem zur Prädiktionierung mit einer regulären Matrix, führt man letztlich einen Basiswechsel durch, der in vielen Fällen eine deutlich stabilere Auswertung erlaubt. Dies zeigt, dass das Problem selbst nicht schlecht konditioniert ist. Stattdessen ist lediglich die Wahl der Darstellung ungeeignet. Basierend auf dieser Idee wird in Kapitel 4 eine Basis des Native Space vorgestellt, die wesentlich bessere Stabilitätseigenschaften als die Standardbasis besitzt. Außerdem hat diese Basis weitere, aus theoretischer Sicht interessante Eigenschaften, die ebenfalls vorgestellt werden. In Kapitel 7 werden dazu ebenfalls numerische Tests durchgeführt.

Kapitel 2

Interpolation mit Kernfunktionen

In diesem Abschnitt werden bereits bekannte Ergebnisse aus der Literatur zitiert, die im späteren Verlauf der Arbeit immer wieder verwendet werden. Ein Großteil der Ergebnisse ist dabei aus [22] entnommen (siehe auch [12]).

Bevor wir uns der Interpolation mit Kernfunktionen zuwenden, sollen noch ein paar Bezeichnungen eingeführt werden.

2.1 Grundlegende Begriffe

Definition 2.1.1. Wir definieren den Abstand eines Punktes $x \in \mathbb{R}^d$ zu der Menge $Y \subset \mathbb{R}^d$ durch

$$\text{dist}(x, Y) := \inf_{y \in Y} \|x - y\|_2.$$

Definition 2.1.2. Für $r \in \mathbb{R}_+$, $x \in \mathbb{R}^d$ definieren wir die Kugel um x mit Radius r durch

$$B_r(x) := \{y \in \mathbb{R}^d : \|x - y\|_2 \leq r\}.$$

Definition 2.1.3. Seien $\Omega, X \subset \mathbb{R}^d$, X endlich. Wir definieren den **Füllabstand** oder auch die **Fülldichte** $h(\Omega, X)$ von X in Ω durch

$$h(\Omega, X) := \sup_{\omega \in \Omega} \min_{x \in X} \|\omega - x\|_2$$

und den **Separationsabstand** $q(X)$ von X durch

$$\min_{x, y \in X} \frac{1}{2} \|x - y\|_2.$$

2.1: Grundlegende Begriffe

Den Füllabstand kann man sich anschaulich vorstellen als den Radius der größten Kugel mit Mittelpunkt im Abschluss von Ω , in deren Innerem kein Punkt aus X liegt. Er gibt an, wie dicht Ω durch X gefüllt ist. Der Separationsabstand bezeichnet den Radius der Kugel, die als Durchmesser den Abstand der beiden Punkte aus X hat, die am dichtesten zusammenliegen.

Definition 2.1.4. Eine Menge $\Omega \subset \mathbb{R}^d$ erfüllt eine **innere Kegelbedingung** mit Winkel $\alpha \in (0, \pi/2)$ und Radius r falls zu jedem $x \in \Omega$ ein Vektor $\xi(x)$, $\|\xi(x)\|_2 = 1$ existiert, so dass der Kegel

$$C(x, \xi(x), \alpha, r) := \left\{ x + \lambda y : y \in \mathbb{R}^d, \|y\|_2 = 1, y^T \xi(x) \geq \cos(\alpha), \lambda \in [0, r] \right\}$$

in Ω enthalten ist.

Definition 2.1.5. Eine Folge von Punktmengen X_n , $n \in \mathbb{N}$ heißt **quasi-uniform** in Ω verteilt, falls gilt

$$\exists C \in \mathbb{R} : Cq(X_n) \geq h(\Omega, X_n), \quad \forall n \in \mathbb{N}. \quad (2.1)$$

Wir sagen auch kurz, die Punkte aus X_n seien quasi-uniform verteilt, was bedeutet, dass der Separationsabstand von X ungefähr proportional zur Füllichte der Punkte aus X in Ω ist.

Satz 2.1.6. Sei $\Omega \subset \mathbb{R}^d$ beliebig, $X_n := \{x_1, \dots, x_n\} \subset \Omega$, $h_n := h(\Omega, X_n)$. Dann gibt es ein $c_\Omega \in \mathbb{R}_+$ mit

$$h_n \geq c_\Omega n^{-1/d}, \quad \forall n \in \mathbb{N}.$$

Beweis: Es gilt

$$\begin{aligned} \text{vol}(\Omega) &\leq \text{vol} \left(\bigcup_{j=1}^n B_{h_n}(x_j) \right) \\ &\leq \sum_{j=1}^n \text{vol}(B_{h_n}(0)) \\ &= n \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} h_n^d. \end{aligned}$$

Mit $c_\Omega^d := \frac{1}{\pi^{d/2}} \text{vol}(\Omega) \Gamma(d/2 + 1)$ folgt daraus

$$h_n^d \geq \frac{c_\Omega^d}{n}.$$

Durch Ziehen der d -ten Wurzel folgt die Behauptung. □

Bei quasi-uniform verteilten Punkten und beschränktem Gebiet Ω lässt sich h_n auch von unten durch n beschränken.

Satz 2.1.7. Sei $\Omega \subset \mathbb{R}^d$ beschränkt und erfülle eine innere Kegelbedingung mit Winkel α und Radius r . Seien $X_n := \{x_1, \dots, x_n\} \subset \Omega$, $h_n := h(\Omega, X_n)$, $q_n := q(X_n)$. Die Punkte aus X_n seien quasi-uniform verteilt, also $\exists M \in \mathbb{R}_+$ mit $\frac{h_n}{q_n} \leq M$, $\forall n \in \mathbb{N}$. Dann existieren $C_{\Omega, M} \in \mathbb{R}$ und $n_0 \in \mathbb{N}$ mit

$$h_n \leq C_{\Omega, M} n^{-1/d}, \quad \forall n \geq n_0.$$

Beweis:

$$\begin{aligned} \text{vol}(\Omega) &\geq \text{vol} \left(\Omega \cap \bigcup_{j=1}^n B_{q_n}(x_j) \right) \\ &= \sum_{j=1}^n \text{vol}(\Omega \cap B_{q_n}(x_j)) \\ &\geq \frac{\alpha}{2\pi} \sum_{j=1}^n \min\{\text{vol}(B_{q_n}(0)), \text{vol}(B_r(0))\} \\ &= n \frac{\alpha}{2\pi} \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} \min\{q_n^d, r^d\}. \end{aligned}$$

Da r und $\text{vol}(\Omega)$ nicht von n abhängen, folgt $q_n \rightarrow 0$ und es existiert ein $n_0 \in \mathbb{N}$ mit

$$\text{vol}(\Omega) \geq n \frac{\alpha}{2\pi} \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} q_n^d, \quad \forall n \geq n_0.$$

Wir definieren C_Ω durch

$$C_\Omega := \frac{\text{vol}(\Omega) 2\pi \Gamma(d/2 + 1)}{\alpha \pi^{d/2}}$$

und erhalten

$$C_\Omega \geq n(q_n)^d \geq n(h_n/M)^d.$$

Also gilt

$$h_n \leq M(C_\Omega/n)^{1/d} = \underbrace{C_\Omega^{1/d} M}_{=: C_{\Omega, M}} n^{-1/d}.$$

Bemerkung 2.1.8. Die Sätze 2.1.6 und 2.1.7 lassen sich in ähnlicher Form als Proposition 14.1. von [22] finden. Sie zeigen, wie bei quasi-uniform verteilten Daten die Anzahl der Datenpunkte zur Fülldichte in Relation gesetzt werden kann. Damit lassen sich h -abhängige Fehlerabschätzungen mit n -abhängigen Fehlerabschätzungen vergleichen. Für ε_X , den maximalen Fehler bei Interpolation an den Punkten aus X , erhält man nämlich nach Satz 2.1.6 die Beziehung

$$\varepsilon_X \leq c n^{-\ell/d} \Rightarrow \varepsilon_X \leq c \left(\frac{h(\Omega, X)}{c_\Omega} \right)^\ell, \quad c \in \mathbb{R}_+, \ell \in \mathbb{R}_+.$$

2.2: Resultate aus der Native Space Theorie

Auf der anderen Seite gilt nach Satz 2.1.7 aber auch

$$\varepsilon_X \leq ch(\Omega, X)^\ell \Rightarrow \varepsilon_X \leq c (C_{\Omega, M} n^{-1/d})^\ell, \quad n \geq n_0, c \in \mathbb{R}_+, \ell \in \mathbb{R}_+,$$

wenn man Quasi-Uniformität voraussetzt. Die Konstante ℓ werden wir in der folgenden Definition als Konvergenzordnung definieren.

Definition 2.1.9. *Seien $\Omega \subset \mathbb{R}^d$, $X_n := \{x_1, \dots, x_n\}$. Ein Verfahren, das jeder Funktion $f : \Omega \rightarrow \mathbb{R}$ eine Funktion $I_{X_n} f : \Omega \rightarrow \mathbb{R}$ zuordnet, liefert Konvergenzordnung der Größe ℓ auf dem Funktionenraum F , falls eine der beiden Aussagen*

$$\varepsilon_{X_n} \leq C_f h(\Omega, X_n)^\ell, \quad n \leq C_f \varepsilon_{X_n}^{-\frac{d}{\ell}},$$

für alle $f \in F$ und alle $n \in \mathbb{N}$ gilt. Dabei darf $C_f \in \mathbb{R}$ nicht von n abhängen. Für $\ell = 1$ sprechen wir von linearer, für $\ell = 2$ von quadratischer Konvergenz.

2.2 Resultate aus der Native Space Theorie

In diesem Abschnitt sollen verschiedene bekannte Resultate der Native Space Theorie aufgelistet werden, auf die später in der Arbeit Bezug genommen wird. Dabei soll zwar eine kurze Einführung basierend auf [22] gegeben werden, aber nur so viel wie benötigt wird, um die verwendeten Aussagen einordnen zu können. Für einen tieferen Einblick sei an dieser Stelle auf [12], [14], [7] und ebenfalls auf [22] verwiesen.

Definition 2.2.1. *Im Folgenden werden stetige Funktionen $K : \Omega \times \Omega \rightarrow \mathbb{R}$ eine zentrale Rolle spielen. Wir wollen diese Funktionen als **Kernfunktionen** bezeichnen, da sie als reproduzierende Kerne von Hilberträumen (siehe [5]) oder außerhalb dieser Arbeit auch als Integalkerne auftreten.*

Definition 2.2.2. *Sei $\Omega \subset \mathbb{R}^d$. Eine stetige Funktion $K : \Omega \times \Omega \rightarrow \mathbb{R}$ heißt **positiv semi-definit**, wenn für alle $n \in \mathbb{N}$, $\alpha \in \mathbb{R}^n$ und alle Mengen paarweise verschiedener Punkte $X_n := \{x_1, \dots, x_n\} \subset \Omega$ gilt*

$$\sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k K(x_j, x_k) \geq 0. \quad (2.2)$$

*K heißt **positiv definit**, falls Ungleichung (2.2) nur für $\alpha = 0$ mit Gleichheit erfüllt ist.*

Für den Aufbau einer Theorie zur Interpolation mit Kernfunktionen gibt es die Möglichkeit, ausgehend von einem Hilbertraum, über den Rieszischen Darstellungssatz zu einer positiv semi-definiten Kernfunktion zu gelangen (siehe auch Satz 10.2 und 10.4 aus [22]). In diesem Fall ist die Kernfunktion genau dann positiv definit, wenn die Punktauswertungsfunktionale linear unabhängig sind.

2.2: Resultate aus der Native Space Theorie

Da wir uns hauptsächlich für die Kernfunktionen interessieren, gehen wir den umgekehrten Weg und starten mit einer beliebigen symmetrischen, positiv definiten Funktion $K : \Omega \times \Omega \rightarrow \mathbb{R}$, $\Omega \subset \mathbb{R}^d$. Nach Satz 10.7. [22] lässt sich auf dem Raum $F_K(\Omega) := \text{span}\{K(\cdot, x) : x \in \Omega\}$ ein Inneres Produkt definieren durch

$$\left(\sum_{j=1}^N \alpha_j K(\cdot, x_j), \sum_{k=1}^M \beta_k K(\cdot, y_k) \right) := \sum_{j=1}^N \sum_{k=1}^M \alpha_j \beta_k K(x_j, y_k).$$

Mit diesem Inneren Produkt ist $F_K(\Omega)$ ein Prä-Hilbertraum mit reproduzierendem Kern K , das heißt

$$\begin{aligned} K(\cdot, x) &\in F_K(\Omega), \quad x \in \Omega, \\ f(x) &= \left(f, K(\cdot, x) \right)_K, \quad f \in F_K(\Omega), x \in \Omega. \end{aligned}$$

Also lässt sich $F_K(\Omega)$ über die Vervollständigung bezüglich $\|\cdot\|_K$ abstrakt zu einem Hilbertraum $\mathcal{F}_K(\Omega)$ ausbauen. Durch die Abbildung $R : \mathcal{F}_K(\Omega) \rightarrow C(\Omega)$, $R(f)(x) := \left(f, K(\cdot, x) \right)_K$ werden den Elementen von $\mathcal{F}_K(\Omega)$ stetige Abbildungen zugewiesen, wodurch auf sinnvolle Weise Punktauswertung in diesem Hilbertraum ermöglicht wird. Der Native Space ist dann gegeben durch $\mathcal{N}_K(\Omega) := R(\mathcal{F}_K(\Omega))$, mit Innerem Produkt

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

Für $\mathcal{N}_K(\Omega)$ gilt nun der folgende Satz.

Satz 2.2.3. *Sei $K : \Omega \times \Omega \rightarrow \mathbb{R}$ ein symmetrischer, positiv definiter Kern. Dann ist der Native Space $\mathcal{N}_K(\Omega)$ ein Hilbertraum von Funktionen und es gilt*

$$\begin{aligned} K(\cdot, x) &\in \mathcal{N}_K(\Omega), \quad x \in \Omega, \\ f(x) &= \left(f, K(\cdot, x) \right)_{\mathcal{N}_K(\Omega)}, \quad f \in \mathcal{N}_K(\Omega), x \in \Omega. \end{aligned}$$

Nun soll die hier betrachtete Interpolation mit Kernen vorgestellt werden.

Definition 2.2.4. *Es seien $\Omega \subset \mathbb{R}^d$, $f : \Omega \rightarrow \mathbb{R}$, $K : \Omega \times \Omega \rightarrow \mathbb{R}$, $X_n := \{x_1, \dots, x_n\} \subset \Omega$ gegeben. Die Interpolante zum Kern K und Datenfunktion f auf der Punktmenge X_n ist definiert durch*

$$\begin{aligned} s_{f, X_n}(x) &= \sum_{j=1}^n \alpha_j K(x, x_j), \quad \alpha \in \mathbb{R}^n, x \in \Omega \text{ und} \\ f(x_i) &= s_{f, X_n}(x_i), \quad x_i \in X_n. \end{aligned}$$

2.2: Resultate aus der Native Space Theorie

Bemerkung 2.2.5. Äquivalent zu den Gleichungen aus Definition 2.2.4 ist das Gleichungssystem

$$A\alpha = \underline{f} \quad \text{mit } A_{ij} = K(x_i, x_j), \underline{f} = (f(x_1), \dots, f(x_n))^T \text{ zu lösen.} \quad (2.3)$$

Die Matrix A wird als Interpolationsmatrix oder Kernmatrix bezeichnet.

Bemerkung 2.2.6. Ist eine Funktion $K : \Omega \times \Omega \rightarrow \mathbb{R}$ positiv definit (siehe Definition 2.2.2), so ist das Gleichungssystem aus (2.3) für beliebiges f eindeutig lösbar, weil dann die Interpolationsmatrix positiv definit ist.

Satz 2.2.7. Sei $\Omega \subseteq \mathbb{R}^d$, $X_n := \{x_1, \dots, x_n\} \subset \Omega$, $s(x) := \sum_{j=1}^n \alpha_j K(x, x_j)$, $\alpha \in \mathbb{R}^n$, $x \in \Omega$. Dann gilt

$$(f, s)_{\mathcal{N}_K(\Omega)} = 0, \quad \text{für alle } f \in \mathcal{N}_K(\Omega) \text{ mit } f|_{X_n} \equiv 0.$$

Beweis: Aus der Linearität des Skalarprodukts und der Reproduktionsgleichung aus Satz 2.2.3 erhalten wir

$$\begin{aligned} (f, s)_{\mathcal{N}_K(\Omega)} &= \sum_{j=1}^n \alpha_j (f, K(\cdot, x_j))_{\mathcal{N}_K(\Omega)} \\ &= \sum_{j=1}^n \alpha_j f(x_j) \\ &= 0. \quad \square \end{aligned}$$

Satz 2.2.8. Die Interpolante $s_{f,X}$ hat minimale Native Space Norm unter allen Funktionen des Native Space, die f auf X interpolieren,

$$\|s_{f,X}\|_{\mathcal{N}_K(\Omega)} \leq \|f\|_{\mathcal{N}_K(\Omega)}, \quad \forall f \in \mathcal{N}_K(\Omega).$$

Beweis: Nach Satz 2.2.7 gilt $(s_{f,X} - f) \perp s_{f,X}$. Daraus ergibt sich mit dem Satz des Pythagoras

$$\begin{aligned} \|s_{f,X} - f\|_{\mathcal{N}_K(\Omega)}^2 + \|s_{f,X}\|_{\mathcal{N}_K(\Omega)}^2 &= \|f\|_{\mathcal{N}_K(\Omega)}^2, \\ \Rightarrow \|s_{f,X}\|_{\mathcal{N}_K(\Omega)}^2 &\leq \|f\|_{\mathcal{N}_K(\Omega)}^2. \quad \square \end{aligned}$$

Satz 2.2.9. Die Native Space Normen der Interpolante und der Fehlerfunktion verhalten sich monoton, denn für $X \subset Y \subset \Omega$ gilt

$$\begin{aligned} \|s_{f,X}\|_{\mathcal{N}_K(\Omega)} &\leq \|s_{f,Y}\|_{\mathcal{N}_K(\Omega)}, \\ \|s_{f,X} - f\|_{\mathcal{N}_K(\Omega)} &\geq \|s_{f,Y} - f\|_{\mathcal{N}_K(\Omega)}. \end{aligned}$$

Beweis: Wegen $(s_{f,Y} - s_{f,X}) \perp s_{f,X}$ (vgl. Satz 2.2.7) gilt im Native Space

$$\|s_{f,Y} - s_{f,X}\|_{\mathcal{N}_K(\Omega)}^2 + \|s_{f,X}\|_{\mathcal{N}_K(\Omega)}^2 = \|s_{f,Y}\|_{\mathcal{N}_K(\Omega)}^2.$$

Daraus folgt, dass die Native Space Norm der Interpolante bei größer werdender Punktmenge monoton steigt. Aus dem Beweis von Satz 2.2.8 folgt damit auch, dass die Native Space Norm der Fehlerfunktion monoton fällt. \square

2.2: Resultate aus der Native Space Theorie

Definition 2.2.10. Zu gegebenen Punkten $X_n := \{x_1, \dots, x_n\} \subset \Omega \subset \mathbb{R}^d$ definieren wir die **Lagrange-Basis** $\{u_j\}_{j=1}^n$ bezüglich des Kerns K durch

$$\begin{aligned} u_j(x_i) &= \delta_{ij}, \quad 1 \leq i, j \leq n, \\ u_j &\in \text{span}\{K(\cdot, x_i) : x_i \in X_n\}, \quad 1 \leq j \leq n. \end{aligned}$$

Definition 2.2.11. Sei $\Omega \subset \mathbb{R}^d$, $K \in C(\Omega \times \Omega)$ symmetrisch, positiv definit auf Ω , $X_n := \{x_1, \dots, x_n\}$. Die **Powerfunktion** P_{K, X_n} oder auch kurz P_n ist definiert durch

$$(P_{K, X_n}(x))^2 = K(x, x) - 2 \sum_{j=1}^n u_j^{X_n}(x) K(x, x_j) + \sum_{i,j=1}^n u_i^{X_n}(x) u_j^{X_n}(x) K(x_i, x_j),$$

dabei ist $u^{X_n}(x)$ Lösung des Gleichungssystems $Au(x) = R(x)$ mit $A_{i,j} := K(x_i, x_j)$ und $R(x) := (K(x, x_1), \dots, K(x, x_n))^T$. Außerdem sei $P_0(x) := K(x, x)$.

Bemerkung 2.2.12. Die Darstellung der Powerfunktion ist etwas vereinfacht, da wir nur positiv definite Kerne betrachten. Aus $Au(x) = R(x)$ erhält man sogar

$$(P_{K, X_n}(x))^2 = K(x, x) - \sum_{j=1}^n u_j^{X_n}(x) K(x, x_j).$$

Nach Satz 11.1 aus [22] ist $u_j^{X_n}$ gerade die j -te Lagrangefunktion zu den Stützstellen X_n , denn es gilt

$$u_j^{X_n} = \sum_{i=1}^n \alpha_i^{(j)} K(\cdot, x_i), \quad \text{mit } u_j^{X_n}(x_k) = \delta_{jk}, \quad k \in \{1, \dots, n\}.$$

Als Spezialfall von Satz 11.4 aus [22] gilt

Satz 2.2.13. Sei $\Omega \subset \mathbb{R}^d$, $X \subset \Omega$, $K \in C(\Omega \times \Omega)$ ein positiv definit, symmetrischer Kern auf Ω . Für $x \in \Omega$ kann der Interpolationsfehler zwischen f und $s_{f, X}$ beschränkt werden durch

$$|f(x) - s_{f, X}(x)| \leq P_{K, X}(x) \|f\|_{\mathcal{N}_{K, \Omega}}. \quad (2.4)$$

Satz 2.2.14. Die Powerfunktion lässt sich darstellen durch (siehe auch [16])

$$P_{K, X_n}(x) = \sup_{f \in \mathcal{N}_{K}(\Omega)} \{f(x) : \|f\|_{\mathcal{N}_{K}(\Omega)} \leq 1, f|_{X_n} \equiv 0\}, \quad (2.5)$$

beziehungsweise durch die Native Space Norm des Fehlerfunctionals

$$P_{K, X_n}(x) = \min_{u \in \mathbb{R}^n} \|\delta_x - \sum_{j=1}^n u_j \delta_{x_j}\|_{\mathcal{N}_K^*} = \min_{u \in \mathbb{R}^n} \|K(\cdot, x) - \sum_{j=1}^n u_j K(\cdot, x_j)\|_{\mathcal{N}_K}. \quad (2.6)$$

2.2: Resultate aus der Native Space Theorie

Dabei bezeichnet δ_x das Punktauswertungsfunktional an der Stelle x .

Beweis: Zunächst soll die erste Gleichung bewiesen werden.

Sei $M_{X_n} := \{f \in \mathcal{N}_K(\Omega) : \|f\|_{\mathcal{N}_K(\Omega)} \leq 1, f|_{X_n} \equiv 0\}$, $f_{X_n, x} := \sup_{f \in M_{X_n}} \{f(x)\}$.

Behauptung: $P_{K, X_n}(x) \geq f_{X_n, x}$.

Wegen Satz 2.2.13 gilt für $f \in \mathcal{N}_K(\Omega)$, $\|f\|_{\mathcal{N}_K(\Omega)} \leq 1$, $f|_{X_n} \equiv 0$ die Abschätzung

$$|f(x)| \leq P_{K, X_n}(x) \|f\|_{\mathcal{N}_K(\Omega)} \leq P_{K, X_n}(x).$$

Behauptung: $P_{K, X_n}(x) \leq f_{X_n, x}$

1. Fall: $P_{K, X_n}(x) = 0$

$$\Rightarrow P_{K, X_n}(x) = 0 \leq f_{X_n, x},$$

da, die Nullfunktion immer in M_{X_n} enthalten ist.

2. Fall: $P_{K, X_n}(x) \neq 0$

Es sei $f := K(\cdot, x) - \sum_{j=1}^n u_j^{X_n}(x) K(\cdot, x_j)$. Das Quadrat der Native Space Norm von f lässt sich umformen zu

$$\begin{aligned} \|f\|_{\mathcal{N}_K(\Omega)}^2 &= \left(K(\cdot, x) - \sum_{j=1}^N u_j^{X_n}(x) K(\cdot, x_j), K(\cdot, x) - \sum_{j=1}^N u_j^{X_n}(x) K(\cdot, x_j) \right)_{\mathcal{N}_K(\Omega)} \\ &= K(x, x) - 2 \sum_{j=1}^N u_j^{X_n}(x) K(x, x_j) + \sum_{i, j=1}^N u_i^{X_n}(x) u_j^{X_n}(x) K(x_i, x_j) \\ &= (P_{K, X_n}(x))^2. \end{aligned}$$

Mit $g := \frac{f}{\|f\|_{\mathcal{N}_K(\Omega)}}$ und wegen $f|_{X_n} \equiv 0$, $\|g\|_{\mathcal{N}_K(\Omega)} = 1$ und $g \in \mathcal{N}_K(\Omega)$ gilt $g \in M_{X_n}$. Außerdem beträgt der Wert von g an der Stelle x

$$\Rightarrow g(x) = \frac{(P_{K, X_n}(x))^2}{P_{K, X_n}(x)} = P_{K, X_n}(x).$$

Die Darstellung der Powerfunktion über die Native Space Norm des Fehlerfunktionals gilt nach der Bemerkung zu Satz 16.3. aus [22]. \square

Aus Satz 2.2.14 folgt sofort, dass die Folge der Powerfunktionen mit wachsender Punktmenge punktweise monoton fallend ist.

Korollar 2.2.15. *Sei $X \subset Y$. Dann gilt*

$$P_{K, X}(x) \geq P_{K, Y}(x), \forall x \in \Omega.$$

Für $x_i \in X$ erhält man direkt durch Einsetzen in Satz 2.2.14

Korollar 2.2.16.

$$P_{K, X}(x_i) = 0.$$

Da die Kernfunktion K nach Voraussetzung positiv definit ist, besitzt die Powerfunktion auch keine weiteren Nullstellen.

Korollar 2.2.17.

$$P_{K,X}(x) > 0, \forall x \in \Omega \setminus X.$$

Beweis: Aus 2.2.14 folgt sofort $P_{K,X}(x) \geq 0$. Angenommen $x \in \Omega \setminus X$ ist eine weitere Nullstelle von $P_{K,X}$. Dann sind die Punktauswertungsfunctionale nach Gleichung (2.6) linear abhängig, was ein Widerspruch dazu ist, dass K positiv definit ist. \square

Wir verwenden im Folgenden den „Powerkern“ K_n zusammen mit einigen Eigenschaften aus [6]. Speziell benutzen wir von dort die Darstellung aus Proposition 4.8., um den Powerkern zu definieren.

Definition 2.2.18. Der **Powerkern** K_n zu den Stützpunkten $X_n := \{x_1, \dots, x_n\}$ und dem positiv definiten Kern K ist definiert durch

$$K_n(x, y) = K(x, y) - \sum_{j=1}^n u_j^{X_n}(x)K(x_j, y) - \sum_{k=1}^n u_k^{X_n}(y)K(x, x_k) + \sum_{j,k=1}^n u_j^{X_n}(x)u_k^{X_n}(y)K(x_j, x_k).$$

Dabei sind die Funktionen $u_j^{X_n}$, $1 \leq j \leq n$, die Lagrange-Funktionen aus Definition 2.2.11.

Satz 2.2.19. Mit den Definitionen aus 2.2.18 gelten für den Powerkern die folgenden Eigenschaften ($x, y \in \Omega$, $n \in \mathbb{N}$):

- a) $P_n^2(x) = K_n(x, x)$,
- b) $K_{n+1}(x, y) = K_n(x, y) - \frac{K_n(x, x_{n+1})K_n(x_{n+1}, y)}{K_n(x_{n+1}, x_{n+1})}$,
- c) $K_n(x_j, x) = K_n(x, x_j) = 0$, $1 \leq j \leq n$,
- d) $K_n(\cdot, x_{n+1}) \in \text{span}\{K(\cdot, x_i) : 1 \leq i \leq n+1\}$,
- e) $K_n(x, x_{n+1}) = u_{n+1}^{X_{n+1}}(x)K_n(x_{n+1}, x_{n+1})$.

Beweis:

- a) wird in Satz 4.10 aus [6] gezeigt, folgt aber auch direkt durch Einsetzen in die hier benutzten Definitionen.
- b) ist Satz 5.19 aus [6].
- c) bekommt man durch Einsetzen in die Definition beziehungsweise nach Korollar 4.7 aus [6].

2.2: Resultate aus der Native Space Theorie

d) gilt, weil nach Definition 2.2.18 $K_n(\cdot, x_{n+1})$ eine Linearkombination der Lagrange-Funktionen $u_j^{X_n}(\cdot)$, $j \in \{1, \dots, n\}$ und der Kernfunktionen $K(\cdot, x_k)$, $k \in \{1, \dots, n+1\}$ ist und die Lagrange-Funktionen $u_j^{X_n}(\cdot)$ nach der Bemerkung zu Definition 2.2.11 in $\text{span}\{K(\cdot, x_i) : 1 \leq i \leq n\}$ liegen.

e) folgt schließlich aus d) zusammen mit $u_{n+1}^{X_{n+1}} = 1$ und

$$K_n(x_i, x_{n+1}) = u_{n+1}^{X_{n+1}}(x_i) = 0, \forall x_i \in X_n. \quad \square$$

Satz 2.2.20. *Es seien $X_n \subset \Omega \subset \mathbb{R}^d$, $n \in \mathbb{N}$, mit $h(\Omega, X_n) \xrightarrow{n \rightarrow \infty} 0$ und es gelte $\mathcal{N}_K(\Omega) \subseteq C(\Omega)$. Dann gilt für die Powerfunktion*

$$P_n(x) \xrightarrow{n \rightarrow \infty} 0, \quad x \in \Omega.$$

Beweis:

Nach Proposition 4.15 aus [6] gilt

$$\|K_n(x, \cdot)\|_{\mathcal{N}_K(\Omega)} = K_n(x, x), \quad x \in \Omega.$$

Wegen $K_n(x, x) = P_{X_n}^2(x)$ (Satz 2.2.19 a)) und der Monotonie der Powerfunktion (Korollar 2.2.15) ist daher $(K_n(x, \cdot))_{n \in \mathbb{N}}$ eine in $\mathcal{N}_K(\Omega)$ beschränkte Folge. Da $\mathcal{N}_K(\Omega)$ ein Hilbertraum und somit reflexiv ist, besitzt diese Folge eine schwach konvergente Teilfolge $(K_{\varphi_n}(x, \cdot))_{n \in \mathbb{N}}$ mit Grenzelement $K_{x, \infty} \in \mathcal{N}_K(\Omega)$. Für $y \in \Omega$ folgt damit punktweise Konvergenz

$$(K_{\varphi_n}(x, y))_{n \in \mathbb{N}} \xrightarrow{n \rightarrow \infty} K_{x, \infty}(y).$$

Nach Satz 2.2.19 c) ist $K_n(x, x_j) = 0$, $x_j \in X_n$ und damit $K_{x, \infty}(y) = 0$ für $y \in \bigcup_{n \in \mathbb{N}} X_n$. Damit verschwindet $K_{x, \infty}$ auf einer dichten Teilmenge von Ω und aus der Stetigkeit von $K_{x, \infty}$ (es gilt $\mathcal{N}_K(\Omega) \subseteq C(\Omega)$ nach Voraussetzung) folgt $K_{x, \infty} \equiv 0$. Für $y = x$ folgt wegen der Monotonie der Powerfunktion sogar Konvergenz für die gesamte Folge, also

$$(K_n(x, x))_{n \in \mathbb{N}} \xrightarrow{n \rightarrow \infty} K_{x, \infty}(x).$$

Damit erhält man die Behauptung aus

$$P_{X_n}^2(x) = K_n(x, x) \xrightarrow{n \rightarrow \infty} K_{x, \infty}(x) = 0, \quad x \in \Omega.$$

Korollar 2.2.21. *Es seien $X_n \subset \Omega \subset \mathbb{R}^d$, $n \in \mathbb{N}$, mit $h(\Omega, X_n) \xrightarrow{n \rightarrow \infty} 0$ und es gelte $\mathcal{N}_K(\Omega) \subseteq C(\Omega)$. Dann gilt für den Interpolationsfehler an beliebiger Stelle $x \in \Omega$*

$$(f - s_{f, X_n})(x) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis:

Nach der Fehlerabschätzung aus Satz 2.2.13 und nach Satz 2.2.20 gilt

$$|(f - s_{f, X_n})(x)| \leq P_{K, X_n}(x) \|f\|_{\mathcal{N}_K(\Omega)} \xrightarrow{n \rightarrow \infty} 0.$$

Kapitel 3

Konvergenzeigenschaften des f -Greedy-Verfahrens

Wir wollen eine Funktion f approximieren, mit $f : \Omega \rightarrow \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$. Dazu verwenden wir einen Approximationsoperator $I_{X_n} : C(\Omega) \rightarrow C(\Omega)$ basierend auf den Funktionswerten von f , ausgewertet an den Punkten aus $X_n := \{x_1, \dots, x_n\} \subset \Omega$.

Die Punkte werden dabei adaptiv über eines der zu Beginn des nächsten Abschnitts definierten Greedy-Verfahren (vgl. [17]) gewählt. Anwendungen zu Greedy-Verfahren finden sich zum Beispiel in [4, 15].

Für die Definition der Greedy-Verfahren spielt es keine Rolle, wie der Approximationsoperator I_{X_n} genau definiert ist. Wir interessieren uns aber speziell für die Interpolation mit Kernfunktionen $K : \Omega \times \Omega \rightarrow \mathbb{R}$, wie sie in Definition 2.2.4 beschrieben ist, also $I_{X_n} f = s_{f, X_n}$. Dabei setzen wir voraus, dass der Kern K positiv definit und symmetrisch ist.¹

3.1 Greedy-Verfahren

Definition 3.1.1. Seien $f : \Omega \rightarrow \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$ gegeben, wie zu Beginn des Kapitels beschrieben. Als **Greedy-Verfahren** bezeichnen wir Methoden zur Auswahl der Datenpunkte $X_n := \{x_1, \dots, x_n\} \subset \Omega$ unter den drei alternativen Bedingungen

$$\begin{aligned} |f(x_1)| &= \|f\|_{L_\infty(\Omega)} \\ |(f - s_{f, X_k})(x_{k+1})| &= \|f - s_{f, X_k}\|_{L_\infty(\Omega)}, \quad 1 \leq k < n, \quad (f\text{-Greedy}) \end{aligned}$$

$$|P_{K, X_k}(x_{k+1})| = \|P_{K, X_k}\|_{L_\infty(\Omega)}, \quad 1 \leq k < n, \quad (P\text{-Greedy})$$

$$\frac{|(f - s_{f, X_k})(x_{k+1})|}{P_{K, X_k}(x_{k+1})} = \left\| \frac{f - s_{f, X_k}}{P_{K, X_k}} \right\|_{L_\infty(\Omega \setminus \{X_k\})}, \quad 1 \leq k < n. \quad (f/P\text{-Greedy})$$

Dabei darf x_1 bei P -Greedy und bei f/P -Greedy beliebig aus Ω gewählt werden.

¹Die Interpolation mit Kernfunktionen wird in Kapitel 2 ausführlich beschrieben. Dort werden auch die im Folgenden verwendeten Begriffe aus der Native Space Theorie erläutert.

3.1: Greedy-Verfahren

Bemerkung 3.1.2. Nach Definition der Greedy-Verfahren wird in jedem Schritt der Datenpunkt an dem jeweiligen Maximum hinzugenommen. Wir setzen hier voraus, dass dieses Maximum in Ω angenommen wird. Dies ist für stetige Funktionen f , K und kompaktes Ω immer der Fall.

Wird das Maximum in verschiedenen Punkten angenommen, könnte man zufällig einen dieser Punkte auswählen. Darauf soll hier aber nicht näher eingegangen werden.

Bemerkungen zu P -Greedy

Bemerkung 3.1.3. Im Gegensatz zu den anderen beiden hier vorgestellten Greedy-Verfahren, werden die Punkte bei P -Greedy f -unabhängig verteilt (siehe auch [3]). Aus der Definition von P -Greedy und in Hinblick auf Gleichung (2.5) ist zu erwarten, dass bei dieser Punktwahl keine großen „Löcher“ in Ω entstehen, in denen kein Punkt ausgewählt wird, da die Punkte gleichmäßig gut für alle Funktionen des Native Space gewählt werden.

Tatsächlich wird in Kapitel 6 deutlich, dass P -Greedy die Punkte sehr gleichmäßig verteilt. Das führt auch zu relativ gut konditionierten Kernmatrizen bei der Rekonstruktion.

Bemerkungen zu f/P -Greedy

Aus Korollar 2.2.17 folgt, dass die Powerfunktion P_{K, X_k} in $\Omega \setminus X_k$ keine weiteren Nullstellen hat und damit f/P -Greedy wohldefiniert ist.

Bemerkung 3.1.4. Für $f \in \mathcal{N}_K(\Omega)$ gilt (siehe Satz 6 in [18] oder in derselben Notation wie hier in [6] Satz 5.24)

$$\|s_{f, X_{k+1}}\|_{\mathcal{N}_K}^2 = \|s_{f, X_k}\|_{\mathcal{N}_K}^2 + \frac{\left((f - s_{f, X_k})(x_{k+1})\right)^2}{P_{K, X_k}^2(x_{k+1})}.$$

Deshalb wird bei f/P -Greedy jeweils der nächste Punkt so ausgewählt, dass die **Native Space Norm der Interpolante maximiert** wird. Wegen $\|f - s_{f, X_k}\|_{\mathcal{N}_K}^2 = \|f\|_{\mathcal{N}_K}^2 - \|s_{f, X_k}\|_{\mathcal{N}_K}^2$ wird damit die Native Space Norm der Fehlerfunktion minimiert und damit auch der Fehler klein gehalten. Es gilt nämlich nach der Reproduktionseigenschaft (Satz 2.2.3)

$$|(f - s_{f, X_k})(x)| = (f - s_{f, X_k}, K(\cdot, x))_{\mathcal{N}_K} \leq \|f - s_{f, X_k}\|_{\mathcal{N}_K} \|K(\cdot, x)\|_{\mathcal{N}_K}.$$

Alternativ kann man auch in Satz 2.2.13 für f die Fehlerfunktion einsetzen und bekommt

$$|(f - s_{f, X_k})(x)| \leq P_{K, X_n}(x) \|f - s_{f, X_k}\|_{\mathcal{N}_K(\Omega)}, \quad x \in \Omega.$$

In Kapitel 6 werden wir diese Greedy-Verfahren miteinander vergleichen. Für den Rest dieses Kapitels betrachten wir aber ausschließlich f -Greedy.

3.2 f -Greedy bei Kernfunktionen

In diesem Abschnitt wollen wir Konvergenzeigenschaften der f -Greedy Methode untersuchen. Im Vergleich zu Verfahren, die ein regelmäßiges Gitter benutzen, beschäftigen wir uns mit der Frage, mit wie vielen Punkten man ein vorgegebenes Fehlerniveau erreichen kann.

Definition 3.2.1. Die Fehlerfunktion nach n Iterationsschritten bezeichnen wir im Folgenden mit $g_n := f - s_{f, X_n}$ und den maximalen Fehler mit $\varepsilon_n := \|g_n\|_{L_\infty(\Omega)}$.

Bemerkungen zu f -Greedy

f -Greedy arbeitet datenabhängig. Charakteristisch ist dabei die Voraussetzung, dass die Fehlerfunktion $f - I_{X_n}f$ auf Ω oder in Anwendungen zumindest auf einer feinen Punktwolke in Ω bekannt ist. Diese Bedingung ist dann erfüllt, wenn man eine sehr große Menge dichtverteilter Daten hat, die man ausdünnen möchte, um eine einfachere Darstellung der Funktion zu haben.

Bei Punktwahl durch f -Greedy werden keine Punkte in Bereichen ausgesucht, wo der Fehler bereits klein ist. Ist global der Fehler noch groß, müssen die ausgewählten Punkte weit auseinander liegen, weil der Fehler an den Interpolationspunkten verschwindet. Dadurch könnte man vermuten, dass es in diesem Fall zu einer Raumfüllung kommt, aber das ist noch zu untersuchen. Unter „Raumfüllung“ wird im Folgenden verstanden, dass eine Punktfolge x_1, x_2, \dots erzeugt wird, so dass die Füllichte $h_n := (\Omega, \{x_1, \dots, x_n\})$ für $n \rightarrow \infty$ gegen Null strebt.

Numerische Experimente (siehe Kapitel 6) zeigen, dass viele Punkte in den Bereichen ausgewählt werden, in denen die Funktion f ihr Verhalten stark ändert, wie zum Beispiel bei hohen Ableitungen oder Kanten. Aber auch am Rand ist der Fehler bei Interpolation mit Kernen oft verhältnismäßig groß, wodurch dort bei Verwendung von f -Greedy ebenfalls viele Punkte ausgesucht werden. Deshalb wird bei f -Greedy nicht unbedingt eine Raumfüllung erreicht, was aber durchaus positiv sein kann, wenn dadurch insgesamt sehr wenige Punkte benötigt werden. Außerdem werden so „kritische“ Regionen von f gefunden.

Das folgende Resultat findet man zum Beispiel als Proposition 5.26 in [6]. Trotzdem wird der Satz hier nochmal mit ähnlichem Beweis aufgeschrieben, da in den einzelnen Beweisschritten viele grundlegende Eigenschaften des Native Space verwendet werden, die dabei vorgestellt werden sollen.

Satz 3.2.2. Sei $f \in \mathcal{N}_K(\Omega)$. Dann konvergiert bei Interpolation mit Kernfunktionen und der Punktauswahl über das Greedy-Verfahren nach Definition 3.1.1 (f -Greedy) der maximale Fehler ε_n für wachsendes n gegen Null, genauer gilt

$$\sum_{i=1}^{\infty} \varepsilon_i^2 < \infty. \quad (3.1)$$

3.2: f -Greedy bei Kernfunktionen

Beweis: Für Funktionen f aus dem Native Space $\mathcal{N}_K(\Omega)$ gelangt man mit Hilfe der Powerfunktion P_{K,X_n} zu der Fehlerabschätzung (vgl. Satz 2.2.13)

$$|f(x) - s_{f,X_n}(x)| \leq P_{K,X_n}(x) \|f\|_{\mathcal{N}_K(\Omega)}, \quad x \in \Omega.$$

Setzt man in dieser Gleichung als Datenfunktion $s_{f,X_{i+1}} - s_{f,X_i}$, $i \in \mathbb{N}$ mit der Nullfunktion als Interpolante auf X_i ein, erhält man die Abschätzung

$$\varepsilon_i = |s_{f,X_{i+1}}(x_{i+1}) - s_{f,X_i}(x_{i+1})| \leq P_{K,X_i}(x_{i+1}) \|s_{f,X_{i+1}} - s_{f,X_i}\|_{\mathcal{N}_K(\Omega)}.$$

Nun betrachten wir die Summe der Fehlerquadrate:

$$\sum_{i=1}^n \varepsilon_i^2 \leq \sum_{i=1}^n \|P_{K,X_i}\|_{L_\infty(\Omega)}^2 \|s_{f,X_{i+1}} - s_{f,X_i}\|_{\mathcal{N}_K(\Omega)}^2.$$

Wegen $(s_{f,X_{i+1}} - s_{f,X_i}) \perp s_{f,X_i}$ (vgl. Satz 2.2.7) gilt im Native Space

$$\|s_{f,X_{i+1}} - s_{f,X_i}\|_{\mathcal{N}_K(\Omega)}^2 + \|s_{f,X_i}\|_{\mathcal{N}_K(\Omega)}^2 = \|s_{f,X_{i+1}}\|_{\mathcal{N}_K(\Omega)}^2.$$

Außerdem ist die Folge der Powerfunktionen punktweise monoton fallend:

$$P_{K,X_i}(x) \geq P_{K,X_{i+1}}(x), \quad i \in \mathbb{N}, \quad x \in \Omega \quad (\text{vgl. Korollar 2.2.15}).$$

Daraus ergibt sich für die Folge der Fehler

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &\leq \|P_{K,X_1}\|_{L_\infty(\Omega)}^2 \sum_{i=1}^n \|s_{f,X_{i+1}} - s_{f,X_i}\|_{\mathcal{N}_K(\Omega)}^2 \\ &\leq \|P_{K,X_1}\|_{L_\infty(\Omega)}^2 \sum_{i=1}^n (\|s_{f,X_{i+1}}\|_{\mathcal{N}_K(\Omega)}^2 - \|s_{f,X_i}\|_{\mathcal{N}_K(\Omega)}^2) \\ &\leq \|P_{K,X_1}\|_{L_\infty(\Omega)}^2 \|s_{f,X_{n+1}}\|_{\mathcal{N}_K(\Omega)}^2 \\ &\leq \|P_{K,X_1}\|_{L_\infty(\Omega)}^2 \|f\|_{\mathcal{N}_K(\Omega)}^2, \quad \forall n \in \mathbb{N}. \quad \square \end{aligned}$$

Wünschenswert wären für f -Greedy Aussagen darüber, wie viele Punkte benötigt werden, um ein vorgegebenes Fehlerlevel

$$\|f - s_{f,X_n}\|_{L_\infty(\Omega)} \leq \varepsilon$$

zu erreichen. Die bekannten Konvergenzsätze, die hohe Konvergenzordnungen für die Kerninterpolation zeigen, lassen sich aber leider nicht anwenden, weil diese gerade von der Raumfüllung abhängen.

Zur Raumfüllung bei f -Greedy wird in dem folgenden Satz gezeigt, dass auf einem festen Gebiet nur endlich oft die Fehlerabschätzung verletzt sein kann, die durch die h -abhängige Asymptotik gegeben ist. Um hieraus allerdings eine Konvergenzaussage der Form $n \leq C\varepsilon_n^{-d/\ell}$ zu bekommen, müsste man wissen, wie schnell $h(\Omega, X_n)$ mit wachsendem n gegen Null konvergiert.

Satz 3.2.3. *Seien f, I_{X_n} wie in Definition 3.1.1 gegeben mit Fehlern $\varepsilon_n := \|f - I_{X_n}f\|_{L^\infty(\Omega)}$, für $n \in \mathbb{N}$. Außerdem gelte für die von n unabhängigen Konstanten $C_{\Omega,f}, h_0 \in \mathbb{R}$*

$$\varepsilon_n \leq C_{\Omega,f} h(\Omega, X_n)^\ell, \quad \text{für } h(\Omega, X_n) \leq h_0.$$

Dann gibt es nur endlich viele $n \in \mathbb{N}$ mit $\varepsilon_n > C_{\Omega,f} h(\Omega, X_n)^\ell$.

Beweis: Aus $\varepsilon_n > C_{\Omega,f} h(\Omega, X_n)^\ell$ folgt nach Voraussetzung $h(\Omega, X_n) > h_0$ und damit $\varepsilon_n > C_{\Omega,f} h_0^\ell$. Wegen Satz 3.2.2 kann dies aber nur in endlich vielen Fällen erfüllt sein, weil sonst die Reihe $\sum_{i=1}^{\infty} \varepsilon_i^2$ divergieren würde. \square

Trotz der genannten Schwierigkeiten soll hier nach Konvergenzaussagen gesucht werden, die den h -abhängigen Abschätzungen möglichst nahe kommen. Dabei ist es ein Fernziel, zu zeigen, dass f -Greedy zur Erreichung eines gegebenen Fehlerlevels weniger Punkte verwendet, als bei gleichmäßiger Raumfüllung nötig sind.

Zumindest in \mathbb{R}^1 kann gezeigt werden, dass f -Greedy mindestens dieselbe Ordnung erreicht, die auch bei äquidistanten Punkten erzielt wird. Bei den numerischen Tests in Kapitel 6 sieht man sogar, dass f -Greedy in der Regel mit deutlich weniger Punkten als bei äquidistanter Verteilung auskommt.

Satz 3.2.2 garantiert nur, dass die Folge der Fehler in ℓ_2 liegt und somit gegen Null konvergiert, es wird aber nichts über die Konvergenzgeschwindigkeit ausgesagt, da vereinzelt große Werte auftreten können.

Immerhin folgt aus der Konvergenz der Reihe, dass es eine Teilfolge $(\varepsilon_{\varphi_n})_{n \in \mathbb{N}}$ mit

$$\varepsilon_{\varphi_n} \leq (\varphi_n)^{-\frac{1}{2}} \quad \text{gibt.} \tag{3.2}$$

Also gibt es unendlich viele $j \in \mathbb{N}$, für die der Gesamtfehler im j -ten Schritt durch $\varepsilon_j \leq j^{-\frac{1}{2}}$ beschränkt werden kann. Ähnliche Konvergenzraten werden auch bei adaptiven Verfahren in der Lerntheorie bewiesen (siehe [21]). Im Vergleich zu den weiteren Abschätzungen (vgl. Gleichung (3.8)) fällt auf, dass die Anzahl der benötigten Punkte nicht von der Raumdimension d abhängt. Nach Definition 2.1.9 beträgt die Konvergenzordnung für diese Teilfolge daher mindestens $\ell = d/2$. Die folgenden Abschnitte verbessern diese Konvergenzaussage.

3.3 Konvergenzaussagen für f -Greedy

In Korollar 3.3.8 wird für f -Greedy quadratische Konvergenz im Inneren von Ω beziehungsweise lineare Konvergenz auf Ω bewiesen. Dazu wird zunächst gezeigt, dass die Punkte aus X_n einen durch den Fehler ε_n determinierten Mindestabstand voneinander haben.

Satz 3.3.1. *Falls Ω konvex und die Fehlerfunktion g_n einmal stetig differenzierbar mit beschränkter erster Ableitung ist mit $M_n := \sup_{\omega \in \Omega} \|\nabla g_n(\omega)\|_2 < \infty$, gilt für die mit*

3.3: Konvergenzaussagen für f -Greedy

Hilfe von f -Greedy (siehe 3.1.1 (**f -Greedy**)) ausgewählten Punkte

$$\text{dist}(x_{n+1}, X_n) \geq \frac{1}{M_n} \varepsilon_n, \quad \forall n \in \mathbb{N}. \quad (3.3)$$

Beweis: Sei $x \in X_n$, $h := x_{n+1} - x$. Nach dem Mittelwertsatz für reellwertige Funktionen gibt es ein $t \in (0, 1)$ mit

$$\underbrace{g_n(x_{n+1})}_{=\varepsilon_n} - \underbrace{g_n(x)}_{=0} = \nabla g_n(x + th)h.$$

Daraus erhält man durch Anwendung der Cauchy-Schwarzschen Ungleichung

$$\begin{aligned} \varepsilon_n &\leq \|\nabla g_n(x + th)\|_2 \|h\|_2 \\ &\leq M_n \cdot \text{dist}(x_{n+1}, x). \end{aligned}$$

Da x beliebig aus X_n gewählt war, folgt daraus die Behauptung. \square

Im Inneren von Ω können wir zusätzlich ausnutzen, dass die Fehlerfunktion g_n an der Stelle x_{n+1} ein lokales Extremum hat und deshalb ihre erste Ableitung dort verschwindet. Damit erhalten wir

Satz 3.3.2. *Es sei $\Omega \subset \mathbb{R}^d$ konvex und die Fehlerfunktion g_n zweimal stetig differenzierbar mit beschränkten zweiten partiellen Ableitungen mit*

$$M_n := \sum_{j,k=1}^d \sup_{\omega \in \Omega} \left| \frac{\partial^2 g_n(\omega)}{\partial x_j \partial x_k} \right|, \quad M_n < \infty.$$

*Bei der Punktauswahl gemäß 3.1.1 (**f -Greedy**) gilt dann für die Punkte x_{n+1} , $n \in \mathbb{N}$, die im Inneren von Ω liegen, die Abschätzung*

$$\text{dist}(x_{n+1}, X_n) \geq \left(\frac{2}{M_n} \varepsilon_n \right)^{1/2}. \quad (3.4)$$

Beweis: Sei $x \in X_n$, $h := x_{n+1} - x$, $h = (h_1, \dots, h_d)^T$. Nach dem Satz von Taylor gibt es ein $t \in (0, 1)$ mit

$$\begin{aligned} \underbrace{g_n(x)}_{=0} &= \underbrace{g_n(x_{n+1})}_{=\varepsilon_n} + \underbrace{\nabla g_n(x_{n+1})h}_{=0} + \frac{1}{2} \sum_{j,k=1}^d \frac{\partial^2 g_n(x_{n+1} + th)}{\partial x_j \partial x_k} h_j h_k, \\ \Rightarrow \varepsilon_n &= \frac{1}{2} \left| \sum_{j,k=1}^d \frac{\partial^2 g_n(x_{n+1} + th)}{\partial x_j \partial x_k} h_j h_k \right|. \end{aligned}$$

Daraus erhalten wir für beliebiges $x \in X_n$ und $h := x_{n+1} - x$

$$\begin{aligned} \varepsilon_n &\leq \frac{\|h\|_\infty^2}{2} \sum_{j,k=1}^d \sup_{\omega \in \Omega} \left| \frac{\partial^2 g_n(\omega)}{\partial x_j \partial x_k} \right| \\ &\leq \frac{\text{dist}(x_{n+1}, x)^2}{2} M_n. \quad \square \end{aligned}$$

Bemerkung 3.3.3. Die Sätze 3.3.1 und 3.3.2 zeigen, dass die Punkte, die durch f -Greedy ausgewählt werden, nicht beliebig dicht beieinander liegen können. Solange der Fehler groß ist, liegen deshalb die ausgewählten Punkte weit auseinander und füllen so den Raum. Nur wenn der Fehler sowieso schon klein ist, können durch f -Greedy Punkte dicht beieinander ausgesucht werden. Bei Interpolation mit Kernfunktionen wirkt sich das auch positiv auf die Kondition der Kernmatrix aus.

Der folgende Satz liefert zwar keine Konvergenzaussage für eine wachsende Anzahl an Interpolationspunkten. Stattdessen wird aber gezeigt, dass im Gegensatz zur Polynominterpolation für hinreichend glattes K und $f \in \mathcal{N}_K(\Omega)$ die Ableitungen der Fehlerfunktion bei Kerninterpolation beschränkt bleiben, so dass die Sätze 3.3.1 und 3.3.2 anwendbar sind. Diese Schranke ist sogar unabhängig von n , was später für die Sätze 3.3.6 und 3.3.13 vorausgesetzt wird. Diese Aussage erhält man auch über die Standardfehlerabschätzungen, allerdings wird dann in der Regel eine bestimmte Füllichte vorausgesetzt.

Satz 3.3.4. Sei $K \in C^{2k}(\Omega \times \Omega)$, $k \in \mathbb{N}$, K positiv definit, symmetrisch, $f \in \mathcal{N}_K(\Omega)$. Dann lässt sich der Betrag der α -ten Ableitung, $\alpha \in \mathbb{N}_0^d$, $|\alpha| \leq k$, der Fehlerfunktion $f - s_{f, X_n}$ beschränken durch

$$|D^\alpha(f - s_{f, X_n})(x)| \leq 2\|f\|_{\mathcal{N}_K(\Omega)}\|D_2^\alpha K(\cdot, x)\|_{\mathcal{N}_K(\Omega)}, \quad \forall x \in \Omega, \forall n \in \mathbb{N}.$$

Dabei wird mit $D_2^\alpha K$ die α -te Ableitung bezüglich des zweiten Arguments von K bezeichnet.

Beweis: Zunächst nutzen wir die Normminimalität der Interpolante s_{f, X_n} aus (vgl. Satz 2.2.8) und erhalten

$$\begin{aligned} \|f - s_{f, X_n}\|_{\mathcal{N}_K(\Omega)} &\leq \|f\|_{\mathcal{N}_K(\Omega)} + \|s_{f, X_n}\|_{\mathcal{N}_K(\Omega)} \\ &\leq 2\|f\|_{\mathcal{N}_K(\Omega)}. \end{aligned} \tag{3.5}$$

Mit Satz 10.45 aus [22] und der Cauchy-Schwarzschen Ungleichung erhalten wir

$$\begin{aligned} |D^\alpha(f - s_{f, X_n})(x)| &\leq |(f - s_{f, X_n}, D_2^\alpha K(\cdot, x))_{\mathcal{N}_K(\Omega)}| \\ &\leq \|f - s_{f, X_n}\|_{\mathcal{N}_K(\Omega)}\|D_2^\alpha K(\cdot, x)\|_{\mathcal{N}_K(\Omega)} \\ &\stackrel{(3.5)}{\leq} 2\|f\|_{\mathcal{N}_K(\Omega)}\|D_2^\alpha K(\cdot, x)\|_{\mathcal{N}_K(\Omega)}. \quad \square \end{aligned}$$

Bisher wurde für das Greedy-Verfahren gezeigt, dass der Abstand der ausgewählten Punkte durch das Minimum der maximalen Fehler auf Ω nach unten beschränkt werden kann, falls die Kernfunktion K hinreichend glatt ist und die Datenfunktion f in $\mathcal{N}_K(\Omega)$ liegt. Ist Ω beschränkt, folgt daraus, dass dieses Minimum für wachsendes n klein werden muss. Dies wollen wir in dem folgenden Satz genauer formulieren.

3.3: Konvergenzaussagen für f -Greedy

Voraussetzungen 3.3.5. Wir setzen voraus, dass $\Omega \subset \mathbb{R}^d$ beschränkt ist und die innere Kugelbedingung mit Winkel $\alpha \in (0, \pi/2)$ und Radius $r > 0$ erfüllt.

Zusätzlich seien die Punkte aus X_n unter Verwendung des Greedy-Verfahrens aus Ω ausgewählt und mit den Bezeichnungen von oben gelte für den Abstand der Datenpunkte $\text{dist}(x_{n+1}, X_n) \geq C\varepsilon_n^{1/\ell}$. Dabei seien $C, \ell \in \mathbb{R}_+$ unabhängig von n .

Satz 3.3.6. Die Voraussetzungen aus 3.3.5 seien erfüllt. Außerdem sei $\bar{\varepsilon}_n := \min_{j=1 \dots n} \varepsilon_j$. Dann gibt es ein $\bar{C} \in \mathbb{R}$, so dass für alle $n \in \mathbb{N}$ gilt:

$$n + 1 \leq \bar{C} \bar{\varepsilon}_n^{-\frac{d}{\ell}}.$$

Beweis: Nach Voraussetzung können wir den Abstand zweier Punkte aus X_{n+1} nach oben abschätzen durch

$$\text{dist}(x_i, x_j) \geq \text{dist}(x_j, X_{j-1}) \geq C\varepsilon_{j-1}^{1/\ell} \geq C\bar{\varepsilon}_n^{1/\ell}, \quad 1 \leq i < j \leq n + 1.$$

Mit der Definition $\hat{\varepsilon}_n := C\bar{\varepsilon}_n^{1/\ell}$ gilt also

$$\hat{\varepsilon}_n \leq \text{dist}(x_i, X_{n+1} \setminus \{x_i\}), \quad \text{für } i \in \{1, \dots, n + 1\}.$$

Durch das folgende Raumfüllargument sieht man, dass dies nur möglich ist, falls $\hat{\varepsilon}_n$ klein beziehungsweise $\text{vol}(\Omega)$ groß genug ist:

$$\begin{aligned} \text{vol}(\Omega) &\geq \text{vol}\left(\Omega \cap \bigcup_{j=1}^{n+1} B_{\frac{\hat{\varepsilon}_n}{2}}(x_j)\right) \\ &= \sum_{j=1}^{n+1} \text{vol}\left(\Omega \cap B_{\frac{\hat{\varepsilon}_n}{2}}(x_j)\right) \\ &\geq \frac{\alpha}{2\pi} \sum_{j=1}^{n+1} \min\left\{\text{vol}\left(B_{\frac{\hat{\varepsilon}_n}{2}}(0)\right), \text{vol}\left(B_r(0)\right)\right\} \\ &= (n + 1) \cdot \underbrace{2^{-d} C_V \frac{\alpha}{2\pi}}_{=: \hat{C}} \min\{\hat{\varepsilon}_n^d, r^d\}. \end{aligned}$$

Dabei ist C_V definiert durch

$$C_V := \frac{\pi^{d/2}}{\Gamma(d/2 + 1)}.$$

Da r nicht von n abhängt, ergibt sich für großes n :

$$\begin{aligned} \text{vol}(\Omega) &\geq (n + 1) \cdot \hat{C} \hat{\varepsilon}_n^d \\ &= (n + 1) \cdot \underbrace{\hat{C} C^d}_{=: C_1} \bar{\varepsilon}_n^{\frac{d}{\ell}}. \end{aligned}$$

Daraus erhalten wir für hinreichend große n die gewünschte Schranke:

$$n + 1 \leq \frac{\text{vol}(\Omega)}{C_1} \bar{\varepsilon}_n^{-\frac{d}{\ell}}.$$

Durch Wahl einer geeigneten Konstante \bar{C} folgt daraus die Behauptung für $n \in \mathbb{N}$.

Bemerkung 3.3.7. Satz 3.3.6 liefert nach Definition 2.1.9 die Konvergenzordnung ℓ . Im Vergleich dazu hatten wir bei Satz 3.2.2 die Konvergenzordnung $d/2$ festgestellt.

Korollar 3.3.8. Sei $\Omega \subset \mathbb{R}^d$ beschränkt und erfülle eine innere Kegelbedingung, $K \in C^2(\Omega \times \Omega)$, symmetrisch, positiv definit, $f \in \mathcal{N}_K(\Omega)$, $\varepsilon_n := \|f - s_{f, X_n}\|_{L^\infty(\Omega)}$, $\bar{\varepsilon}_n := \min_{j=1, \dots, n} \varepsilon_j$. Dann gibt es bei Punktwahl mit f -Greedy ein $C \in \mathbb{R}_+$ mit

$$\bar{\varepsilon}_n \leq Cn^{-\frac{1}{d}}, \quad n \in \mathbb{N}. \quad (3.6)$$

Sei zusätzlich $K \in C^4(\Omega \times \Omega)$ und I_n bezeichne die Anzahl der Punkte aus X_n , die im Inneren von Ω liegen. Dann gibt es ein $C \in \mathbb{R}_+$ mit

$$\bar{\varepsilon}_n \leq C(I_n)^{-\frac{2}{d}}, \quad n \in \mathbb{N}. \quad (3.7)$$

Beweis: Nach Satz 3.3.4 lassen sich die Ableitungen der Fehlerfunktion durch eine von n unabhängige Konstante nach oben abschätzen. Aus Satz 3.3.1 folgt daher, dass es eine Konstante $C_1 \in \mathbb{R}_+$ gibt, so dass ein Mindestabstand der Punkte aus X_n , $n \in \mathbb{N}$, gegeben ist durch

$$\text{dist}(x_{n+1}, X_n) \geq C_1 \varepsilon_n.$$

Liegt x_{n+1} im Inneren von Ω , so gilt nach Satz 3.3.2 sogar

$$\text{dist}(x_{n+1}, X_n) \geq C_1 (\varepsilon_n)^{1/2}.$$

Die Aussagen folgen nun durch Anwendung von Satz 3.3.6. □

Bemerkung 3.3.9. Aus Satz 3.3.6 folgt zwar nicht, für welches $j \in \{1, \dots, n\}$ der Gesamtfehler ε_j minimal wird. Durch das Hinzunehmen weiterer Punkte kann der maximale Fehler auch für einige Schritte ansteigen. Allerdings muss man sowieso den Fehler in jedem Schritt auswerten, um x_{n+1} zu bestimmen. Man könnte demnach den Algorithmus solange durchführen, bis der Fehler das gewünschte Niveau erreicht hat und dann abbrechen.

Zumindest gibt es eine Teilfolge (vgl. Gleichung (3.2)) $(\varepsilon_{\varphi_n})_{n \in \mathbb{N}}$ mit

$$\varepsilon_{\varphi_n} \leq \left(\frac{\text{vol}(\Omega)}{\varphi_n \cdot C_1} \right)^{\ell/d}. \quad (3.8)$$

In Bemerkung 2.1.8 wurde gezeigt, wie diese n -abhängigen Abschätzungen zu den sonst üblichen, von der Füllichte abhängigen Konvergenzraten in Beziehung gesetzt werden können.

3.3: Konvergenzaussagen für f -Greedy

Der folgende Satz zeigt, dass die quadratische Fehlerabschätzung aus Gleichung (3.7) nicht nur für f -Greedy gilt, sondern auch allgemein auf Approximation mit Kernfunktionen auf quasi-uniformen Daten angewendet werden kann. f -Greedy sorgt für eine gute Punktwahl, die aber nicht unbedingt raumfüllend ist und wird daher in der Regel mit deutlich weniger Punkten auskommen (siehe Abschnitt 6.3).

Satz 3.3.10. *Seien $f \in \mathcal{N}_K(\Omega)$, $\Omega \subseteq \mathbb{R}^d$, $X_n := \{x_1, \dots, x_n\} \subset \Omega$, $g_n := f - s_{f, X_n}$, $K \in C^4(\Omega \times \Omega)$, K symmetrisch und positiv definit. Falls der maximale Fehler $\|g_n\|_{L_\infty(\Omega)}$ an einer Stelle y im Inneren von Ω angenommen wird, gibt es ein $C \in \mathbb{R}_+$ unabhängig von n , so dass gilt*

$$\|g_n\|_{L_\infty(\Omega)} \leq Ch(\Omega, X_n)^2 \|f\|_{\mathcal{N}_K(\Omega)},$$

wobei $h(\Omega, X_n)$ die Füllsdichte von X_n in Ω ist.

Beweis: Mit $h^{(i)} := x_i - y$ für ein festes $y \in \Omega$ und dem Satz von Taylor erhalten wir

$$g_n(x_i) = g_n(y) + \nabla g_n(y) h^{(i)} + \frac{1}{2} \sum_{j,k=1}^d \frac{\partial^2 g_n(y + t_i h^{(i)})}{\partial^2 x_j x_k} h_j^{(i)} h_k^{(i)},$$

für geeignete $t_i \in (0, 1)$.

Wir interessieren uns für den maximalen Fehler. Wird dieser an der Stelle y im Inneren von Ω angenommen, so verschwindet der Gradient von g_n und wir erhalten für den Fehler

$$|g_n(y)| \leq |g_n(x_i)| + \frac{1}{2} \left| \sum_{j,k=1}^d \frac{\partial^2 g_n(y + t_i h^{(i)})}{\partial^2 x_j x_k} h_j^{(i)} h_k^{(i)} \right|.$$

Nach Satz 3.3.4 lassen sich die zweiten partiellen Ableitungen der Fehlerfunktion unabhängig von n beschränken,

$$\sum_{j,k=1}^d \sup_{\omega \in \Omega} \frac{\partial^2 g_n(\omega)}{\partial^2 x_j x_k} \leq M \|f\|_{\mathcal{N}_K(\Omega)},$$

und man bekommt die Fehlerabschätzung

$$|g_n(y)| \leq \underbrace{\|g_n\|_{L_\infty(X_n)}}_{=0} + C_M h(\Omega, X_n)^2 \|f\|_{\mathcal{N}_K(\Omega)},$$

mit einer Konstanten $C_M \in \mathbb{R}_+$, die von M aber nicht von n abhängt. \square

Der eigentliche Vorteil des Greedy-Verfahrens besteht darin, dass in den Bereichen, wo f sich „brav“ verhält, das heißt mit wenigen Punkten gut approximiert werden kann, auch nur wenige Punkte verwendet werden. An kritischen Stellen wird dagegen feiner gesampled, so dass mit relativ wenigen Punkten ein kleiner globaler Fehler erreicht wird.

Um dies theoretisch zu untermauern, wird nun Satz 3.3.6 auf den Fall übertragen, dass Ω sich in verschiedene Bereiche Ω_i , $i \in \{1, \dots, M\}$ aufteilen lässt, mit Konstanten $C_i, \ell_i \in \mathbb{R}$, so dass jeweils Voraussetzung 3.3.5 erfüllt ist.

Definition 3.3.11. Es seien $\Omega \subseteq \mathbb{R}^d$, $\Omega = \bigcup_{i=1}^M \Omega_i$, $\Omega_i \cap \Omega_j = \emptyset$, für $i \neq j$, $\text{vol}(\Omega_i) > 0$, $X_n = \{x_1, \dots, x_n\} \subseteq \Omega$, $X_n^{(i)} = X_n \cap \Omega_i$, $X_n^{(i)} = \{x_1^{(i)}, \dots, x_{n_i}^{(i)}\}$, $\sum_{i=1}^M n_i = n$.

Außerdem definieren wir noch:

$$\begin{aligned} \varepsilon_n &:= \|f - I_{X_n} f\|_{\infty, \Omega}, \quad \varepsilon_n^{(i)} := \|f - I_{X_n} f\|_{\infty, \Omega_i}, \\ \bar{\varepsilon}_n^{(i)} &:= \min\{\varepsilon_j^{(i)}, j = 1, \dots, n \text{ mit } \varepsilon_j = \varepsilon_j\}. \end{aligned} \quad (3.9)$$

Voraussetzungen 3.3.12. Für jedes Teilgebiet Ω_i soll die folgende Abstandsbedingung erfüllt sein:

$$\exists C_i, \ell_i \in \mathbb{R}_+ : x_{n+1} \in \Omega_i \Rightarrow \text{dist}(x_{n+1}, X_n^{(i)}) \geq C_i (\varepsilon_n^{(i)})^{\frac{1}{\ell_i}}. \quad (3.10)$$

Zusätzlich erfülle Ω_i die innere Kegelbedingung mit Winkel α_i und Radius r_i .

Satz 3.3.13. Mit den Bezeichnungen aus Definition 3.3.11, der Punktauswahl mit Hilfe von f -Greedy und unter Voraussetzung 3.3.12 existieren Konstanten $\bar{C}_i \in \mathbb{R}$, so dass für alle $n \in \mathbb{N}$ und $i \in \{1, \dots, M\}$ für die Anzahl der Punkte in $X_{n+1}^{(i)}$ gilt:

$$(n+1)_i \leq \bar{C}_i \bar{\varepsilon}_n^{(i)^{(-d/\ell_i)}}.$$

Beweis: Für ein festes n sei i so gewählt, dass x_{n+1} in Ω_i liegt, also $\varepsilon_n^{(i)} = \varepsilon_n$.

Als Abschätzung für den Abstand $\text{dist}(x_{n+1}, X_n^{(i)})$ gilt nach Voraussetzung:

$$\text{dist}(x_{n+1}, X_n^{(i)}) \geq C_i (\varepsilon_n^{(i)})^{\frac{1}{\ell_i}}. \quad (3.11)$$

Um eine untere Schranke für den Abstand zweier Punkte aus Ω_i zu bestimmen, genügt es, wenn wir nur diejenigen $\varepsilon_j^{(i)}$ betrachten, die gleich dem maximalen Fehler auf Ω bei Approximation mit X_j sind. Denn nur in diesem Fall wird im nächsten Schritt ein weiterer Punkt in Ω_i hinzugefügt.

Wir definieren $\hat{\varepsilon}_n^{(i)} := C_i \bar{\varepsilon}_n^{(i)^{\frac{1}{\ell_i}}}$ und bekommen

$$\hat{\varepsilon}_n^{(i)} \leq \text{dist}(x_k^{(i)}, X_{(n+1)}^{(i)} \setminus \{x_k^{(i)}\}), \quad \forall k \in \{1, \dots, (n+1)_i\}.$$

Damit erhält man

$$\begin{aligned} \text{vol}(\Omega_i) &\geq \text{vol} \left(\Omega_i \cap \bigcup_{j=1}^{(n+1)_i} B_{\frac{\hat{\varepsilon}_n^{(i)}}{2}}(x_j^{(i)}) \right) \\ &= \sum_{j=1}^{(n+1)_i} \text{vol} \left(\Omega_i \cap B_{\frac{\hat{\varepsilon}_n^{(i)}}{2}}(x_j^{(i)}) \right) \\ &\geq \frac{\alpha_i}{2\pi} \sum_{j=1}^{(n+1)_i} \min\{\text{vol}(B_{\frac{\hat{\varepsilon}_n^{(i)}}{2}}(0)), \text{vol}(B_{r_i}(0))\} \\ &= (n+1)_i \cdot \hat{C}_i \min\{(\hat{\varepsilon}_n^{(i)})^d, r_i^d\}, \quad \text{für geeignetes } \hat{C}_i \in \mathbb{R}. \end{aligned}$$

3.3: Konvergenzaussagen für f -Greedy

Für großes n_i und geeignetes $\bar{C}_i \in \mathbb{R}$ gilt:

$$\begin{aligned} \text{vol}(\Omega_i) &\geq (n+1)_i \cdot \widehat{C}_i \left(\widehat{\varepsilon}_n^{(i)}\right)^d \\ &= (n+1)_i \cdot \bar{C}_i \bar{\varepsilon}_n^{(i)(d/\ell_i)}, \\ \Rightarrow (n+1)_i &\leq \frac{\text{vol}(\Omega_i)}{\bar{C}_i} \bar{\varepsilon}_n^{(i)(-d/\ell_i)}. \quad \square \end{aligned}$$

Anstelle der Abstandsbedingung aus Voraussetzung 3.3.12 kann man auch fordern, dass das Greedy-Verfahren auf Ω_i die Punkte quasi-uniform verteilt und dass der Fehler auf Ω_i bis auf einen konstanten Faktor durch die ℓ_i -te Potenz der Füllichte beschränkt werden kann.

Voraussetzungen 3.3.14. Die Punkte aus $X_n^{(i)} := X_n \cap \Omega_i$ seien quasi-uniform verteilt (vgl. 2.1.5):

$$\exists D_i : D_i q(X_n^{(i)}) \geq h(\Omega_i, X_n^{(i)}), \quad \forall n \in \mathbb{N}. \quad (3.12)$$

Zusätzlich gelte für jedes Teilgebiet $\Omega_i \subset \mathbb{R}^d$ die Fehlerabschätzung

$$\exists C_i, \ell_i \in \mathbb{R} : \|f - I_{X_n} f\|_{\infty, \Omega_i} \leq C_i h(\Omega_i, X_n^{(i)})^{\ell_i}, \quad \forall n \in \mathbb{N}. \quad (3.13)$$

Außerdem erfülle Ω_i die innere Kegelbedingung mit Winkel α_i und Radius r_i .

Satz 3.3.15. Mit den Bezeichnungen aus Definition 3.3.11, der Punktauswahl mit Hilfe des Greedy-Verfahrens und unter Voraussetzung 3.3.14 existieren Konstanten $\bar{C}_i \in \mathbb{R}$, so dass für alle $n \in \mathbb{N}$, $i \in \{1, \dots, M\}$ gilt:

$$n_i \leq \bar{C}_i \bar{\varepsilon}_n^{(i)(-d/\ell_i)}.$$

Beweis: Die Aussage folgt aus Satz 3.3.13, denn für $x_{n+1} \in \Omega_i$ gilt:

$$\begin{aligned} \text{dist}(x_{n+1}, X_n^{(i)}) &\geq q(X_{(n+1)}^{(i)}) \\ &\stackrel{(3.12)}{\geq} \frac{h(\Omega_i, X_{(n+1)}^{(i)})}{D_i} \\ &\stackrel{(3.13)}{\geq} \left(\frac{\|f - I_{X_{n+1}} f\|_{\infty, \Omega_i}}{C_i D_i^{\ell_i}} \right)^{1/\ell_i} \\ &\stackrel{\text{Def. 3.3.11}}{=} \left(\frac{\varepsilon_{n+1}^{(i)}}{C_i D_i^{\ell_i}} \right)^{1/\ell_i}. \quad \square \end{aligned}$$

Bemerkung 3.3.16. Falls die Fehlerabschätzung (3.13) nur für $h(\Omega_i, X_n^{(i)}) < h^{(i)}$ mit gegebenen $h^{(i)} \in \mathbb{R}_+$ gilt, erhält man für den Abstand zweier Punkte aus $X_{n+1}^{(i)}$ die Schranke

$$\|x_i - x_j\|_2 \geq \min \left\{ h^{(i)}, C_i \bar{\varepsilon}_n^{(i) \frac{1}{\ell_i}} \right\}.$$

Da $h^{(i)}$ nicht von n abhängt, gilt in diesem Fall aber genauso die Aussage des Satzes.

3.3: Konvergenzaussagen für f -Greedy

Im Folgenden wollen wir untersuchen, wie sich der Diskretisierungsfehler auf das Greedy-Verfahren auswirkt. Anstatt die Datenpunkte aus Ω auszuwählen, wird man in der Praxis die Fehlerfunktion höchstens auf einer Punktwolke, zum Beispiel auf einem feinen Gitter, kennen. Für hinreichend kleine Fülldichte dieser Punktwolke in Ω wollen wir eine ähnliche Konvergenzaussage wie in Gleichung (3.6) zeigen, wo das Extremum der Fehlerfunktion auf Ω gesucht wird. Dabei ist es wichtig, dass wir zwischen der Fülldichte der ausgewählten Punkte $X_n \subset \Omega$ und der Fülldichte der zur Verfügung stehenden Punkte unterscheiden (siehe auch Definition 3.1.1 (f -Greedy)).

Definition 3.3.17. Seien f, I_{X_n} wie zu Beginn von Kapitel 3.1.1 gegeben mit Fehlerfunktion $g_0 := f$, $g_n := f - I_{X_n}f$, für $n \in \mathbb{N}$. Außerdem sei P_δ eine beliebige, diskrete Teilmenge von Ω mit Fülldichte $h(P_\delta, \Omega) = \delta$. Als **Diskretes f -Greedy** auf der Punktmenge P_δ bezeichnen wir jede Methode zur Auswahl der Datenpunkte $X_n := \{x_1, \dots, x_n\}$ aus P_δ mit der Eigenschaft

$$|g_n(x_{n+1})| = \|g_n\|_{L_\infty(P_\delta)}, \quad n \in \mathbb{N}_0.$$

Wir wählen eine beliebige Extremstelle der Fehlerfunktion g_n aus Ω und bezeichnen sie durch ω_{n+1} , $n \in \mathbb{N}_0$,

$$|g_n(\omega_{n+1})| = \|g_n\|_{L_\infty(\Omega)}, \quad n \in \mathbb{N}_0.$$

Schließlich definieren wir noch

$$\begin{aligned} \varepsilon_{n, P_\delta} &:= |g_n(x_{n+1})|, \\ \varepsilon_{n, \Omega} &:= |g_n(\omega_{n+1})|, \quad n \in \mathbb{N}_0. \end{aligned}$$

Satz 3.3.18. Sei $n \in \mathbb{N}$, $c \in \mathbb{R}_+$ und für die Fülldichte des Diskreten Greedy-Verfahrens aus Definition 3.3.17 gelte $\delta \leq c\varepsilon_{n, \Omega}$.

Außerdem sei $M_n := \sup_{\omega \in \Omega} \|\nabla g_n(\omega)\|_2 < \infty$. Dann ist der Abstand der ausgewählten Punkte nach unten beschränkt durch

$$\|x_{n+1} - x_i\| \geq \left(\frac{1}{M_n} - c \right) \varepsilon_{n, \Omega}, \quad 1 \leq i \leq n.$$

Beweis: Nach Voraussetzung existiert ein $p_{n+1} \in P_\delta$ mit $\|p_{n+1} - \omega_{n+1}\|_2 \leq \delta$. Daraus erhalten wir nach dem Mittelwertsatz

$$|g_n(p_{n+1}) - g_n(\omega_{n+1})| \leq \delta \sup_{\omega \in \Omega} \|\nabla g_n(\omega)\|_2 \leq \delta M_n.$$

Da x_{n+1} als Extremstelle der Fehlerfunktion g_n auf P_δ gewählt wird, gilt also für $i \in \{1, \dots, n\}$

$$|g_n(x_{n+1}) - g_n(x_i)| = |g_n(x_{n+1})| \geq |g_n(p_{n+1})| \geq |g_n(\omega_{n+1})| - \delta M_n.$$

Wiederum durch Anwenden des Mittelwertsatzes sieht man

$$\|x_{n+1} - x_i\|_2 \geq \frac{\varepsilon_{n, \Omega}}{M_n} - \delta$$

und für $\delta \leq c\varepsilon_{n, \Omega}$ ergibt sich die Behauptung. □

3.4 Höhere Konvergenzordnungen bei f -Greedy in \mathbb{R}

Die Konvergenzaussage aus Satz 3.3.13 möchten wir gerne im Allgemeinen auf das f -Greedy-Verfahren mit Kernen übertragen. Dabei stellt sich heraus, dass die Annahmen aus Voraussetzung 3.3.12 zu stark sind, da ein bestimmter Mindest-Füllabstand vorausgesetzt wird.

Die Abschätzungen gelten auf einer Menge Ω nur dann, wenn $h(X_n, \Omega) \leq h_0$ gilt. Die Konstante h_0 hängt dabei unglücklicherweise von Ω ab. Dagegen werden aber Aussagen über die Abhängigkeit des Fehlers von dem Separationsabstand $q(X_n)$ benötigt, damit die Voraussetzungen aus 3.3.12 gezeigt werden können.

Numerische Tests zeigen (siehe Kapitel 6), dass f -Greedy nur dort viele Punkte setzt, wo die Funktion hohe Ableitungen hat, ihr Verhalten ändert (Extremstellen) oder zum Beispiel wegen des Gebietsrandes nicht gut approximiert werden kann. Dieses wünschenswerte Verhalten sorgt aber dafür, dass die Mengen, die gefüllt werden, nahezu beliebig aussehen können. Damit lassen sich die h -abhängigen Abschätzungen nicht anwenden, da nicht klar ist, wie sich die Füllichte mit steigender Punktzahl verhält. Im 1-dimensionalen Fall ist die Situation einfacher. Da es keine Bereiche kleinerer Dimension gibt, in denen sich Punkte anhäufen können, werden automatisch Teilbereiche gefüllt, falls sich irgendwo Punkte sammeln. Deshalb soll dieser Fall im Folgenden separat betrachtet werden.

Für den 1-dimensionalen Fall, $\Omega \subset \mathbb{R}$, möchten wir nun einen Ausweg vorstellen, mit dessen Hilfe sich die vollen Konvergenzordnungen² für f -Greedy zeigen lassen. Dazu treffen wir zunächst Annahmen über das Approximationsverhalten auf Würfeln im \mathbb{R}^d , unter denen sich die volle Konvergenzordnung zeigen lässt. Anschließend wird bewiesen, dass diese Annahmen bei Verwendung von Kernfunktionen auf Intervallen erfüllt sind.

Definition 3.4.1. Der Würfel $W_r(x)$ mit Seitenlänge $2r$ und Mittelpunkt x sei definiert durch

$$W_r(x) = \{y \in \mathbb{R}^d : \|x - y\|_\infty \leq r\}.$$

Voraussetzungen 3.4.2. Es existiere ein $M \in \mathbb{N}$ und ein $C = C(f, M) \in \mathbb{R}_+$, so dass die Fehlerabschätzung

$$\|f - I_X f\|_{L_\infty(W_r(x))} \leq Cr^\ell$$

gilt, für alle $r > 0$, $x \in \Omega$ und Würfel $W_r(x) \subset \Omega$, in denen mindestens M Datenpunkte aus X liegen.

Satz 3.4.3. Es seien $\Omega = [a, b]^d$ mit $a, b \in \mathbb{R}$, $b > a$, $f \in \mathcal{N}_K(\Omega)$ und $\varepsilon_n := \|f - I_{X_n} f\|_{L_\infty(\Omega)}$ für $n \in \mathbb{N}$. Die Datenpunkte $X_n := \{x_1, \dots, x_n\} \subset \Omega$ seien nach

²„Volle Konvergenzordnung“ soll hier bedeuten, dass bis auf einen konstanten Faktor nicht mehr Punkte als bei einem Gitter für ein gegebenes Fehlerlevel benötigt werden (siehe auch 2.1.9).

3.4: Höhere Konvergenzordnungen bei f -Greedy in \mathbb{R}

f -Greedy (siehe Definition 3.1.1 (**f -Greedy**)) ausgewählt. Außerdem gelte die Voraussetzung 3.4.2. Dann gibt es eine Teilfolge $(\varphi_n)_{n \in \mathbb{N}}$, für die gilt

$$\varphi_n \leq \overline{C} (\varepsilon_{\varphi_n})^{-\frac{d}{\ell}},$$

wobei die Konstante $\overline{C} \in \mathbb{R}_+$ nur von M , Ω , d und C abhängt.

Beweis: Sei $k \in \mathbb{N}$, $r = \frac{b-a}{k}$. Ω lässt sich wie folgt in Würfel aufteilen

$$\begin{aligned} \Omega &= \bigcup_{z \in \mathbb{Z}^d} (W_{\frac{r}{2}}(rz) \cap \Omega) \\ &= \bigcup_{\substack{z \in \mathbb{Z}^d \\ rz \in [a - \frac{r}{2}, b + \frac{r}{2}]^d}} (W_{\frac{r}{2}}(rz) \cap \Omega). \end{aligned}$$

Die Anzahl dieser Würfel lässt sich beschränken durch

$$\#\{z \in \mathbb{Z}^d : rz \in [a - \frac{r}{2}, b + \frac{r}{2}]^d\} \leq (k+2)^d.$$

Demnach sind für $n = M(k+2)^d$ Punkte entweder in jedem Würfel genau M Datenpunkte oder es gibt mindestens einen Würfel W , der mehr als M Punkte enthält. Im ersten Fall kann man für jeden Würfel die Voraussetzung anwenden und bekommt

$$\varepsilon_n \leq Cr^\ell.$$

Im zweiten Fall gibt es eine Teilmenge $X_{n_0} \subseteq X_n$ mit $n_0 \leq n$ Punkten und $\varepsilon_{n_0} \leq Cr^\ell$. Denn für ein $n_0 \leq n$ sind in W bereits M Punkte enthalten und trotzdem wird ein weiterer Punkt hinzugefügt. Mit der Voraussetzung erhält man

$$\varepsilon_{n_0} = \|f - I_{X_{n_0}} f\|_{L^\infty(W)} \leq Cr^\ell.$$

Für ein beliebiges $r \in (0, 1]$ gibt es deshalb in beiden Fällen ein $n(r) \in \mathbb{N}$ mit $\varepsilon_{n(r)} \leq Cr^\ell$ und

$$n(r) \leq M(k+2)^d \leq M \left(\frac{b-a+2}{r} \right)^d \leq M \left(\frac{b-a+2}{(\frac{\varepsilon_{n(r)}}{C})^{1/\ell}} \right)^d.$$

Damit gilt für ein geeignetes \overline{C} , das nicht von $n(r)$ oder r abhängt,

$$n(r) \leq \overline{C} (\varepsilon_{n(r)})^{-\frac{d}{\ell}}.$$

Die Aussage erhalten wir nun, indem wir r immer kleiner werden lassen und jeweils ein geeignetes $n(r)$ auswählen. \square

Satz 3.4.4. Sei $[a, b] \subseteq \Omega \subseteq \mathbb{R}$, $K \in C^{2\ell}(\Omega \times \Omega)$, $f \in \mathcal{N}_K(\Omega)$, $h := b - a$ und $X_n := \{x_1, \dots, x_n\}$, wobei in $X_n \cap [a, b]$ mindestens ℓ verschiedene Punkte, x_1, \dots, x_ℓ ,

3.4: Höhere Konvergenzordnungen bei f -Greedy in \mathbb{R}

liegen. Dann gilt für die Fehlerfunktion $g_n := s_{f, X_n} - f$ mit ℓ -ter Ableitung $g_n^{(\ell)}$ die Abschätzung

$$|g_n(x)| \leq h^\ell \|g_n^{(\ell)}\|_{L_\infty([a,b])}, \quad \forall x \in [a, b]. \quad (3.14)$$

Beweis:

Teil 1: Zu $r \in \{0, \dots, \ell - 1\}$, $k \in \{1, \dots, \ell - r\}$ gibt es $y_k^{(r)} \in [a, b]$ mit

$$y_i^{(r)} \neq y_j^{(r)} \text{ für } i \neq j \text{ und } g_n^{(r)}(y_k^{(r)}) = 0. \quad (3.15)$$

Beweis Teil 1:

$r = 0$: Setze $y_k^{(0)} := x_k$ nach Voraussetzung gilt $g_n^{(0)}(x_k) = 0$.

$r > 0$: Wegen $g_n^{(r-1)}(y_k^{(r-1)}) = g_n^{(r-1)}(y_{k+1}^{(r-1)}) = 0$ für $1 \leq k \leq \ell - r$ gilt nach dem Satz von Rolle, dass es ein $y_k^{(r)} \in (y_k^{(r-1)}, y_{k+1}^{(r-1)})$ mit $g_n^{(r)}(y_k^{(r)}) = 0$ gibt.

Teil 2: Zu $x \in [a, b]$ und $r \in \{0, \dots, \ell - 1\}$ gibt es ein $\xi^{(r)} \in [a, b]$ mit

$$|g_n^{(r)}(x)| \leq |g_n^{(r+1)}(\xi^{(r)})| h.$$

Beweis Teil 2: Wir entwickeln $g_n^{(r)}$ um die Nullstelle $y_1^{(r)}$ aus dem ersten Beweisteil. Nach der Taylorformel gibt es ein $\xi^{(r)}$ zwischen x und $y_1^{(r)}$ mit

$$\begin{aligned} g_n^{(r)}(x) &= \underbrace{g_n^{(r)}(y_1^{(r)})}_{=0} + g_n^{(r+1)}(\xi^{(r)})(y_1^{(r)} - x) \\ \Rightarrow |g_n^{(r)}(x)| &\leq |g_n^{(r+1)}(\xi^{(r)})| h. \end{aligned}$$

Durch wiederholtes Anwenden von Teil 2 erhalten wir die Behauptung. \square

Nach Satz 3.3.4 lassen sich die Ableitungen der Fehlerfunktion durch die Native Space Norm von f und die Ableitungen der Kernfunktion beschränken, sind also unabhängig von X_n . Zusammen mit Satz 3.4.3 und 3.4.4 ergibt sich daraus das folgende Korollar.

Korollar 3.4.5. Sei Ω ein reelles Intervall, $K \in C^{2\ell}(\Omega \times \Omega)$ und $f \in \mathcal{N}_K(\Omega)$. Dann gibt es bei Punktwahl durch f -Greedy eine Teilfolge $(\varphi_n)_{n \in \mathbb{N}}$ mit

$$\varphi_n \leq C (\varepsilon_{\varphi_n})^{-\frac{1}{\ell}}.$$

Dabei ist $\varepsilon_n := \|f - s_{f, X_n}\|_{L_\infty(\Omega)}$ und $C \in \mathbb{R}_+$ hängt nur von f , Ω und K , aber nicht von φ_n ab.

Beweis: Voraussetzung 3.4.2 ist mit $M = \ell$ und $C = 2\|f\|_{\mathcal{N}_K(\Omega)} \|D_2^\alpha K(\cdot, x)\|_{\mathcal{N}_K(\Omega)}$ erfüllt. Denn sei $[a, b] \subseteq \Omega$ ein Intervall der Länge r und $[a, b]$ enthalte ℓ verschiedene Punkte aus X_n . Dann gilt für $x \in [a, b]$

$$|g_n^{(\ell)}(x)| \stackrel{3.4.4}{\leq} \|g_n^{(\ell)}\|_{L_\infty([a,b])} r^\ell \leq \|g_n^{(\ell)}\|_{L_\infty(\Omega)} r^\ell \stackrel{3.3.4}{\leq} 2\|f\|_{\mathcal{N}_K(\Omega)} \|D_2^\alpha K(\cdot, x)\|_{\mathcal{N}_K(\Omega)} r^\ell.$$

Die Behauptung folgt damit aus Satz 3.4.3. \square

Bemerkung 3.4.6. Das Korollar zeigt, dass mit der Punktwahl durch f -Greedy bei Interpolation auf einem Intervall mindestens die gleiche Konvergenzordnung erzielt wird wie bei der Verwendung äquidistanter Punkte (vgl. Satz 11.13 aus [22]).

Allerdings beobachtet man häufig bei der Interpolation mit Radialen Basisfunktionen einen sogenannten Randeffect, wodurch die Konvergenz am Rand deutlich schlechter ist, als im Inneren des Gebiets (das scheint auch plausibel, weil am Rand wesentlich weniger Punkte liegen als im Inneren). Dieser Effekt tritt bei f -Greedy nicht auf, da automatisch an den Stellen, an denen der Fehler groß ist, mehr Punkte ausgesucht werden. Durch dieses Oversampling am Rand wird also der Randeffect ausgeglichen.

Bei den Tests in Abschnitt 6.2 konnte bei Verwendung von f -Greedy im Vergleich zu äquidistanten Punkten tatsächlich die doppelte Konvergenzordnung festgestellt werden.

Bemerkung 3.4.7. Leider lässt sich Satz 3.3.4 nicht lokal anwenden. Wenn f nur auf einer Teilmenge Γ von Ω sehr glatt ist, möchten wir gerne daraus schließen, dass hohe Ableitungen der Fehlerfunktion auf Γ beschränkt bleiben. Allerdings beeinflussen die Datenpunkte global die Interpolante. Dies kann sogar dazu führen, dass durch das Hinzufügen eines Interpolationspunktes der Gesamtfehler vergrößert wird (siehe auch Abschnitt 6.2). Es ist zu erwarten, dass speziell die Punkte mit kleinem Abstand zu Γ Einfluss auf die Fehlerfunktion in Γ haben. Allerdings ist unklar, wie groß dieser Einfluss in Abhängigkeit des jeweiligen Abstandes ist.

Kapitel 4

Eine Newton-Basis des Native Space

4.1 Einführung

Ein Problem bei der Interpolation mit Radialen Basisfunktionen besteht darin, dass es leicht zu Instabilitäten kommen kann. Wir unterscheiden dabei zwei Arten von Instabilitäten.

Zum Einen kann es passieren, dass die Kondition der Interpolationsmatrix mit steigender Punktzahl dramatisch ansteigt. Deutlich wird dies vor allem, wenn zwei Interpolationspunkte dicht zusammenliegen. In diesem Fall hat die Interpolationsmatrix zwei nahezu identische Zeilen und somit eine hohe Konditionszahl.

Aber auch sonst steigt die Kondition bei wachsender Punktzahl, besonders bei Verwendung glatter Basisfunktionen. Hierbei gibt es einen „Trade-off“ zwischen der Interpolationsgüte und der Stabilität. In [11] wird gezeigt, dass es prinzipiell nicht möglich ist, gleichzeitig die Powerfunktion (und damit die obere Schranke des Interpolationsfehlers) und die Kondition der Interpolationsmatrix beliebig klein zu haben. Es gilt nämlich

$$\frac{P_{K, X_n}^2(x)}{\lambda(x)} \geq 1. \quad (4.1)$$

Dabei bezeichnet $\lambda(x)$ den kleinsten Eigenwert der Interpolationsmatrix aus Definition 2.2.4 für die Punkte x_1, \dots, x_n, x .

Etwas ähnliches lässt sich auch beobachten, wenn die Skalierung der Basisfunktion verändert wird. Mit flacher Basisfunktion steigt in der Regel die Approximationsgüte, während bei steileren Basisfunktionen die Kondition der Matrix besser ist.

Die zweite Instabilität, die wir hier betrachten, entsteht durch betragsgroße Koeffizienten beim Lösen des Gleichungssystems und tritt ebenfalls auf, wenn die Matrix schlecht konditioniert ist. Große Koeffizienten mit entgegengesetzten Vorzeichen führen zu Auslöschungen und damit kommt es zu **Auswertungsinstabilitäten**. Dieser

Zusammenhang soll hier nochmal kurz ausgeführt werden.

Wir betrachten die Auswertungsabbildung $x \mapsto \sum_j \alpha_j K(x, x_j)$ an den Interpolationspunkten $X_n = \{x_1, \dots, x_n\}$ im Fall von *relativen Fehlern* $|\varepsilon_j|, |\delta_j| \leq \varepsilon$ in den Koeffizienten und bei Auswertung der Basisfunktion. Der absolute Auswertungsfehler lässt sich dann beschränken durch

$$\begin{aligned} & \left\| \sum_j \alpha_j K(\cdot, x_j) - \sum_j \alpha_j (1 + \varepsilon_j) K(\cdot, x_j) (1 + \delta_j) \right\|_{L_\infty(\Omega)} \\ & \leq \sum_j |\varepsilon_j + \delta_j| |\alpha_j| |K(\cdot, x_j)| + O(\varepsilon^2) \\ & \leq 2\varepsilon \sum_j |\alpha_j| |K(\cdot, x_j)| + O(\varepsilon^2). \end{aligned}$$

Hier wird deutlich, dass betragsmäßig große Koeffizienten zu schlechteren Schranken führen. Außerdem spielt aber auch die sogenannte **Lebesguekonstante**, definiert durch

$$L_{X_n, K} := \sum_{j=1}^n \|K(\cdot, x_j)\|_{L_\infty(\Omega)} \quad (4.2)$$

eine wichtige Rolle, weil der obige absolute Fehler sich durch $2\varepsilon \|\alpha\|_\infty L_{X_n, K} + O(\varepsilon^2)$ abschätzen lässt.

Die genannten Stabilitätsprobleme treten nicht so stark bei rauheren Basisfunktionen auf. Auch bei Verwendung von Basisfunktionen mit kompaktem Träger (siehe [22]) stellt sich heraus, dass sich das auftretende Gleichungssystem relativ stabil lösen lässt.

Allerdings liefern gerade die glatten Basisfunktionen gute Approximationseigenschaften. Deshalb soll hier untersucht werden, wodurch diese Instabilitäten entstehen und wie sie vermieden werden können. Die Arbeit folgt hier im Wesentlichen der Publikation [8].

Stellt man die Interpolante s_{f, X_n} an $f \in \mathcal{N}_K(\Omega)$ in der Lagrange-Basis dar, $s_{f, X_n} = \sum_{j=1}^n f(x_j) u_j(\cdot)$, erhält man offensichtlich beschränkte Koeffizienten. Nach Satz 1 aus [2] gilt damit die folgende Aussage.

Satz 4.1.1. *Sei $\Omega \subset \mathbb{R}^d$ beschränkt und erfülle eine innere Kegelbedingung. Die Datenpunkte aus X_n seien quasi-uniform verteilt. Außerdem sei der Native Space $\mathcal{N}_K(\Omega)$ normäquivalent zu W_2^τ , $\tau \in \mathbb{N}$, $\tau > d/2$. Dann gibt es eine Konstante $C > 0$, die von K , Ω und d abhängt, so dass gilt*

$$\begin{aligned} \sum_{j=1}^n |f(x_j)| \|u_j(\cdot)\|_{L_\infty(\Omega)} & \leq \|f\|_{L_\infty(X_n)} \sum_{j=1}^n \|u_j(\cdot)\|_{L_\infty(\Omega)} \\ & \leq C \sqrt{n} \|f\|_{L_\infty(X_n)}. \end{aligned}$$

4.2: Die Newton-Basis

Als Interpolationsmatrix entsteht die Einheitsmatrix, hat also optimale Kondition. Auf diese Weise sieht man, dass die auftretenden Instabilitäten durch einen geschickten Basiswechsel gelöst werden können.

Allerdings ist das Aufstellen der Lagrange-Basis genauso schwierig, als würde man das Interpolationsproblem direkt lösen. Deshalb suchen wir im Folgenden nach einem Mittelweg zwischen der einfach zu berechnenden Kernbasis $\{K(\cdot, x_j)\}_j$ und der stabilen Lagrange-Basis $\{u_j(\cdot)\}_j$. Dazu erinnern wir uns an die Eigenschaften der Newton-Basis zur Polynominterpolation und definieren eine Newton-Basis für den Native Space.

Der Rechenaufwand wird sich dabei als nicht wesentlich geringer als für die Lagrange-Basis herausstellen, aber die Newton-Basis erfordert bei Hinzufügen neuer Punkte keine Neuberechnung der „alten“ Basisfunktionen. Das macht sie für Greedy-Verfahren besonders interessant. Darüberhinaus hat die Newton-Basis unerwartete theoretische Eigenschaften, die im Folgenden dargestellt werden sollen.

4.2 Die Newton-Basis

Definition 4.2.1. Sei $\Omega \subset \mathbb{R}^d$, Ω beschränkt, $K : \Omega \times \Omega \rightarrow \mathbb{R}$ positiv definit, symmetrisch. Wir definieren die **Newton-Basis** $\{v_j\}_{j=1}^n$ bezüglich des Kerns K durch

$$\begin{aligned} v_j(x_i) &= 0, & 1 \leq i < j \leq n, \\ v_j(x_j) &= 1, & 1 \leq j \leq n, \end{aligned}$$

mit $v_j \in U_{X_j, K} := \text{span}\{K(\cdot, x_\ell) : x_\ell \in X_j\}$, $1 \leq j \leq n$.

Bemerkung 4.2.2.

- Die so definierte Basis hängt zwar von der Sortierung der Elemente aus X_n ab, ist aber ansonsten wegen der positiven Definitheit des Kerns K eindeutig bestimmt (vgl. [22], Satz 11.1).
- Die v_j , $1 \leq j \leq n$, sind linear unabhängig, denn sonst gäbe es k und j , $k < j$ mit $v_j \in U_{X_k, K}$. Dann wäre aber v_j wegen $v_j(x_i) = 0$, $1 \leq i < j \leq n$, die Nullfunktion und die Bedingung $v_j(x_j) = 1$ wäre nicht erfüllt.

Definition 4.2.3. Sei $\Omega \subset \mathbb{R}^d$. Für $f \in \mathcal{N}_K(\Omega)$ definieren wir das Koeffizientenfunktional $\lambda_j(f)$, $1 \leq j \leq n$, durch die Gleichungen

$$f(x_j) = \sum_{k=1}^j \lambda_k(f) v_k(x_j), \quad 1 \leq j \leq n. \quad (4.3)$$

Satz 4.2.4. Die Koeffizientenfunktionale $\lambda_j(\cdot)$, $1 \leq j \leq n$, lassen sich rekursiv bestimmen durch

$$\begin{aligned}\lambda_1(f) &= f(x_1), \\ \lambda_j(f) &= f(x_j) - \sum_{k=1}^{j-1} \lambda_k(f)v_k(x_j).\end{aligned}$$

Insbesondere gilt $\lambda_i(v_j) = \delta_{ij}$, $1 \leq i \leq j$.

Beweis: Der erste Teil des Satzes folgt aus

$$f(x_j) = \lambda_j(f) \underbrace{v_j(x_j)}_{=1} + \sum_{k=1}^{j-1} \lambda_k(f)v_k(x_j)$$

und $\lambda_i(v_j) = \delta_{ij}$ gilt wegen

$$\begin{aligned}\lambda_1(v_j) &= v_j(x_1) = \delta_{1j} \text{ nach Definition von } v_j, \\ \lambda_i(v_j) &= v_j(x_i) - \sum_{k=1}^{i-1} \underbrace{\lambda_k(v_j)}_{=0} v_k(x_j) = \delta_{ij}. \quad \square\end{aligned}$$

Satz 4.2.5. Die Interpolante s_{f,X_n} zu $f \in \mathcal{N}_K(\Omega)$, $\Omega \subset \mathbb{R}^d$ ist gegeben durch

$$s_{f,X_n}(x) = \sum_{k=1}^n \lambda_k(f)v_k(x).$$

Beweis: Es gilt $s_{f,X_n}, \sum_{k=1}^n \lambda_k(f)v_k(\cdot) \in U_{X_n,K}$ (siehe 4.2.1) und

$$s_{f,X_n}(x_j) = f(x_j) = \sum_{k=1}^n \lambda_k(f)v_k(x_j), \quad 1 \leq j \leq n,$$

wegen $v_k(x_j) = 0$ für $k > j$. Durch n Punkte ist aber jede Funktion aus $U_{X_n,K}$ eindeutig festgelegt, da K positiv definit ist. \square

Durch Einsetzen von 4.2.5 in 4.2.4 erhalten wir

Korollar 4.2.6. Die Koeffizientenfunktionale $\lambda_j(\cdot)$, $1 \leq j \leq n$, lassen sich rekursiv bestimmen durch

$$\begin{aligned}\lambda_1(f) &= f(x_1), \\ \lambda_j(f) &= f(x_j) - s_{f,X_{j-1}}(x_j).\end{aligned}$$

4.2: Die Newton-Basis

Bemerkung 4.2.7. Bei Polynominterpolation in \mathbb{R} lässt sich die Interpolante P_{f, X_n} in der üblichen Notation darstellen durch

$$P_{f, X_n} = \sum_{k=0}^n \underbrace{\left(\prod_{j=0}^{k-1} (x - x_j) \right)}_{=: v_k(x)} \underbrace{[x_0, \dots, x_k] f}_{=: \lambda_k(f)}$$

was Satz 4.2.5 entspricht. Korollar 4.2.6 zeigt, dass die λ_k genau wie die dividierten Differenzen einer Rekursionsformel genügen.

Nun beschäftigen wir uns mit der Frage, wie sich die Elemente der Newton-Basis bestimmen lassen.

Satz 4.2.8. Die Elemente der Newton-Basis sind gegeben durch

$$v_1(x) = \frac{K(x, x_1)}{\beta_{11}},$$

$$v_j(x) = \frac{K(x, x_j)}{\beta_{jj}} - \sum_{k=1}^{j-1} \frac{\beta_{jk}}{\beta_{jj}} v_k(x), \quad \beta_{jk} \in \mathbb{R}, \quad 1 \leq k < j \leq n,$$

dabei erhält man die Koeffizienten β_{jk} aus der LR-Zerlegung der Kernmatrix

$$(K(x_i, x_j))_{i,j} = \begin{pmatrix} \beta_{11} & & 0 \\ \vdots & \ddots & \\ \beta_{n1} & \dots & \beta_{nn} \end{pmatrix} \begin{pmatrix} v_1(x_1) & \dots & v_1(x_n) \\ & \ddots & \vdots \\ 0 & & v_n(x_n) \end{pmatrix}.$$

Die positive Definitheit der Matrix garantiert, dass die Koeffizienten β_{jj} von Null verschieden sind, außerdem ist wegen $v_j(x_j) = 1$ die LR-Zerlegung eindeutig.

Beweis: Da $\{v_k\}_{k=1}^j$ eine Basis von $U_{X_j, K}$ (vgl. Definition 4.2.1) ist, gibt es Koeffizienten β_{jk} , $1 \leq k \leq j \leq n$, mit

$$K(x, x_j) = \sum_{k=1}^j \beta_{jk} v_k(x). \quad (4.4)$$

Die angegebene Matrixdarstellung erhält man aus

$$K(x_j, x_i) = K(x_i, x_j) = \sum_{k=1}^j \beta_{jk} v_k(x_i). \quad \square \quad (4.5)$$

Da die Kernmatrix A positiv definit ist, ist es möglich statt einer LR-Zerlegung die Cholesky-Zerlegung $A = LL^T$ anzuwenden. Ausgehend von dieser Zerlegung sollen nun die Koeffizienten der Newton-Basis bestimmt werden.

Korollar 4.2.9. Sei $A = LL^T$ die Cholesky-Zerlegung der Kernmatrix mit $A = (K(x_i, x_j))_{ij}$, $L = (L_{ij})_{ij}$. Mit den Bezeichnungen aus Satz 4.2.8 gilt dann

$$\beta_{ij} = L_{jj}L_{ij}, \quad v_j(x_i) = \frac{L_{ij}}{L_{jj}}, \quad 1 \leq i \leq n, 1 \leq j \leq n.$$

Beweis: Es sei $A = \bar{L}\bar{R}$ die LR-Zerlegung von A bei der \bar{L} lauter Einsen auf der Diagonalen hat. Damit gilt für die Koeffizienten aus Satz 4.2.8 wegen $A = A^T = \bar{R}^T\bar{L}^T$ auch

$$\bar{L}^T = (v_i(x_j))_{ij}, \quad \bar{R}^T = (\beta_{ij})_{ij}.$$

Mit $D = \text{diag}(\beta_{11}, \dots, \beta_{nn})$ gilt (siehe [23])

$$\begin{aligned} \bar{R}^T &= D\bar{L}^T, \text{ also } \beta_{ij} = \beta_{jj}v_j(x_i) \quad \text{und} \\ L &= \bar{L}D^{1/2}, \text{ also } L_{ij} = \bar{L}_{ij}\sqrt{\beta_{jj}} = v_j(x_i)\sqrt{\beta_{jj}}. \end{aligned}$$

Aus $v_j(x_j) = 1$, $1 \leq j \leq n$ folgt $L_{jj} = \sqrt{\beta_{jj}}$ und damit die Behauptung. \square

Korollar 4.2.10. Mit den Bezeichnungen aus Korollar 4.2.9 lassen sich die v_j , β_{ij} beschränken durch

$$|v_j(x_i)| \leq \frac{\sqrt{K(x_i, x_i)}}{L_{jj}}, \quad |\beta_{ij}| \leq \sqrt{K(x_j, x_j)K(x_i, x_i)}, \quad 1 \leq j \leq i \leq n.$$

Beweis: Weil K positiv definit ist, gilt

$$0 < K(x_j, x_j) = \sum_{k=1}^j L_{jk}^2.$$

Damit erhält man $|L_{jk}| \leq \sqrt{K(x_j, x_j)}$, für $k \leq j$. Durch Einsetzen in Korollar 4.2.9 folgt die Behauptung. \square

Zur Berechnung der v_j wurde bei den bisherigen Darstellungen immer rekursiv auf die Elemente der Kernbasis zugegriffen. Deshalb soll jetzt ausgerechnet werden, wie die v_j durch die Kernbasis darstellbar sind.

Satz 4.2.11. Die Koeffizienten a_ℓ^k , $\ell \leq k$, der Darstellung

$$v_k(x) = \sum_{\ell=1}^k a_\ell^k K(x, x_\ell)$$

lassen sich berechnen durch

$$\begin{aligned} a_j^k &= - \sum_{\ell=j+1}^k a_\ell^k v_j(x_\ell), \quad j < k, \\ a_k^k K(x_k, x_k) &= 1 - \sum_{\ell=1}^{k-1} a_\ell^k K(x_k, x_\ell). \end{aligned}$$

4.3: Eigenschaften der Newton-Basis

Beweis: Sei $j < k$, dann gilt

$$\begin{aligned} 0 = (v_k, v_j)_{\mathcal{N}_K} &= \sum_{\ell=1}^k a_\ell^k \underbrace{\sum_{r=1}^j a_r^j K(x_\ell, x_r)}_{v_j(x_\ell)} \\ &= a_j^k + \sum_{j+1}^k a_\ell^k v_j(x_\ell) \quad (\text{wegen Def. 4.2.1}). \end{aligned}$$

Die letzte Gleichung zur Berechnung von a_k^k folgt aus $1 = v_k(x_k) = \sum_{\ell=1}^k a_\ell^k K(x_k, x_\ell)$. \square

4.3 Eigenschaften der Newton-Basis

Die Besonderheit der in 4.2.1 definierten Basis gegenüber der herkömmlichen Newton-Basis auf Polynomräumen besteht darin, dass sich die bekannten Aussagen der Native Space Theorie darauf anwenden lassen. Damit lassen sich verschiedene nützliche Eigenschaften der Newton-Basis in kernbasierten Räumen zeigen, die wir nun vorstellen möchten.

Satz 4.3.1. *Die Elemente der Newton-Basis aus Definition 4.2.1 stehen senkrecht bezüglich $(\cdot, \cdot)_{\mathcal{N}_K}$ aufeinander,*

$$(v_j, v_k)_{\mathcal{N}_K} = 0, \quad 1 \leq k < j \leq n.$$

Außerdem gilt $(v_j, v_j)_{\mathcal{N}_K} = 1/\beta_{jj}$, mit β_{jj} aus Satz 4.2.8.

Beweis: Nach Definition gilt $v_j(x_i) = 0$, $1 \leq i \leq j$ und $v_k \in U_{X_k, K}$. Damit folgt die erste Aussage direkt aus Satz 2.2.7.

Aus der Darstellung von v_j aus 4.2.8, zusammen mit der Reproduktionseigenschaft (Satz 2.2.3) und $(v_j, v_k) = 0$, $k < j$, folgt

$$\begin{aligned} \beta_{jj}(v_j, v_j)_{\mathcal{N}_K} &= (v_j, K(\cdot, x_j) - \sum_{k=0}^{j-1} \beta_{jk} v_k)_{\mathcal{N}_K} \\ &= (v_j, K(\cdot, x_j))_{\mathcal{N}_K} \\ &= v_j(x_j) \\ &= 1. \quad \square \end{aligned}$$

Damit ist $\left\{ \frac{v_j}{\|v_j\|_{\mathcal{N}_K}} \right\}_{j=1}^n$ eine Orthonormalbasis von $U_{X_n, K}$ und $\lambda_j(f) \|v_j\|_{\mathcal{N}_K}$ ist der j -te Koeffizient von f bezüglich dieser Basis. Dies nutzen wir aus, um eine weitere Darstellung der Koeffizientenfunktionale $\lambda_j(f)$ zu finden.

Satz 4.3.2. Die Koeffizienten $\lambda_j(f)$, $1 \leq j \leq n$, lassen sich darstellen durch

$$\begin{aligned}\lambda_j(f) &= \left(s_{f, X_n}, \frac{v_j}{\|v_j\|_{\mathcal{N}_K}^2} \right)_{\mathcal{N}_K} \\ &= \left(f, \frac{v_j}{\|v_j\|_{\mathcal{N}_K}^2} \right)_{\mathcal{N}_K}.\end{aligned}$$

Beweis: Nach Satz 4.2.5 gilt $s_{f, X_n}(x) = \sum_{k=1}^n \lambda_k(f) v_k(x)$. Aus der Orthogonalität der v_k folgt die erste Gleichung wegen $\lambda_j(f) \|v_j\|_{\mathcal{N}_K} = \left(s_{f, X_n}, \frac{v_j}{\|v_j\|_{\mathcal{N}_K}^2} \right)_{\mathcal{N}_K}$. Außerdem gilt

$$\left(s_{f, X_n} - f, \frac{v_j}{\|v_j\|_{\mathcal{N}_K}^2} \right)_{\mathcal{N}_K} = 0,$$

da $(s_{f, X_n} - f)(x_i) = 0$, für $x_i \in X_n$ und $v_j \in U_{X_j, K} \subset U_{X_n, K}$ (vergleiche Satz 2.2.7). Daraus folgt die zweite Gleichheit. \square

Bemerkung 4.3.3. Satz 4.3.2 liefert nicht nur eine Darstellung des Koeffizientenfunktionalen. Gleichzeitig wird nämlich bewiesen, dass $v_j / \|v_j\|_{\mathcal{N}_K}^2$ gerade der **Riesz Darsteller** von $\lambda_j(\cdot) \in \mathcal{N}_K^*$ ist.

Korollar 4.3.4. Die Koeffizienten $\lambda_j(f)$, $1 \leq j \leq n$, sind beschränkt durch

$$|\lambda_j(f)| \leq \|f\|_{\mathcal{N}_K} \sqrt{K(x_j, x_j)}.$$

Beweis: Aus Satz 4.3.2 und der Cauchy-Schwarzschen Ungleichung erhalten wir

$$|\lambda_j(f)| = \left| \left(f, \frac{v_j}{\|v_j\|_{\mathcal{N}_K}^2} \right)_{\mathcal{N}_K} \right| \leq \frac{\|f\|_{\mathcal{N}_K}}{\|v_j\|_{\mathcal{N}_K}}.$$

Mit $\|v_j\|_{\mathcal{N}_K}^2 = 1/\beta_{jj}$ (siehe Satz 4.3.1) und Korollar 4.2.10 folgt daraus

$$|\lambda_j(f)| \leq \|f\|_{\mathcal{N}_K} \sqrt{\beta_{jj}} \leq \|f\|_{\mathcal{N}_K} \sqrt{K(x_j, x_j)}. \quad \square$$

Satz 4.3.5. Sei $x \in \Omega$, $g_x := K(\cdot, x)$. Dann gilt für die Koeffizienten der Newton-Basis

$$\lambda_j(g_x) = \frac{v_j(x)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2}.$$

Beweis: Nach Satz 4.3.2 folgt unter Verwendung der Reproduktionseigenschaft des Kerns K (siehe Satz 2.2.3)

$$\lambda_j(g_x) = \left(K(\cdot, x), \frac{v_j(\cdot)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2} \right)_{\mathcal{N}_K(\Omega)} = \frac{v_j(x)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2}. \quad \square$$

4.3: Eigenschaften der Newton-Basis

Satz 4.3.6. Für die Native Space Norm der Interpolante gilt:

$$\|s_{f,X_n}\|_{\mathcal{N}_K}^2 = \sum_{j=1}^n \lambda_j^2(f) \|v_j\|_{\mathcal{N}_K}^2.$$

Beweis: Wegen $s_{f,X_n} \in U_{X_n,K}$ und weil $\left\{ \frac{v_j}{\|v_j\|_{\mathcal{N}_K}} \right\}_{j=1}^n$ eine Orthonormalbasis von $U_{X_n,K}$ ist, gilt nach der Parsevalschen Gleichung

$$\begin{aligned} \|s_{f,X_n}\|_{\mathcal{N}_K}^2 &= \sum_{j=1}^n \left(s_{f,X_n}, \frac{v_j}{\|v_j\|_{\mathcal{N}_K}} \right)_{\mathcal{N}_K}^2 \\ &= \sum_{j=1}^n \lambda_j^2(f) \|v_j\|_{\mathcal{N}_K}^2, \quad (\text{nach Satz 4.3.2}). \quad \square \end{aligned}$$

Bemerkung 4.3.7. Aus Satz 4.3.6 und der Normminimalität der Interpolante folgt

$$\sum_{j=1}^n \lambda_j^2(f) \|v_j\|_{\mathcal{N}_K}^2 = \|s_{f,X_n}\|_{\mathcal{N}_K}^2 \leq \|f\|_{\mathcal{N}_K}^2, \quad \forall n \in \mathbb{N}.$$

Deshalb ist sichergestellt, dass im Falle von unendlich vielen Punkten die Reihe aus Satz 4.3.6 beschränkt bleibt und man kann den Grenzwert für $n \rightarrow \infty$ betrachten.

Nun nutzen wir Satz 4.3.6, um für die Newton-Basis eine Stabilitätsaussage wie in Satz 4.1.1 herzuleiten.

Satz 4.3.8. Für die Darstellung der Interpolante aus 4.2.5 mit Hilfe der Newton-Basis gilt die Abschätzung

$$\sum_{j=1}^n |\lambda_j(f)| |v_j(x)| \leq c\sqrt{n} \|f\|_{\mathcal{N}_K}, \quad \forall x \in \Omega,$$

unter der Voraussetzung, dass sich der Native Space $\mathcal{N}_K(\Omega)$ durch $\|g\|_{L_\infty(\Omega)} \leq c \|g\|_{\mathcal{N}_K}$, für $g \in \mathcal{N}_K(\Omega)$, in den Raum der stetigen Funktionen $C(\Omega)$ einbetten lässt.

Beweis:

$$\begin{aligned} \sum_{j=1}^n |\lambda_j(f)| \|v_j\|_{L_\infty(\Omega)} &\leq c \sum_{j=1}^n |\lambda_j(f)| \|v_j\|_{\mathcal{N}_K} \\ &\leq c \sqrt{\sum_{j=1}^n 1^2} \underbrace{\sqrt{\sum_{j=1}^n \lambda_j^2(f) \|v_j\|_{\mathcal{N}_K}^2}}_{\leq \|f\|_{\mathcal{N}_K}}. \end{aligned}$$

Dabei wird im zweiten Schritt die Cauchy Schwarzsche Ungleichung benutzt. Die Abschätzung durch $\|f\|_{\mathcal{N}_K}$ gilt nach Bemerkung 4.3.7. \square

Bemerkung 4.3.9. Die Voraussetzung $\|g\|_{L_\infty(\Omega)} \leq c\|g\|_{\mathcal{N}_K}$ ist zum Beispiel erfüllt, wenn Ω kompakt und K stetig (und damit auch beschränkt) ist, denn es gilt für $g \in \mathcal{N}_K(\Omega)$ nach der Reproduktionseigenschaft von K

$$\begin{aligned} |g(x)| &= \left| (g, K(\cdot, x))_{\mathcal{N}_K} \right| \\ &\leq \|g\|_{\mathcal{N}_K} \|K(\cdot, x)\|_{\mathcal{N}_K} \\ &= \|g\|_{\mathcal{N}_K} \sqrt{K(x, x)}, \quad \forall x \in \Omega. \end{aligned}$$

Im Vergleich zur Abschätzung bezüglich der Lagrange-Basis in Satz 4.1.1 ist zu beachten, dass in Satz 4.3.8 auf der rechten Seite die Native Space Norm anstatt der L_∞ Norm benutzt wurde.

4.4 Reihenentwicklungen

Aus Satz 4.3.1 folgt noch mehr, als wir bisher benutzt haben. Falls der Interpolationsfehler für wachsendes n punktweise gegen Null konvergiert, erhalten wir nämlich für alle $f \in \mathcal{N}_K(\Omega)$ durch die Newton-Basis eine Darstellung in einer orthogonalen Reihe.

Satz 4.4.1. Unter der Voraussetzung $|(f - s_{f, X_n})(x)| \xrightarrow{n \rightarrow \infty} 0$, für $x \in \Omega$ folgt

$$f(x) = \sum_{j=1}^{\infty} \lambda_j(f) v_j(x), \quad x \in \Omega.$$

Beweis: Nach Satz 4.2.5 gilt $s_{f, X_n}(x) = \sum_{j=1}^n \lambda_j(f) v_j(x)$ und damit nach Voraussetzung

$$0 = \lim_{n \rightarrow \infty} |(f - s_{f, X_n})(x)| = |(f - \sum_{j=1}^{\infty} \lambda_j(f) v_j)(x)|. \quad \square$$

Korollar 4.4.2. Sei $f \in \mathcal{N}_K(\Omega)$ und die Interpolationpunkte seien mit f -Greedy (siehe Definition 3.1.1 (**f -Greedy**)) ausgewählt. Dann gilt

$$f(x) = \sum_{j=1}^{\infty} \lambda_j(f) v_j(x), \quad x \in \Omega.$$

Beweis: Nach Satz 3.2.2 gilt

$$\sum_{j=1}^{\infty} \varepsilon_j^2 < \infty. \quad (4.6)$$

Damit konvergiert der Interpolationsfehler bei f -Greedy (sogar gleichmäßig) gegen Null und aus Satz 4.4.1 folgt, dass f sich wie behauptet in einer Reihe durch die Elemente der Newton-Basis darstellen lässt. \square

4.5: Reihenentwicklung des Kerns mit der Powerfunktion

Die Reihendarstellung aus Korollar 4.4.2 erhält man auch, wenn man die Punkte so wählt, dass die Füllichte gegen Null konvergiert. Denn aus Satz 4.4.1 und Korollar 2.2.21 folgt sofort

Korollar 4.4.3. *Sei $f \in \mathcal{N}_K(\Omega) \subset C(\Omega)$. Es gelte $h(\Omega, X_n) \xrightarrow{n \rightarrow \infty} 0$. Dann lässt sich f in einer **orthogonalen Reihe** darstellen durch*

$$f(x) = \sum_{j=1}^{\infty} \lambda_j(f) v_j(x), \quad x \in \Omega.$$

Bemerkung 4.4.4. Im Vergleich der beiden Korollare 4.4.2 und 4.4.3 lässt sich Folgendes feststellen:

Werden die Punkte durch f -Greedy verteilt, muss Ω nicht unbedingt gefüllt werden, aber dennoch gilt die Reihenentwicklung für diese spezielle Funktion. Liegen dagegen die Interpolationspunkte dicht in Ω , gilt die Reihendarstellung für alle $f \in \mathcal{N}_K(\Omega)$.

Korollar 4.4.5. *Sei $X_n \subset \Omega \subset \mathbb{R}^d$, $n \in \mathbb{N}$, mit $h(\Omega, X_n) \xrightarrow{n \rightarrow \infty} 0$. Dann lässt sich der Interpolationskern K als Reihe darstellen durch*

$$K(x, y) = \sum_{j=1}^{\infty} \frac{v_j(x)v_j(y)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2}.$$

Beweis: Weil die $v_j/\|v_j\|_{\mathcal{N}_K(\Omega)}$, $j = 1, 2, \dots$ ein Orthonormalsystem des Native Space bilden, gilt die Aussage nach [5]. Mit den Methoden dieser Arbeit folgt auch ein direkter Beweis. Sei $x \in \Omega$, $g_x := K(\cdot, x)$. Nach Korollar 4.4.3 gilt

$$g_x(y) = \sum_{j=1}^{\infty} \lambda_j(g_x) v_j(y),$$

und nach Satz 4.3.5 folgt die Behauptung. □

Bemerkung 4.4.6. *Nach Korollar 4.4.5 gilt insbesondere*

$$\sum_{j \in \mathbb{N}} \frac{v_j^2(x)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2} < \infty.$$

4.5 Reihenentwicklung des Kerns mit der Powerfunktion

Satz 4.5.1. *Zwischen dem Funktional $\lambda_j(\cdot)$ und der Powerfunktion (siehe Definition 2.2.11) besteht folgende Beziehung,*

$$\|\lambda_j(\cdot)\|_{\mathcal{N}_K^*} = P_{X_{j-1}}(x_j).$$

4.5: Reihenentwicklung des Kerns mit der Powerfunktion

Beweis: Für das Funktional $\lambda_j(\cdot)$ gilt nach Korollar 4.2.6

$$\lambda_j(f) = f(x_j) - s_{f, X_{j-1}}(x_j).$$

Auf der anderen Seite ist der Wert der Powerfunktion an einer Stelle $x \in \Omega$ nach Satz 2.2.14 gerade die Native Space Norm des zugehörigen Fehlerfunktionals. Damit stimmen die beiden Ausdrücke überein. \square

Für die folgenden Überlegungen benötigen wir die Eigenschaften des Powerkerns K_n aus Definition 2.2.18.

Satz 4.5.2. *Für die Powerfunktion (siehe 2.2.11) und die Newton-Basis $\{v_j\}_{j=1}^{n+1}$ zu gegebenem positiv definiten, symmetrischen Kern K gilt folgende Rekursionsformel:*

$$P_{n+1}^2(x) = P_n^2(x) - v_{n+1}^2(x)P_n^2(x_{n+1}), \quad n \in \mathbb{N}_0. \quad (4.7)$$

Beweis:

$n = 0$: Nach Definition 2.2.11 gilt (beachte $u_n^{X_n} = v_n$)

$$P_1^2(x) = K(x, x) - 2v_1(x)K(x, x_1) + v_1^2(x)K(x_1, x_1).$$

Mit $v_1(x) = K(x, x_1)/K(x_1, x_1)$ folgt daraus

$$P_1^2(x) = K(x, x) - v_1^2(x)P_0^2(x_1).$$

$n > 1$: Nach Satz 2.2.19 b) gilt

$$K_{n+1}(x, x) = K_n(x, x) - \frac{K_n^2(x, x_{n+1})}{K_n(x_{n+1}, x_{n+1})}.$$

Daraus folgt unter Verwendung von 2.2.19 e) und $u_n^{X_n} = v_n$

$$\begin{aligned} P_{n+1}^2(x) &= P_n^2(x) - \frac{v_{n+1}^2(x) \cdot K_n^2(x_{n+1}, x_{n+1})}{P_n^2(x_{n+1})} \\ &= P_n^2(x) - v_{n+1}^2(x)P_n^2(x_{n+1}). \quad \square \end{aligned}$$

Satz 4.5.2 ist eine verschärfte quantitative Form von Korollar 2.2.15.

Satz 4.5.3. *Für die Powerfunktion P_n zu den Stützstellen x_1, \dots, x_n (vgl. 2.2.11) und die Newton-Basis $\{v_j\}_{j=1}^n$ zu gegebenem positiv definitem, symmetrischem Kern K gilt die Rekursionsformel*

$$P_n^2(x) = K(x, x) - \sum_{j=1}^n v_j^2(x)P_{j-1}^2(x_j), \quad n \in \mathbb{N}, x \in \Omega.$$

4.5: Reihenentwicklung des Kerns mit der Powerfunktion

Beweis: Durch wiederholtes Anwenden von Satz 4.5.2 folgt

$$\begin{aligned}
 P_n^2(x) &= P_{n-1}^2(x) - v_n^2(x)P_{n-1}^2(x_n) \\
 &= P_{n-2}^2(x) - \sum_{j=n-1}^n v_j^2(x)P_{j-1}^2(x_j) = \dots \\
 &= P_0^2(x) - \sum_{j=1}^n v_j^2(x)P_{j-1}^2(x_j).
 \end{aligned}$$

Korollar 4.5.4. Sei $X_n \subset \Omega \subset \mathbb{R}^d$, $n \in \mathbb{N}$, mit $h(\Omega, X_n) \xrightarrow{n \rightarrow \infty} 0$. Dann lässt sich die Kernfunktion K „auf der Diagonalen“ folgendermaßen als Reihe darstellen:

$$K(x, x) = \sum_{j=1}^{\infty} v_j^2(x)P_{j-1}^2(x_j), \quad x \in \Omega. \quad (4.8)$$

Beweis: Nach Satz 4.5.3 gilt

$$K(x, x) = P_n^2(x) + \sum_{j=1}^n v_j^2(x)P_{j-1}^2(x_j), \quad n \in \mathbb{N}, x \in \Omega.$$

Wegen $h(\Omega, X_n) \xrightarrow{n \rightarrow \infty} 0$ gilt nach Satz 2.2.20 aber auch

$$\begin{aligned}
 0 &= \lim_{n \rightarrow \infty} P_n^2(x) \text{ und damit} \\
 K(x, x) &= \lim_{n \rightarrow \infty} \left(\sum_{j=1}^n v_j^2(x)P_{j-1}^2(x_j) \right). \quad \square
 \end{aligned}$$

Kapitel 5

Die Newton-Basis mit Greedy-Punktwahl

In diesem Kapitel werden die beiden vorherigen Kapitel zusammengeführt, da wir Newton-Basen betrachten, deren Punktfolgen durch Greedy-Verfahren bestimmt sind.

Die Elemente der Newton-Basis hängen stark von der verwendeten Punktwahl ab. In diesem Kapitel werden wir zeigen, dass die Elemente der Newton-Basis bei Punktwahl durch P -Greedy beschränkt bleiben. Außerdem benutzen wir Eigenschaften der Newton-Basis, um weitere Aussagen für die Greedy-Verfahren aus Definition 3.1.1 zu gewinnen. Dazu untersuchen wir zunächst unabhängig von der Punktwahl, wie die Powerfunktion mit der Newton-Basis zusammenhängt.

5.1 Beziehungen zwischen der Powerfunktion und der Newton-Basis

Wegen der Orthogonalität der Basiselemente v_j , $j \in \mathbb{N}$ der Newton-Basis sind die zugehörigen Koeffizienten in den Reihendarstellungen eindeutig. Setzt man nun in Korollar 4.4.5 $y := x$ und vergleicht die Koeffizienten mit denen aus Korollar 4.5.4, so folgt direkt

Korollar 5.1.1. *Für die Elemente der Newton-Basis und die Powerfunktion gilt*

$$P_{j-1}(x_j) = \frac{1}{\|v_j\|_{\mathcal{N}_K}}.$$

Mit Satz 4.5.1 gilt somit auch

$$\|\lambda_j(\cdot)\|_{\mathcal{N}_K^*} = \frac{1}{\|v_j\|_{\mathcal{N}_K}}.$$

Im Gegensatz zu Korollar 4.5.4 muss hierbei die Menge der Interpolationspunkte nicht dicht liegen, da sich die Basisfunktionen durch das Hinzufügen weiterer Punkte nicht

5.2: Eigenschaften der Newton-Basis bei P -Greedy

verändern, die Punktmenge also immer zu einer dichten Teilmenge ergänzt werden könnte.

Mit Hilfe des vorhergehenden Korollars lässt sich nun eine weitere nützliche Beziehung zwischen der Powerfunktion und den Basisfunktionen herstellen, die wir in Korollar 5.2.1 benutzen werden, um für P -Greedy die Beschränktheit der Newton-Basis zu zeigen.

Satz 5.1.2. Für $f \in \Omega$, $x_1, \dots, x_j \in \Omega$, $j \in \mathbb{N}$ mit zugehöriger Powerfunktion P_j gilt:

$$\frac{v_j^2(x)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2} = P_{j-1}^2(x) - P_j^2(x) \leq P_{j-1}^2(x).$$

Beweis: Nach Satz 4.5.2 gilt

$$P_{j-1}^2(x) - P_j^2(x) = v_j^2(x)P_{j-1}^2(x_j) \stackrel{5.1.1}{=} \frac{v_j^2(x)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2}.$$

Da nach Korollar 2.2.15 $P_{j-1} \geq P_j$ gilt, folgt auch der zweite Teil der Behauptung. \square

5.2 Eigenschaften der Newton-Basis bei P -Greedy

Als eine Variante der Greedy-Verfahren wurde in Definition 3.1.1 P -Greedy vorgestellt. Diese Art der Punktwahl fügt in jedem Schritt einen Interpolationspunkt dort ein, wo die Powerfunktion (siehe 2.2.11) ein Maximum hat (siehe auch [3]). Für diese Methode zeigen wir nun, dass die Elemente der Newton-Basis global im Absolutbetrag durch 1 beschränkt werden können.

Korollar 5.2.1. Werden die Interpolationspunkte mit Hilfe von P -Greedy ausgewählt, so gilt

$$|v_j(x)| \leq 1, \quad x \in \Omega, \quad j \in \mathbb{N}.$$

Beweis: Aus vorherigen Überlegungen wissen wir

$$1 = v_j(x_j), \quad (\text{Def. 4.2.1})$$

$$\frac{v_j^2(x)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2} \leq P_{j-1}^2(x), \quad x \in \Omega, \quad (\text{Satz 5.1.2})$$

$$\frac{v_j^2(x_j)}{\|v_j\|_{\mathcal{N}_K(\Omega)}^2} = P_{j-1}^2(x_j) \geq P_{j-1}^2(x), \quad x \in \Omega. \quad (\text{Satz 2.2.16, Def. 3.1.1})$$

Daraus ergibt sich $v_j^2(x_j) \geq v_j^2(x)$ und damit

$$1 = v_j(x_j) \geq |v_j(x)|, \quad x \in \Omega. \quad \square$$

5.3: Beziehungen zwischen der Determinante und der Powerfunktion

Bemerkung 5.2.2. In Korollar 5.2.1 wurde gezeigt, dass v_j an der Stelle x_j ein globales Maximum besitzt. Falls x_j im Inneren von Ω liegt, müssen daher auch alle Richtungsableitungen von v_j bei x_j verschwinden.

Das ist sehr ungewöhnlich, da die Lagrangefunktionen bei allgemeiner Interpolation nicht ihr Maximum in dem jeweiligen ausgezeichneten Datenpunkt annehmen müssen. Außerdem ist zum Beispiel bei Polynomen zu beobachten, dass die Lagrangefunktionen viel stärker oszillieren und nicht beschränkt bleiben, wenn die Punkte dicht zusammenliegen. Bei der Punktwahl durch P -Greedy verhalten sich die Elemente der Newton-Basis dagegen sehr stabil, wie auch in den Plots in Kapitel 7 zu beobachten ist.

5.3 Beziehungen zwischen der Determinante und der Powerfunktion

Im folgenden Satz zeigen wir, wie sich die Determinante der Kernmatrix als Produkt von Funktionswerten der Powerfunktion schreiben lässt. Diese Aussage gilt unabhängig von der verwendeten Basis, aber für den Beweis können wir verschiedene Eigenschaften benutzen, die wir für die Newton-Basis gezeigt haben. Für die Punktwahl durch P -Greedy werden wir die Aussage noch weiter umformulieren.

Satz 5.3.1. Sei A_{X_n} die Interpolationsmatrix zu den Stützstellen $X_n := \{x_1, \dots, x_n\}$ aus Gleichung (2.3). Dann gilt

$$\det(A_{X_n}) = \prod_{j=1}^n P_{j-1}^2(x_j).$$

Beweis: Für die Zerlegung aus 4.2.8 gilt nach Satz 4.3.1

$$(v_j, v_j)_{\mathcal{N}_K} = 1/\beta_{jj}.$$

Nach Korollar 4.2.9 folgt daraus für die Diagonalelemente L_{jj} der Cholesky-Zerlegung

$$\|v_j\|_{\mathcal{N}_K}^2 = (v_j, v_j)_{\mathcal{N}_K} = 1/L_{jj}^2.$$

Wegen der positiven Definitheit von K gilt außerdem $L_{jj} > 0$ und aus Korollar 5.1.1 erhält man damit

$$P_{j-1}(x_j) = L_{jj}.$$

Für die Diagonaleinträge der Cholesky-Zerlegung gilt aber gerade

$$\det(A_{X_n}) = \prod_{j=1}^n L_{jj}^2. \quad \square$$

5.4: Fehlerabsch. für Greedy-Verfahren mit Hilfe der Newton-Basis

Korollar 5.3.2. *Bei der Punktwahl durch P -Greedy gilt*

$$\det(A_{X_n}) = \prod_{j=1}^n \|P_{j-1}^2\|_{L_\infty(\Omega)}.$$

Bemerkung 5.3.3. Ähnlich wie in Gleichung (4.1) kann man hier folgern, dass nicht gleichzeitig die Determinante der Kernmatrix groß (gute Stabilität) und die Powerfunktion klein (kleiner Fehler) sein kann.

Außerdem ist bemerkenswert, dass die Determinante nicht von der Reihenfolge der Punkte abhängt, was damit auch für das Produkt der Werte der aufeinander folgenden Powerfunktionen gilt.

Bemerkung 5.3.4. Aus den beiden Korollaren folgt, dass bei P -Greedy in jedem Schritt der nächste Punkt so gewählt wird, dass die Determinante der Interpolationsmatrix maximal wird. Außerdem lassen numerische Tests vermuten, dass bei spaltenpivotisierter LR-Zerlegung der durch P -Greedy gewonnenen Kernmatrix keine Zeilenvertauschungen nötig sind. Da auch die Elemente der Newton-Basis sich aus der LR-Zerlegung ergeben, kann P -Greedy deshalb benutzt werden, um zu einer gegebenen Menge X an Interpolationspunkten eine gute Reihenfolge für die Berechnung der Newton-Basis festzulegen, indem P -Greedy auf X durchgeführt wird. Das kann man als eine Pivotisierungsmethode für eine LR-Zerlegung ansehen.

5.4 Fehlerabschätzungen für Greedy-Verfahren mit Hilfe der Newton-Basis

Im Vergleich zu Satz 3.2.2 leiten wir nun etwas stärkere Aussagen für das Verhalten des Interpolationsfehlers bei Verwendung von f -Greedy und f/P -Greedy her.

Satz 5.4.1. *Sei $f \in \mathcal{N}_K(\Omega)$, Ω beschränkt und die Datenpunkte seien mit Hilfe von f -Greedy gewählt. $\varepsilon_n := \|f - s_{f, X_n}\|_{L_\infty(\Omega)}$ sei der Interpolationsfehler, der an der Stelle x_{ε_n} angenommen wird. Dann gilt*

$$\sum_{j=1}^{\infty} \frac{\varepsilon_j^2}{P_{j-1}^2(x_{\varepsilon_j})} = \|f\|_{\mathcal{N}_K}^2.$$

Beweis: Nach Satz 4.3.1 und Korollar 4.4.2 ist $\{\frac{v_j}{\|v_j\|_{\mathcal{N}_K}}\}_{j=1}^{\infty}$ eine Orthonormalbasis von \mathcal{N}_K . Deshalb gilt nach der Parsevalschen Gleichung

$$\|f\|_{\mathcal{N}_K}^2 = \sum_{j=1}^{\infty} \left(f, \frac{v_j}{\|v_j\|_{\mathcal{N}_K}}\right)_{\mathcal{N}_K}^2 \stackrel{4.3.2}{=} \sum_{j=1}^{\infty} \lambda_j^2(f) \|v_j\|_{\mathcal{N}_K}^2.$$

5.4: Fehlerabsch. für Greedy-Verfahren mit Hilfe der Newton-Basis

Damit folgt aus den Korollaren 5.1.1 und 4.2.6

$$\begin{aligned} \sum_{j=1}^n \lambda_j^2(f) \|v_j\|_{\mathcal{N}_K}^2 &= \sum_{j=1}^n \frac{\lambda_j^2(f)}{P_{j-1}^2(x_j)} \\ &= \sum_{j=1}^n \frac{(f - s_{f, X_{j-1}})^2(x_j)}{P_{j-1}^2(x_j)}. \end{aligned}$$

Nach der Definition von f -Greedy gilt aber gerade $x_j = x_{\varepsilon_j}$. □

Bemerkung 5.4.2. Satz 5.4.1 zeigt, dass der Fehler bei f -Greedy genau dann schnell fällt, wenn die Punkte dort ausgewählt werden, wo die Powerfunktion klein ist. Das bedeutet aber, dass die Punkte gerade nicht gleichmäßig verteilt werden.

Interessant ist das auch deshalb, weil die vom Füllabstand abhängigen Abschätzungen ebenfalls durch das Abklingverhalten der Powerfunktion gezeigt werden können. Da die Aussage des Satzes mit Gleichheit gilt, ist bei f -Greedy auch das Abklingverhalten der Powerfunktion für die Größe des Interpolationsfehlers von großem Einfluss.

Satz 5.4.3. Sei $f \in \mathcal{N}_K(\Omega)$, Ω und die Datenpunkte seien mit Hilfe von f/P -Greedy gewählt. $\varepsilon_n := \|f - s_{f, X_n}\|_{L^\infty(\Omega)}$ sei der Interpolationsfehler, der an der Stelle x_{ε_n} angenommen wird. Dann gilt

$$\sum_{j=1}^{\infty} \frac{\varepsilon_j^2}{P_{j-1}^2(x_{\varepsilon_j})} \leq \|f\|_{\mathcal{N}_K}^2.$$

Beweis: Nach Definition von f/P -Greedy gilt

$$(f - s_{f, X_{j-1}}(x_j))^2 \geq \frac{(f - s_{f, X_{j-1}}(x_{\varepsilon_j}))^2}{P_{j-1}^2(x_{\varepsilon_j})} P_{j-1}^2(x_j).$$

Damit folgt wie in Satz 5.4.1

$$\begin{aligned} \sum_{j=1}^n \lambda_j^2(f) \|v_j\|_{\mathcal{N}_K}^2 &= \sum_{j=1}^n \frac{(f - s_{f, X_{j-1}})^2(x_j)}{P_{j-1}^2(x_j)} \\ &\geq \sum_{j=1}^{\infty} \frac{\varepsilon_j^2}{P_{j-1}^2(x_{\varepsilon_j})}. \quad \square \end{aligned}$$

Bemerkung 5.4.4. Bei f -Greedy und f/P -Greedy ist leider nicht bekannt, ob und gegebenenfalls wie schnell $P_{j-1}(x_j)$ mit wachsendem j gegen Null strebt. Daher bleibt an dieser Stelle unklar, ob die obigen Resultate schnellere Konvergenz als $\sum_{j=1}^{\infty} \varepsilon_j^2 < \infty$ liefern. Allerdings ist bekannt, dass P_j bei P -Greedy beziehungsweise bei Raumfüllung gleichmäßig gegen Null konvergiert und es ist zu vermuten, dass die Werte der Powerfunktion in Bereichen, in denen die Interpolationspunkte eng zusammen liegen, sogar noch kleiner sind.

Kapitel 6

Tests verschiedener Greedy-Verfahren

Die numerischen Tests gliedern sich, wie auch die Arbeit selbst, in zwei große Bereiche. In diesem Kapitel geht es um dünne Approximationen, dabei wird vor allem das Greedy-Verfahren aus Kapitel 3 untersucht. Das folgende Kapitel behandelt dagegen die Newton-Basis (siehe Kapitel 4) unter der Fragestellung der stabilen Approximation.

Bei der Untersuchung der Greedy-Methoden geht es um die Fragestellung, wie viele Punkte benötigt werden, um eine vorgegebene Approximationsgüte zu erreichen. Natürlich lässt sich darauf pauschal keine Antwort geben, weil das Approximationsverhalten von zahlreichen, völlig verschiedenen Parametern abhängt: Profil der Datenfunktion, verwendeter Kern, das Gebiet Ω , sowie die Verteilung und Anzahl der Punkte.

Ziel ist es, den Einfluss der genannten Parameter auf die Approximation sichtbar zu machen. Dabei wird deutlich, dass zum Teil große Unterschiede zwischen den asymptotischen Abschätzungen und dem tatsächlichen numerischen Verhalten bestehen. Während für die Asymptotik vor allem die Glattheitsklasse der Datenfunktion und des Kerns entscheidend ist, spielt in der Anwendung auch die Größe der Ableitungen eine wichtige Rolle. Außerdem werden in der Theorie der Numerik vor allem „worst-case“ Szenarien betrachtet. Diese liefern deutlich pessimistischere Konvergenzraten, als die Beobachtungen bei Verwendung der Greedy-Verfahren vermuten lassen würden. Gerade weil durch die unregelmäßige Punkteverteilung f -Greedy nur schwer theoretisch analysiert werden kann (siehe Kapitel 3), sind hier numerische Ergebnisse besonders interessant.

Als Datenfunktionen benutzen wir die Betragsfunktion $f(x, y) = |x|$ und Frankes Testfunktion, gegeben durch

$$\begin{aligned}
 f(x, y) &= f_1(x, y) + f_2(x, y) + f_3(x, y) - f_4(x, y), & \text{mit} & & (6.1) \\
 f_1 &= 0.75 \exp(-((9x - 2)^2 + (9y - 2)^2)/4), \\
 f_2 &= 0.75 \exp(-((9x + 1)^2/49 + (9y + 1)^2/10)), \\
 f_3 &= 0.5 \exp(-((9x - 7)^2 + (9y - 3)^2)/4), \\
 f_4 &= 0.2 \exp(-((9x - 4)^2 + (9y - 7)^2)),
 \end{aligned}$$

jeweils in $[-1, 1]^2$. Es werden 100 Punkte ausgewählt, und die Wendlandfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4(4\|x - y\|_2 + 1)$ wird als Kern verwendet. Die Punkte werden aus einem Gitter mit Gitterweite 0,02 ausgesucht.

Es werden drei Arten der Greedy-Approximation miteinander verglichen (siehe Definition 3.1.1):

- **f -Greedy**, das den nächsten Punkt dort auswählt, wo der Fehler am größten ist.
- **P -Greedy**, das f -unabhängig Punkte bestimmt, an denen die Powerfunktion maximal ist.
- **f/P -Greedy**, eine Mischung aus den beiden anderen Methoden, bei der die Native Space Norm der Interpolante maximiert wird.

6.1 Untersuchung der Punkteverteilungen

Im ersten Schritt vergleichen wir die Punktfolgen, die von den drei oben genannten Verfahren erzeugt werden.

Beobachtungen: Aus den Abbildungen 6.1 und 6.2 wird deutlich, dass f/P -Greedy die Punkte sehr stark clustert, was auch zu schlecht konditionierten Gleichungssystemen führt (siehe Abschnitt 6.5). Dies kommt daher, dass in Definition 3.1.1 (**f/P -Greedy**) die Powerfunktion im Nenner steht und diese in der Nähe der Datenpunkte sehr klein ist. Für $f(x, y) = |x|$ ist bei $x \approx 0$ wegen der schlechten Konvergenz $f - s_{f, X_k}$ groß im Vergleich zu P_{K, X_k} . Bei Frankes Testfunktion tritt dieses Problem weniger stark auf, weil dort die Konvergenz besser ist. In diesem Fall werden kreisförmig um die lokalen Extremstellen Punkte ausgesucht, ansonsten werden in weiten Teilen nur sehr wenige Punkte ausgewählt.

f -Greedy wählt bei der Betragsfunktion auch viele Punkte um $x = 0$, weil die Funktion dort einen Knick hat, allerdings viel breiter gestreut als bei f/P -Greedy. Auch im zweiten Test wurden von f -Greedy viele Punkte dort ausgewählt, wo die Funktion ihr Verhalten stark ändert. Dort, wo die Datenfunktion nahezu konstant ist, werden dagegen kaum Punkte verwendet.

P -Greedy verteilt dagegen die Punkte f -unabhängig und sehr gleichmäßig, anscheinend quasi-uniform.

6.1: Untersuchung der Punkteverteilungen

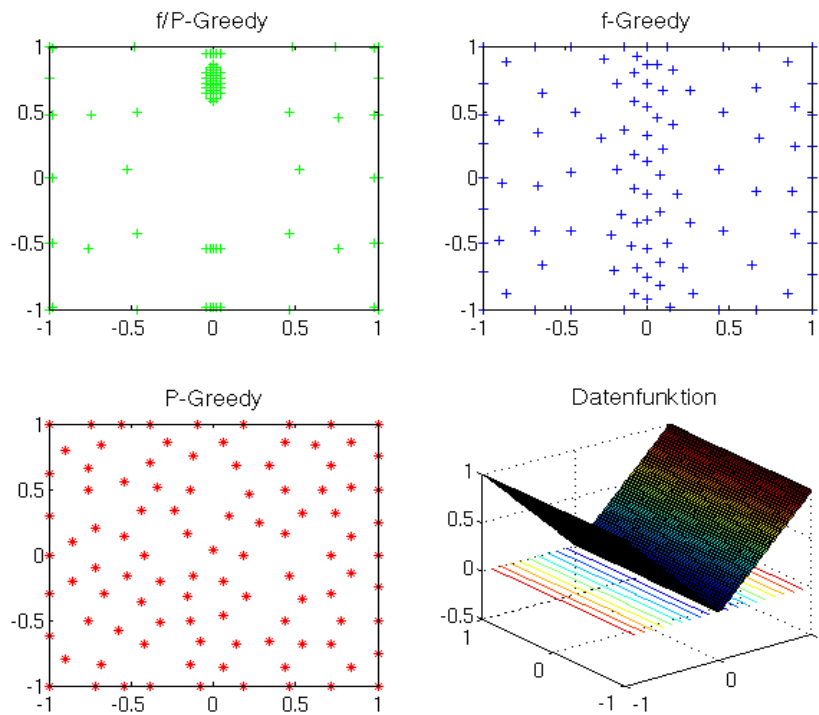


Abbildung 6.1: Punkteverteilung mit Betragsfunktion als Datenfunktion

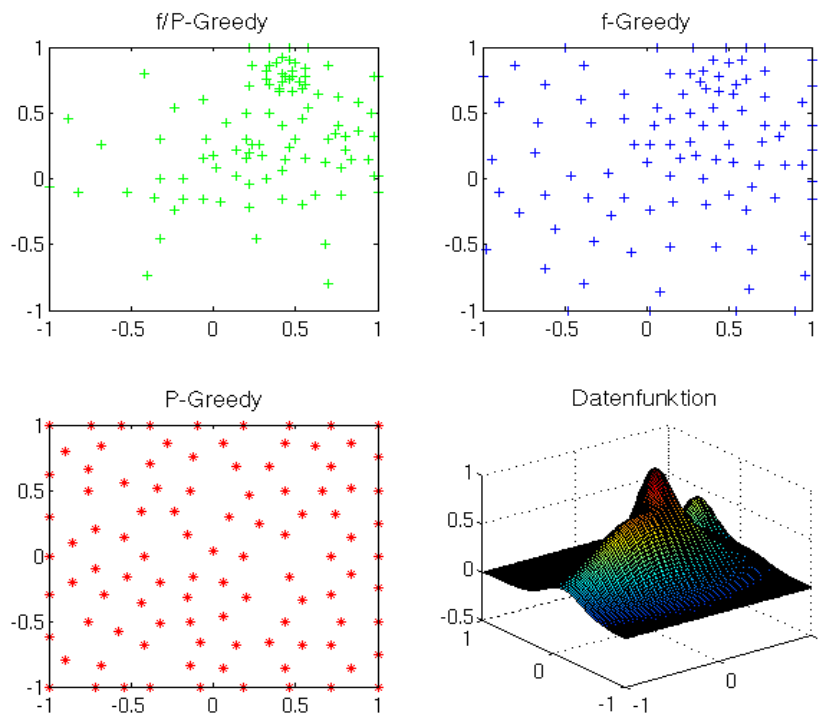


Abbildung 6.2: Punkteverteilung mit Franke's Testfunktion als Datenfunktion

6.2 Fehlerkurven der Greedy-Methoden

Bei den Fehlerabschätzungen wurde jeweils der L_∞ -Fehler ausgewertet (Abb. 6.3, 6.4). Dabei schneidet f -Greedy von den betrachteten Verfahren mit Abstand am besten ab. Allerdings sieht man, dass der Fehler nicht monoton fällt, sondern durch das Hinzufügen eines weiteren Punktes kurzfristig ansteigen kann. Im Gegensatz dazu fällt die Native Space Norm der Fehlerfunktion monoton (siehe Satz 2.2.9).

Bei hinreichend großer Glätte der Datenfunktion ist der Fehler von f/P -Greedy zwischen dem von f -Greedy und P -Greedy. Es wurde auch eine Greedy-Version getestet, bei der x_{k+1} als das Maximum von $|(s_{f, X_k} - f)(\cdot)|_{P_{K, X_k}(\cdot)}$ ausgewählt wird. Dabei ist der Fehler kleiner als bei P -Greedy und die Kondition ist besser als bei f -Greedy, da die Punkte gleichmäßiger verteilt werden. Diese Methode kombiniert die Eigenschaften von f -Greedy (kleiner Fehler) mit denen von P -Greedy (gleichmäßige Punktverteilung) und kann damit auch dazu benutzt werden, um zwischen Fehler und Kondition auszubalancieren.

Bemerkenswert ist, dass bei P -Greedy und f/P -Greedy „treppenartige“ Fehlerplots entstehen. Bei P -Greedy kann man beobachten, dass Schritt für Schritt die größeren punktfreien Löcher gefüllt werden, ohne dass sich der globale Fehler stark verändert. In dem Schritt in dem der Füllabstand fällt, fällt dann auch deutlich der globale Fehler. Bei f/P -Greedy ist dagegen unklar, woher dieses Verhalten kommt. Es wird wieder das Setting verwendet, das am Anfang des Kapitels beschrieben ist. Der erste Plot zeigt die Fehlerkurven für die Betragsfunktion und der zweite die für Frankes Testfunktion.

6.2: Fehlerkurven der Greedy-Methoden

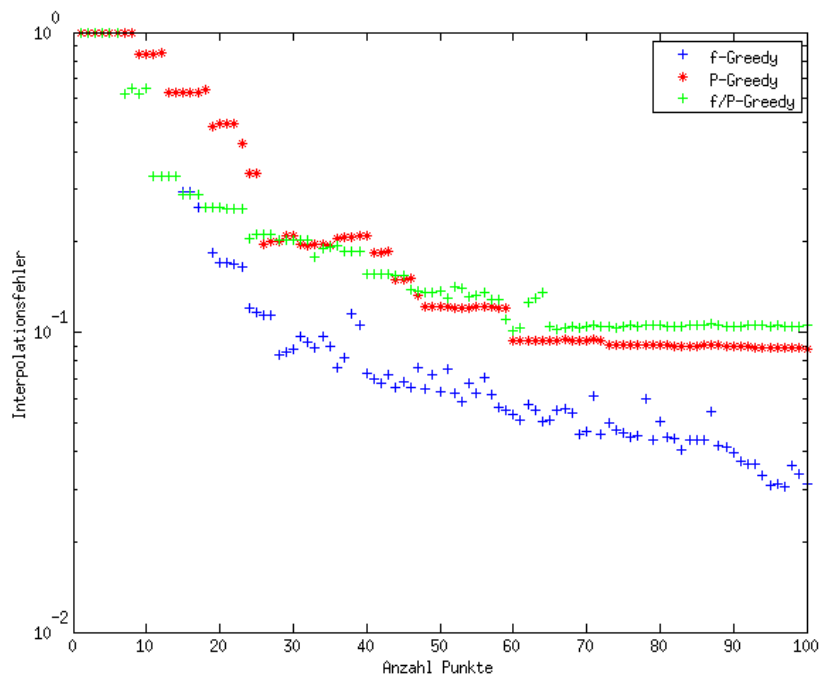


Abbildung 6.3: Fehlerkurven der Greedy-Verfahren bei Betragsfunktion

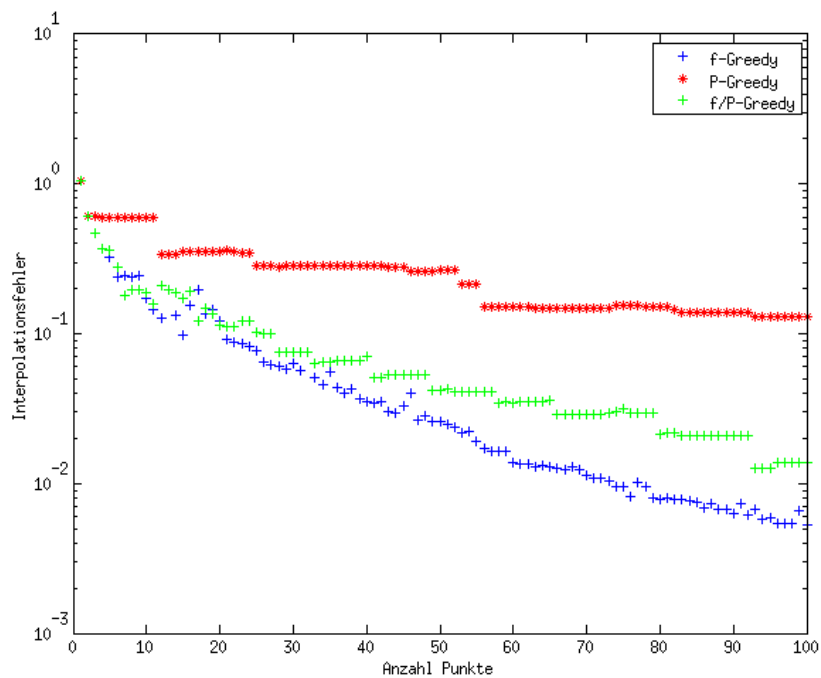


Abbildung 6.4: Fehlerkurven der Greedy-Verfahren bei Frankes Testfunktion

Für f -Greedy werden nun die Fehlerkurven noch genauer untersucht. Zur Schätzung der Konvergenzgeschwindigkeit wurde jeweils ein „least-squares fit“ mit einer Geraden an die geeignet transformierten Daten gemacht. Für die Wendlandfunktionen wurde jeweils sowohl die Anzahl der Punkte $i = 10, \dots, n$ als auch der zugehörige Fehler ε_i logarithmiert (die ersten Punkte wurden nicht berücksichtigt, weil sich dort das Verhalten der Fehlerkurve noch zu stark geändert hat). Durch folgende Beziehung lässt sich über die Steigung m der Geraden anschließend die Ordnung bestimmen

$$\begin{aligned} \ln(\varepsilon_n) - \ln(\varepsilon_1) &= m (\ln(n) - \ln(1)) \\ \Rightarrow \ln\left(\frac{\varepsilon_n}{\varepsilon_1}\right) &= m \ln(n) \\ \Rightarrow \varepsilon_n &= \varepsilon_1 n^m. \end{aligned}$$

Dagegen wurde bei dem Gauss-Kern nur der Fehlervektor logarithmiert, um über folgende Gleichung exponentielle Ordnung zu zeigen

$$\begin{aligned} \ln\left(\frac{\varepsilon_n}{\varepsilon_1}\right) &= m(n-1) \\ \Rightarrow \varepsilon_n &= \varepsilon_1 \exp(m(n-1)). \end{aligned} \tag{6.2}$$

1. $f(x) = 1/(25x^2 + 1)$, Kern $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$ (Abb. 6.5). Von 10000 äquidistanten Punkten aus $[0, 1]$ wurden 100 mit f -Greedy ausgewählt. Die Gerade, mit der der oben genannte least-squares fit durchgeführt wurde, hatte dabei eine Steigung von

$$m = \frac{-18,0714 + 4,72084}{5 - 2} = -4,4502 \quad (\text{siehe Grafik}).$$

Entsprechend den Abschätzungen aus Kapitel 3 ($\varepsilon_n \leq c n^{-\ell/d}$) entspricht das einer Konvergenzordnung von $\ell = 4,45$ mit Raumdimension $d = 1$. Im Vergleich dazu lieferte derselbe Test mit äquidistanten Daten $\ell = 2.14$ (siehe auch die Bemerkung zu Greedy und dem Randeffect in 3.4.6). Nach Satz 11.17 in [22] gilt für diese Basisfunktion im Native Space die Abschätzung¹

$$\begin{aligned} \|f - s_{f,X}\|_{L^\infty(\Omega)} &\leq Ch(X, \Omega)^{\frac{3}{2}} \|f\|_{\mathcal{N}_K(\Omega)} \\ &\stackrel{2.1.7}{\leq} C_2 n^{-\frac{3}{2d}} \|f\|_{\mathcal{N}_K(\Omega)}, \quad \text{also } \ell = 1, 5. \end{aligned}$$

¹Dort ist die Abschätzung für $d = 3$ gezeigt, was aber bei der Berechnung von ℓ berücksichtigt wird.

6.2: Fehlerkurven der Greedy-Methoden

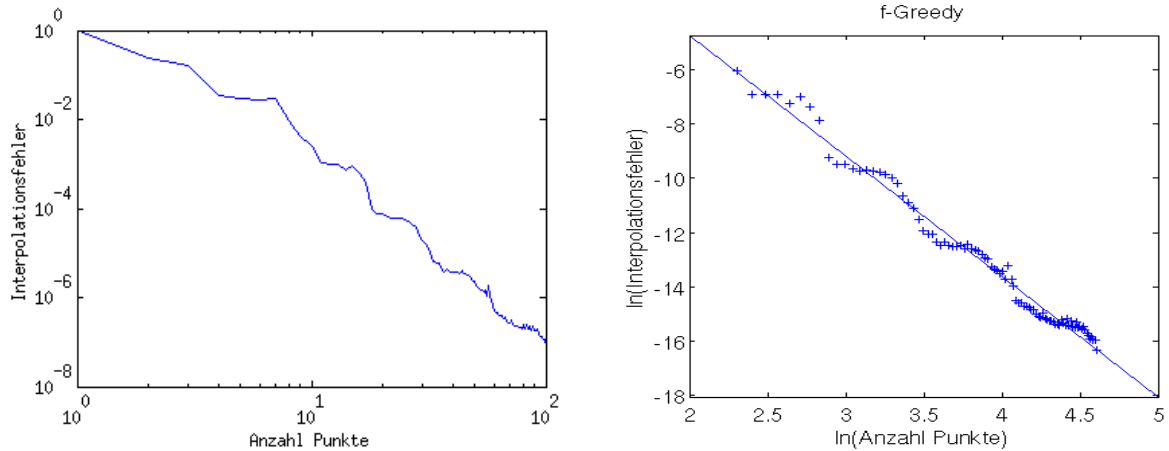


Abbildung 6.5: Fehlerkurve bei Approximation von $f(x) = 1/(25x^2 + 1)$, Kernfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$

2. $f(x, y) = peaks(x, y)$, Kern $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$ (Abb. 6.6). Es wurden 100 Punkte aus einem äquidistantem Gitter mit insgesamt 10000 Punkten aus $[0, 1]^2$ ausgewählt. Die Geradensteigung des least-squares fit betrug

$$m = \frac{-5,93153 - 0,193169}{5 - 2} = -2,0416.$$

Da in diesem Beispiel die Raumdimension $d = 2$ beträgt, entspricht dies der Konvergenzordnung $\ell = 4,08$.

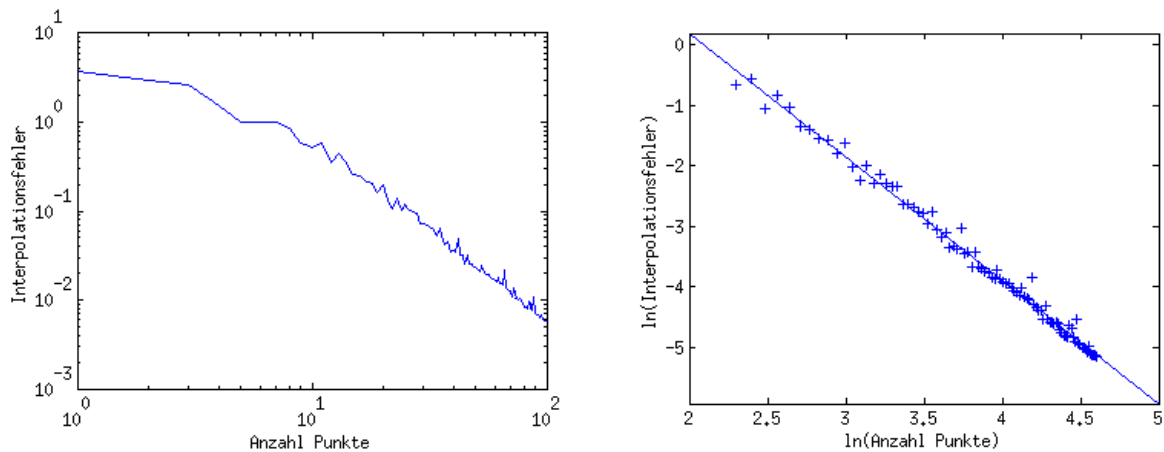


Abbildung 6.6: Fehlerkurve bei Approximation von $f(x, y) = peaks(x, y)$, Kernfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$

3. $f(x, y) = peaks(x, y)$, Kern $K(x, y) = \exp(-\|x - y\|_2^2)$ (Abb. 6.7). Statt 100 Punkte, wie in den vorhergehenden Beispielen, wurden wegen der schlechten Konditionierung nur

6.3: Vergleich von Greedy-Verfahren zu Gitterpunkten

60 Punkte des selben Gitters wie bei 2. ausgewählt. Die Geradensteigung des least-squares fit im Sinne von (6.2) betrug

$$m = \frac{-13,3591 + 1,74283}{60 - 10} = -0,2323,$$

so dass ein Fehlerverhalten $\varepsilon_n \approx \varepsilon_1 \exp(-0,2323(n - 1))$ resultiert.

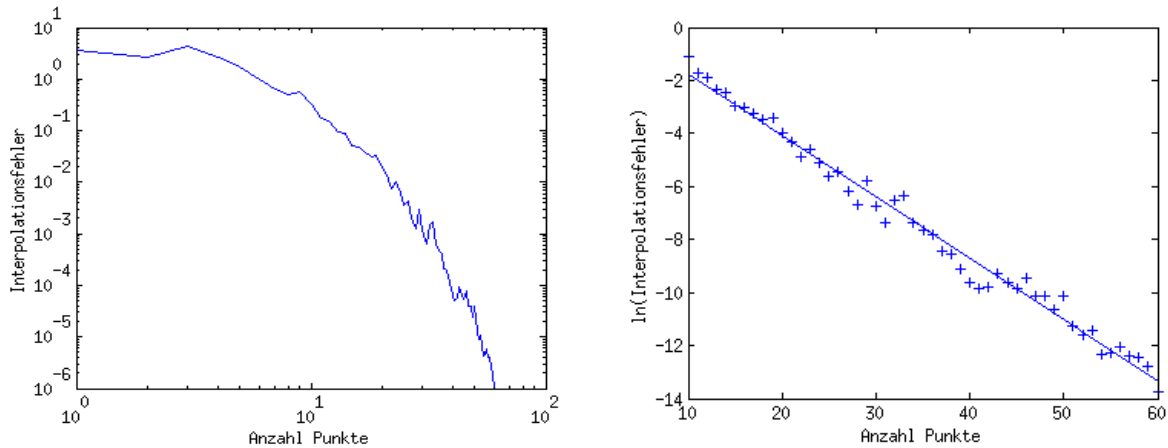


Abbildung 6.7: Fehlerkurve bei Approximation von $f(x, y) = peaks(x, y)$, Kernfunktion $K(x, y) = \exp(-\|x - y\|_2^2)$

6.3 Vergleich von Greedy-Verf. zu Gitterpunkten

In den folgenden numerischen Tests soll das Approximationsverhalten bei Punktwahl durch f -Greedy und P -Greedy im Vergleich zur Verwendung eines regelmäßigen Gitters untersucht werden. Dabei interessieren wir uns sowohl für den Fehler in Abhängigkeit der verwendeten Punkte als auch für die Stabilität, ausgedrückt durch die Kondition der Interpolationsmatrix. Die Punkte werden jeweils aus $[-1, 1]^2$ ausgewählt.

Bei f -Greedy und P -Greedy wird jeweils ein Gitter mit Punktabstand von 0,02 angeboten. Als dritte Punktwahl wird ein regelmäßiges Gitter mit $\lceil \sqrt{n} \rceil^2$ Punkten im n -ten Schritt verwendet. Es wird jeweils der L_∞ -Fehler auf dem Gitter mit Punktabstand 0,02 verglichen, was sich als hinreichend genau herausgestellt hat.

1. Im ersten Beispiel benutzen wir Franke's Testfunktion (siehe (6.1)) als Datenfunktion und interpolieren mit dem skalierten Gauss-Kern $K(x, y) = \exp(-50(\|x - y\|_2^2))$ (Abb. 6.8). Man sieht deutlich, dass f -Greedy wesentlich kleinere Fehler erzeugt als die beiden anderen Verfahren. Von daher ist zu erwarten, dass f -Greedy asymptotisch mindestens genauso kleine Fehler liefert wie ein regelmäßiges Gitter. In der Anwendung sah man, dass f -Greedy besser die Daten approximiert als ein regelmäßiges Gitter, da die Punktwahl viel besser an lokale Eigenschaften der Datenfunktion angepasst wird.

6.3: Vergleich von Greedy-Verfahren zu Gitterpunkten

Allerdings werden die Punkte bei f -Greedy ungleichmäßiger verteilt als bei den anderen beiden Methoden, was sich durch eine deutlich schlechtere Kondition der Interpolationsmatrix bemerkbar macht.

Im Vergleich zwischen P -Greedy und dem regelmäßigen Gitter sehen die Kurven ähnlich aus. Der absolute Fehler scheint bei Interpolation auf dem Gitter etwas kleiner als bei P -Greedy zu sein. Dafür ist in diesem Fall die Kondition der Interpolationsmatrix bei größeren Punktzahlen schlechter als bei P -Greedy. Es ist zu beachten, dass die Berechnung der Gitterpunkte deutlich schneller ist. Dafür lassen sich die P -Greedy Punkte flexibler einsetzen, da Punktzahl und Interpolationsgebiet frei gewählt werden können.

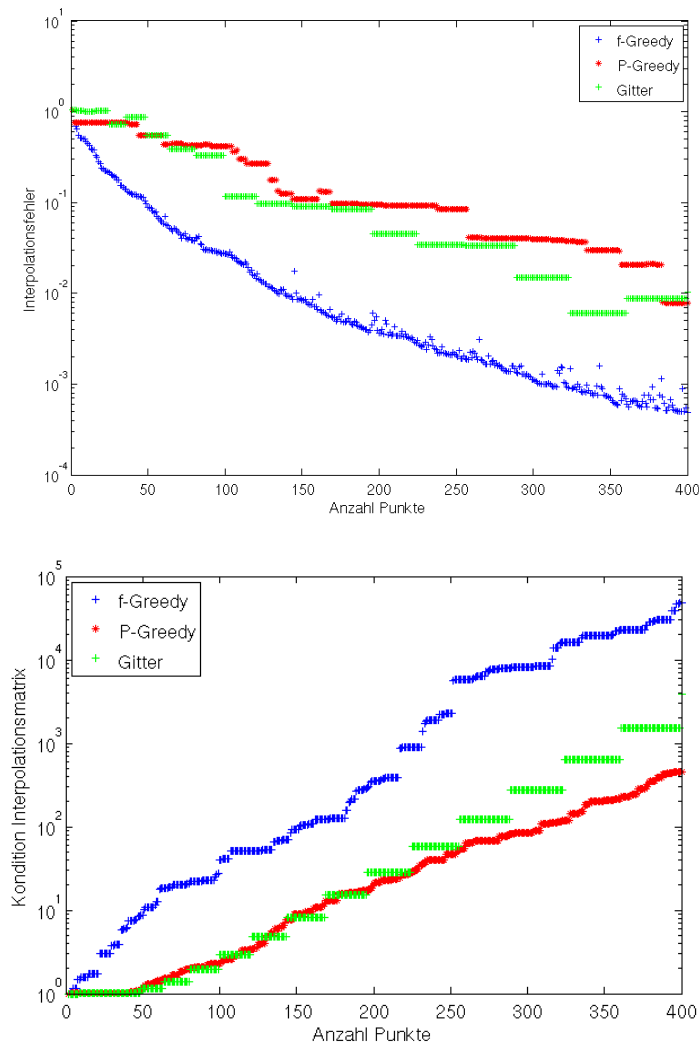


Abbildung 6.8: Fehler und Kondition bei Approximation von Frankes Testfunktion unter Verwendung von $K(x, y) = \exp(-50(\|x - y\|_2^2))$

6.3: Vergleich von Greedy-Verfahren zu Gitterpunkten

2. Im zweiten Test wurde als Datenfunktion die Matlab[©]-interne ([20]) „peaks“-Funktion benutzt (Abb. 6.9). In diesem Fall war die Kondition bei f -Greedy besser als im ersten Test, aber immer noch deutlich größer als bei den beiden anderen Verfahren. Da diese Verfahren f -unabhängig sind, wurde die Kondition hier nicht mehr dargestellt.

Insgesamt verhalten sich die Fehlerkurven schlechter als im ersten Beispiel. Für eine kleine Anzahl Punkte liegt das Gitter-Verfahren dicht bei f -Greedy, bei Verwendung vieler Punkte ist der Fehler aber sehr ähnlich wie bei P -Greedy. f -Greedy schneidet hier auch mit Abstand am besten ab.

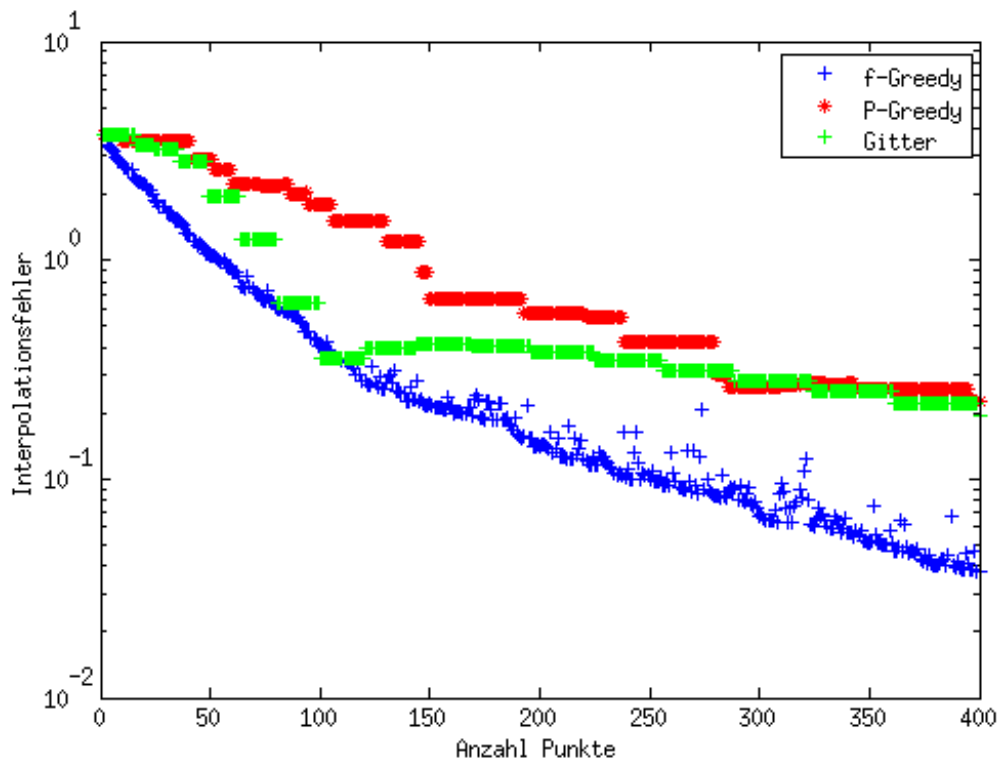


Abbildung 6.9: Fehler bei $K(x, y) = \exp(-50(\|x - y\|_2^2))$ und „peaks“-Funktion

6.3: Vergleich von Greedy-Verfahren zu Gitterpunkten

3. Als rauhere Datenfunktion wurde schließlich noch die Betragsfunktion $f(x, y) = |x|$ getestet (Abb. 6.10). Der Test fiel ähnlich aus wie bei der „peaks“-Funktion. Bemerkenswert ist aber, dass sich f -Greedy bei größerer Punktzahl (> 200 Punkte) in diesem Beispiel sehr instabil verhalten hat, obwohl die Kondition kleiner als 10^5 war. Dies scheint an der im Vergleich zum glatten Gauss-Kern sehr rauhen Datenfunktion zu liegen. Bei einer Skalierung des Gauss-Kerns mit Faktor 100 trat dieser Effekt nicht mehr auf.

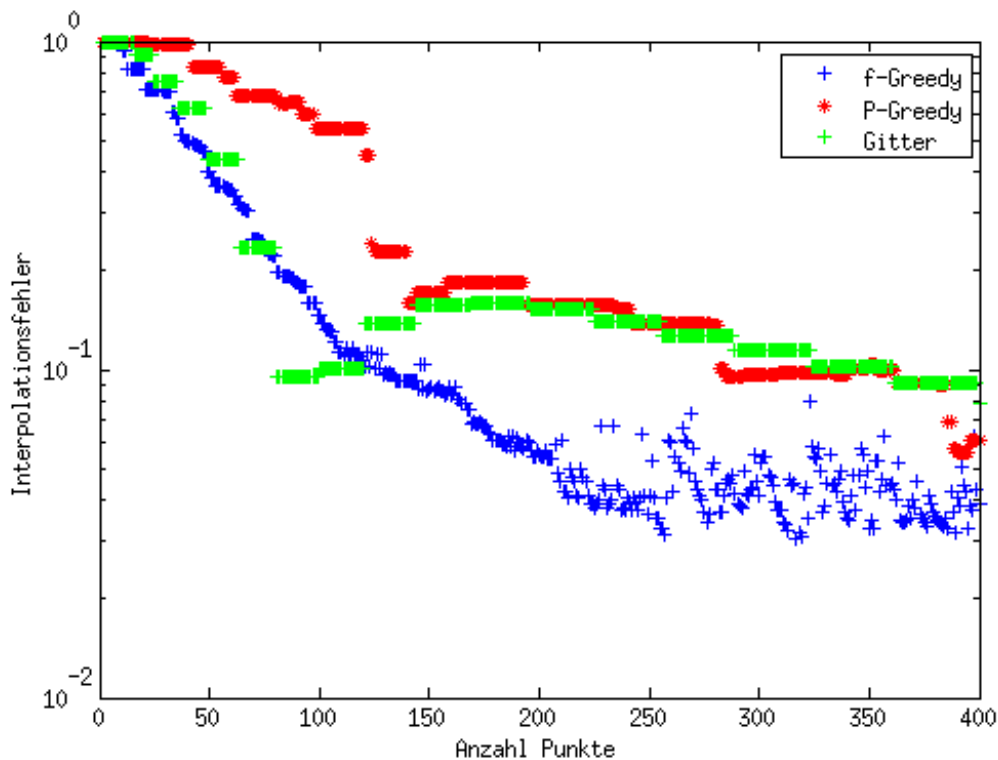


Abbildung 6.10: Fehler bei $K(x, y) = \exp(-50(\|x - y\|_2^2))$ und $f(x, y) = |x|$

Die Beispiele der Interpolation mit dem Gauss-Kern werden nun nochmal mit der Wendlandfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4(4\|x - y\|_2 + 1)$ als Kernfunktion präsentiert. Die Ergebnisse sind für die Wendlandfunktion noch klarer als für den Gauss-Kern. Man sieht, dass die Approximation mit f -Greedy am besten funktioniert, aber dafür liefert dieses Verfahren auch die schlechtesten Konditionszahlen. P -Greedy und die Interpolation auf dem Gitter verhalten sich sehr ähnlich, sowohl bei der Fehlerkurve als auch bei den Konditionszahlen.

6.3: Vergleich von Greedy-Verfahren zu Gitterpunkten

4. Der erste Plot zeigt die Fehlerkurven bei Verwendung von Frankes Testfunktion als Datenfunktion.

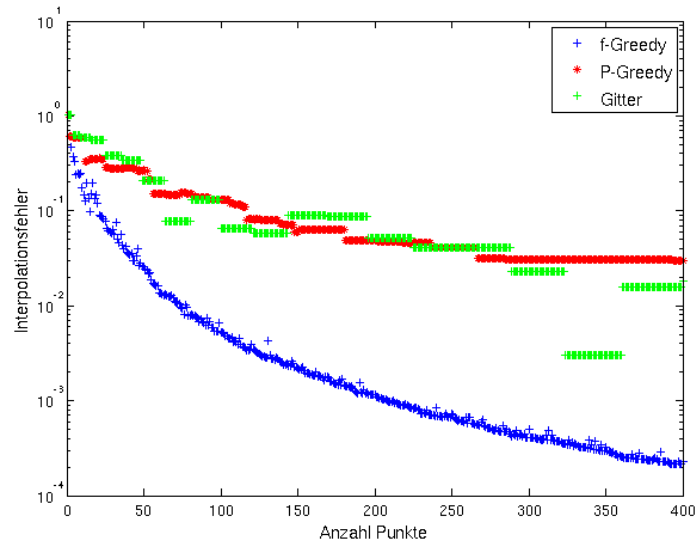


Abbildung 6.11: Fehler bei Approximation von Frankes Testfunktion mit der Wendlandfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4(4\|x - y\|_2 + 1)$

5. Die folgende Grafik zeigt die Konditionszahlen bei der Punktwahl durch P -Greedy, f -Greedy und bei Verwendung eines regelmäßigen Gitters. Als Datenfunktion wurde Frankes Testfunktion benutzt, was aber nur die Kurve von f -Greedy beeinflusst.

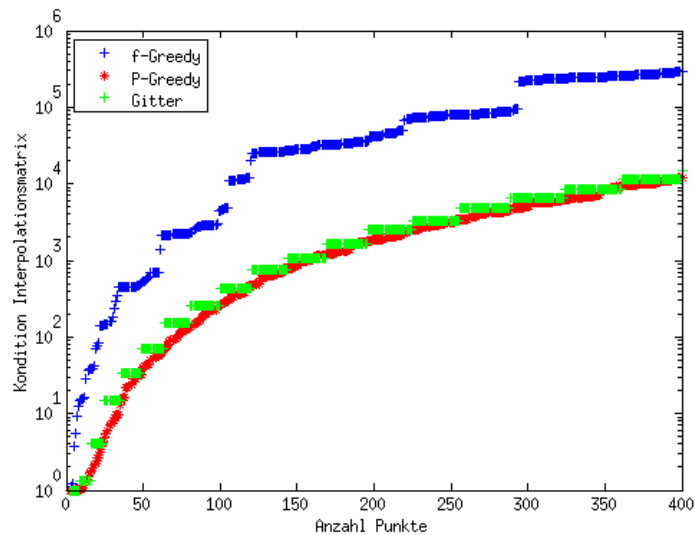


Abbildung 6.12: Kondition bei Approximation von Frankes Testfunktion mit der Wendlandfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4(4\|x - y\|_2 + 1)$

6.3: Vergleich von Greedy-Verfahren zu Gitterpunkten

6. Die beiden letzten Grafiken dieses Abschnitts (Abb. 6.13, 6.14) vergleichen die verschiedenen Methoden zur Punktauswahl für die „peaks“-Funktion und als Beispiel für eine rauhere Funktion für die Betragsfunktion. Bei der Betragsfunktion fällt auf, dass bei der Gitter-Methode der Fehler oszilliert, abhängig davon, ob die Datenpunkte bei dem Knick der Datenfunktion liegen oder nicht.

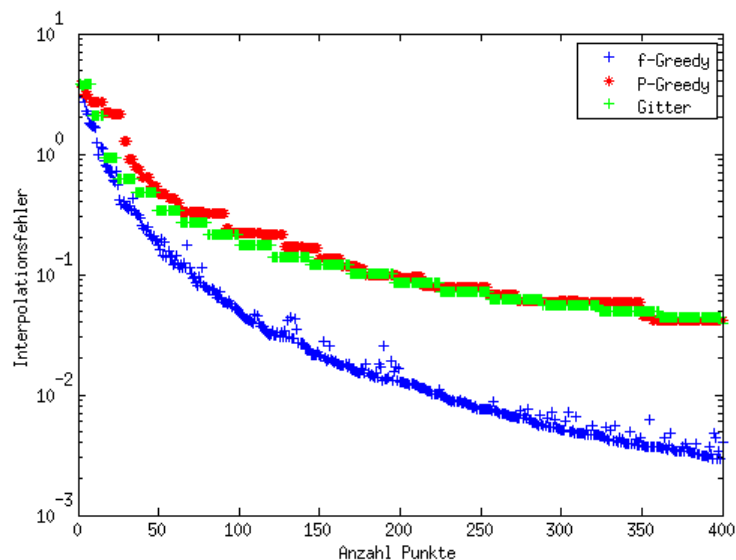


Abbildung 6.13: Fehler bei Approximation der „peaks“-Funktion mit der Wendlandfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$

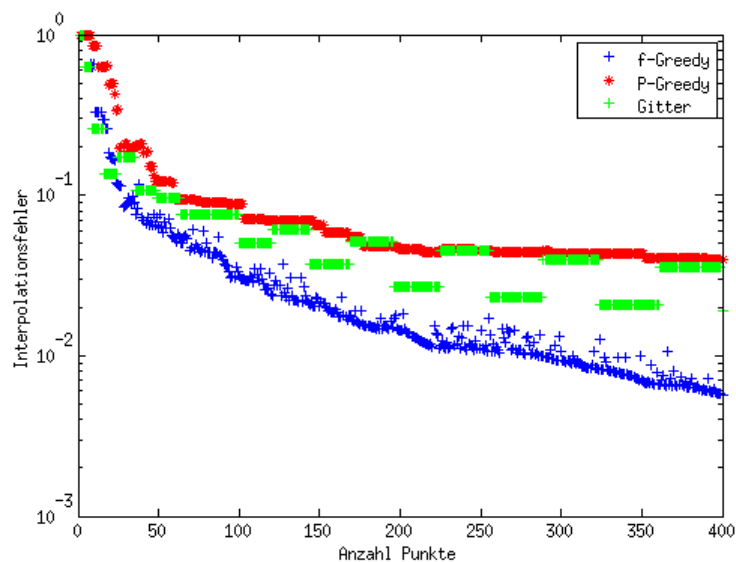


Abbildung 6.14: Fehler bei Approximation von $f(x, y) = |x|$ mit der Wendlandfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$

6.4 Vergleich von f -Greedy mit Optimallösung

Zum Abschluss der Untersuchung des Fehlerverhaltens von f -Greedy machen wir noch einen Vergleich mit der optimalen Lösung des Minimierungsproblems

$$\min_{\alpha \in \mathbb{R}^n} \|\alpha\|_{\ell_0} \quad \text{mit} \quad \|f(\cdot) - \sum_{j=1}^n \alpha_j K(\cdot, x_j)\|_{L_\infty(X_n)} \leq \varepsilon.$$

dabei ist $\|\alpha\|_{\ell_0}$ definiert als die Anzahl der von Null verschiedenen Einträge von α .

Dieses Problem ist vergleichbar mit „Support Vector Machines“ (siehe [1]), und es ist interessant, ob einfache Greedy-Verfahren mit diesen erheblich komplexeren Optimierungsverfahren aus der Lerntheorie mithalten können.

Das Optimierungsproblem wurde mit Xpress[©][24] gelöst. Dazu ist es nötig $|\alpha| \leq M$ für eine feste Schranke $M \in \mathbb{R}$ festzulegen, damit das Problem als lineares Programm umformuliert werden kann. Es wurde $M = 100$ gewählt und in den gerechneten Beispielen wurde diese Schranke auch nicht erreicht. Allerdings ließ sich dieses Problem für den Gauss-Kern nur in wenigen Fällen stabil rechnen. Deshalb wurden als Kerne die Wendlandfunktionen $K(x, y) = (1 - |x - y|)_+^4 (4|x - y| + 1)$ (W2) und $K(x, y) = (1 - |x - y|)_+$ (W0) verwendet. Das zugehörige Programm ist in Anhang A zu finden.

Aus Komplexitätsgründen verwendeten wir $X_n \subset [0, 1]$ mit $n = 80$ äquidistanten Punkten. Bei Einträgen, die mit (*) markiert sind, konnten nur 60 Punkte verwendet werden. Die Anzahl der benötigten Datenpunkte bei vorgegebenen Fehlerlevel ist für verschiedene Datenfunktionen f in Tabelle 6.1 abgebildet.

	$x^2 - 3x$		$\sin(x)$		$\sin(5x)$		$ 0,5 - x $	
Fehlerlevel	W2	W0	W2	W0	W2	W0	W2	W0
0,1	3	2	2	1	3	4	3	3
0,05	3	3	2	1	3	6	3	3
0,01	6	5	5	3	4	11(*)	7	3
0,005	6	7	5	4	6	16(*)	7	4
0,001	8	13(*)	7	7	8	36	10	4

Tabelle 6.1: Anzahl Punkte der Optimallösung

Im Vergleich dazu listen wir in Tabelle 6.2 die Anzahl der Punkte auf, die f -Greedy aus X_n benötigte, um dasselbe Fehlerlevel auf X_n wie die Optimallösung zu erreichen.

6.5: Stabilität der Greedy-Verfahren

Fehlerlevel	$x^2 - 3x$		$\sin(x)$		$\sin(5x)$		$ 0,5 - x $	
	W2	W0	W2	W0	W2	W0	W2	W0
0,1	4	3	2	1	4	6	3	3
0,05	6	5	3	2	5	10	7	3
0,01	9	9	6	5	7	22	11	4
0,005	10	9	9	5	8	25	13	4
0,001	13	17	13	11	13	67	20	4

Tabelle 6.2: Anzahl Punkte bei f -Greedy

Auch bei diesem Test ist festzustellen, dass f -Greedy erstaunlich gut abschneidet. Im Vergleich zur Optimallösung macht sich bemerkbar, dass bei der Punktwahl mit f -Greedy interpoliert wurde, die Optimallösung diese Bedingung aber nicht erfüllen muss. Außerdem kann bei f -Greedy ein Punkt, der einmal ausgewählt wurde, nicht wieder entfernt werden. Allerdings kommt f -Greedy auch schon mit sehr wenigen Punkten aus und die Optimallösung lässt sich auch nur für extrem kleine Beispiele effizient berechnen. Inwiefern sich Greedy-Verfahren als suboptimale Löser von Lernproblemen eignen, ist ein interessantes Forschungsproblem.

6.5 Stabilität der Greedy-Verfahren

Durch die starke Clusterung sorgt die Punktauswahl durch f/P -Greedy für eine deutlich höhere Konditionszahl der Interpolationsmatrix. Insgesamt scheint dieses Verfahren schlecht geeignet. Zwar wird in jedem einzelnen Schritt die Native Space Norm der Fehlerfunktion minimiert, aber sowohl Fehler als auch Kondition der Interpolationsmatrix verhalten sich schlecht.

Im Vergleich dieser drei Verfahren wurde die beste Kondition durch P -Greedy erzielt, weil damit die Punkte sehr gleichmäßig verteilt werden. Außerdem ist P -Greedy f -unabhängig. Auf unförmigen Gebieten, bei denen man nicht einfach ein Gitter verwenden kann, ist P -Greedy geeignet, falls f unbekannt ist.

Bei f -Greedy ist die Kondition der Interpolationsmatrix deutlich höher als bei P -Greedy, dafür verhält sich aber der Interpolationsfehler sehr gut, wodurch nur wenige Punkte verwendet werden müssen. Es wurde hier auch wieder die Interpolationsaufgabe gelöst, die zu Beginn des Kapitels vorgestellt wurde. Im ersten Beispiel wurde die Betragsfunktion als Datenfunktion benutzt und im zweiten Beispiel Frankes Testfunktion.

6.5: Stabilität der Greedy-Verfahren

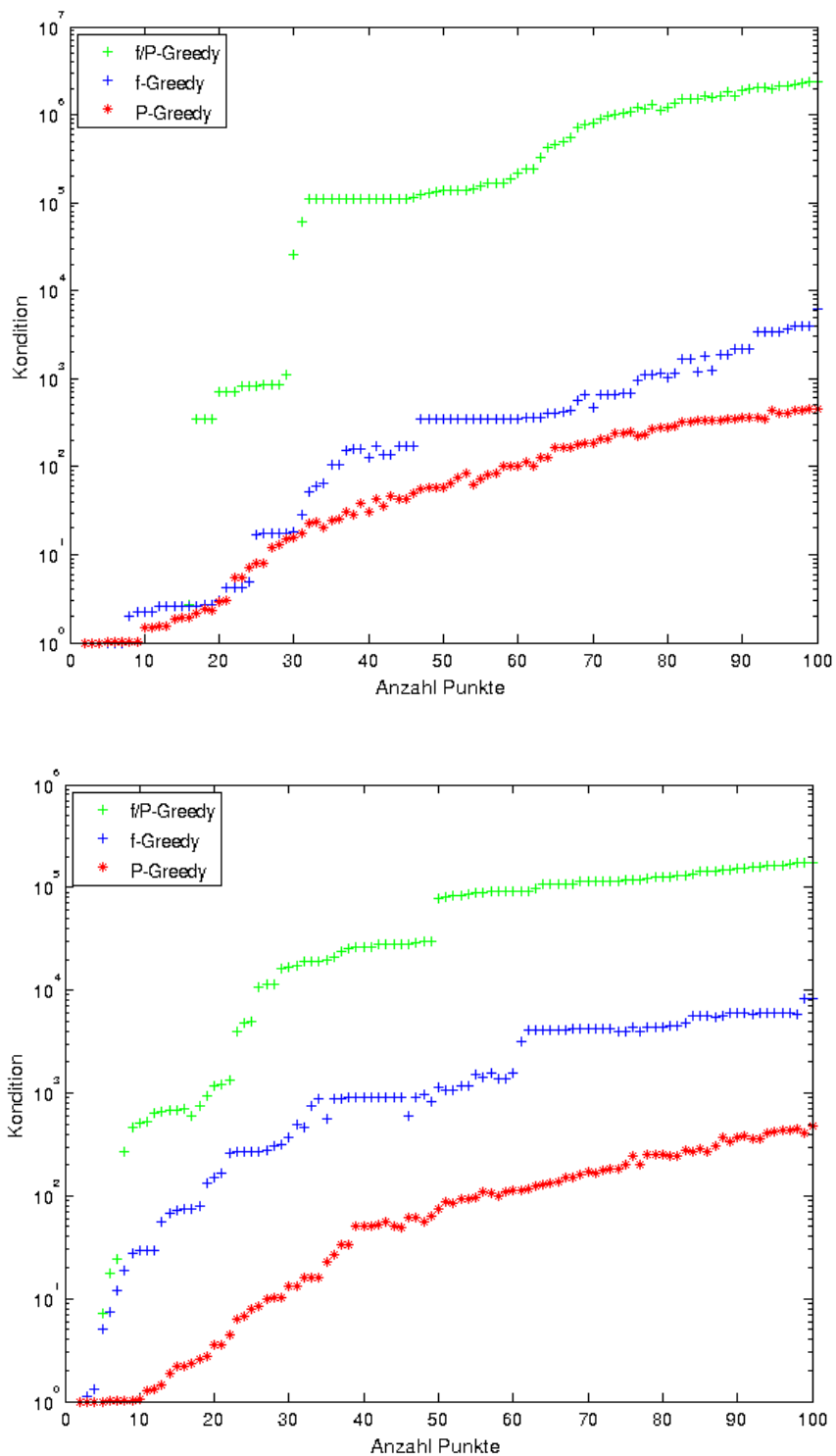


Abbildung 6.15: Konditionszahlen bei Greedy-Punktwahl und Approximation der Betragsfunktion (oben) bzw. Frankes Testfunktion (unten)

Kapitel 7

Tests zur Newton-Basis

7.1 Stabilitätsverhalten

Wie in Kapitel 4 gezeigt wurde, besitzt die Newton-Basis gute Stabilitätseigenschaften (siehe z.B. Satz 4.3.4 und Satz 4.3.8). Dazu sollen nun auch numerische Tests durchgeführt werden. Wir interessieren uns dabei für die Größe der Koeffizienten, die Konditionszahl der Interpolationsmatrix und $\sum_j |b_j| |B_j(x)|$ (siehe auch 4.3.8), wobei B_j das j -te Element der verwendeten Basis mit zugehörigem Koeffizient b_j ist. Als Basen verwenden wir die Kern-Basis, die Newton-Basis (siehe Definition 4.2.1) und die Lagrange-Basis (siehe Definition 2.2.10).

Zunächst folgt aber noch eine kurze Bemerkung zur Sortierung der Datenpunkte. Wie bereits in 4.2.2 festgestellt wurde, hängt die Newton-Basis im Gegensatz zur Kernbasis von der Sortierung ab. Wie in Abbildung 7.1 zu sehen ist, führte eine Sortierung derselben Punkte nach x - bzw. y -Koordinate dazu, dass die Konditionszahl der Matrix bei kleiner Anzahl an Punkten riesengroß war. Weil die Koeffizienten für die Newton-Basis rekursiv berechnet werden, ist in diesem Fall das Verfahren unbrauchbar. Das ist erstaunlich, weil mit denselben Punkten bei geeigneter Reihenfolge die Newton-Basis deutlich stabiler als die Standardbasis ist. Falls die Punktmenge bereits gegeben ist, muss deshalb bei der Berechnung der Newton-Basis besonders auf die Sortierung der Punkte geachtet werden. Nach Bemerkung 5.3.4 kann man die Punkte zum Beispiel dadurch sortieren, dass man auf der gegebenen Punktmenge P -Greedy durchführt.

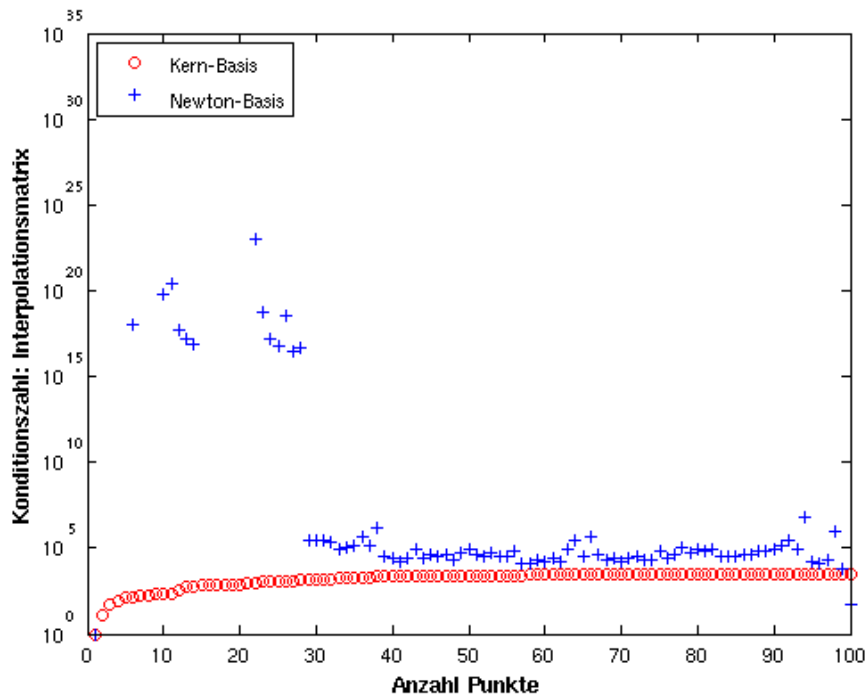


Abbildung 7.1: Kondition bei Verwendung der Newton-Basis und ungünstig sortierten Punkten

Bei den Tests wird jeweils die Kernbasis $\{K(\cdot, x_j)\}_{j=1}^n$ mit der entsprechenden Newton-Basis (siehe 4.2.1) des selben Unterraums verglichen. Die Datenpunkte wurden durch P -Greedy (siehe Def. 3.1.1) mit der Kernfunktion $K(x, y) = (1 - \frac{\|x-y\|_2}{100})_+^4 (4 \frac{\|x-y\|_2}{100} + 1)$ auf einem Gitter in $[-3, 3]^2$ mit Gitterweite 0,02 gewählt. Als Datenfunktion wurde die Matlab[©]-interne [20] Funktion „peaks“ verwendet, die durch Kombination von Gauss-Kernen erzeugt wird. Für die Interpolation mit der Wendlandfunktion $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$ wurden 200 Punkte verwendet, bei Interpolation mit dem skalierten Gauss-Kern $K(x, y) = \exp(-\frac{\|x-y\|_2^2}{25})$ wegen schlechter Konditionierung nur 60 Punkte.

Bei den Koeffizienten sieht man große Unterschiede zwischen der Kernbasis und der Newton-Basis (Abb. 7.2 und 7.3). Während die Koeffizienten bei Wahl der Newton-Basis nach Satz 4.3.4 beschränkt bleiben, lassen die Plots bei Verwendung des Gauss-Kerns für die Koeffizienten der Standardbasis exponentielles Wachstum vermuten. Bei Verwendung der Wendlandfunktionen mit kompaktem Träger sind die Unterschiede nicht ganz so groß, aber dennoch deutlich zu erkennen. Es sieht aber so aus, als würden in diesem Fall bei beiden Basen die Koeffizienten beschränkt bleiben.

7.1: Stabilitätsverhalten

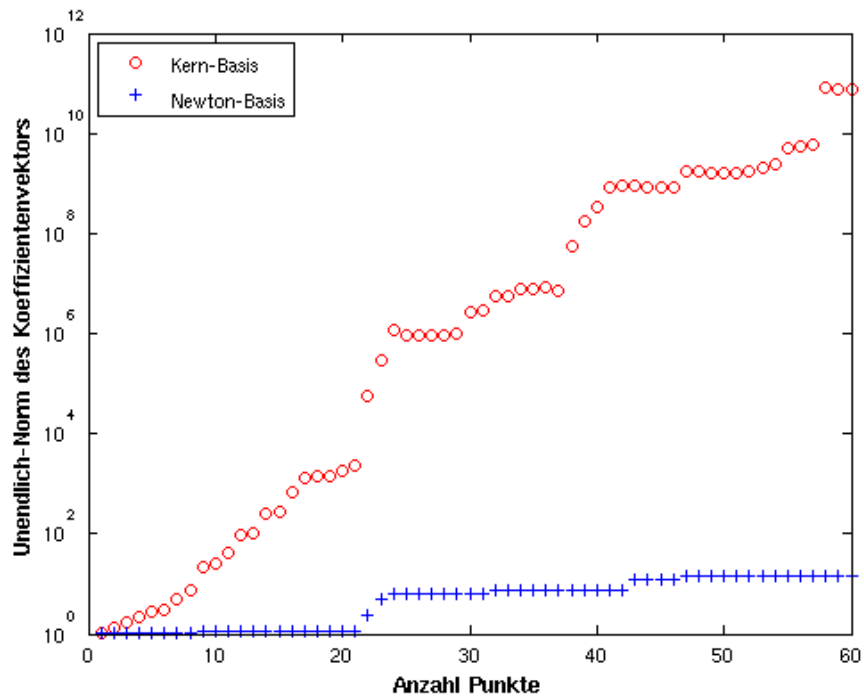


Abbildung 7.2: Koeffizienten bei Interpolation mit Gauss-Kern

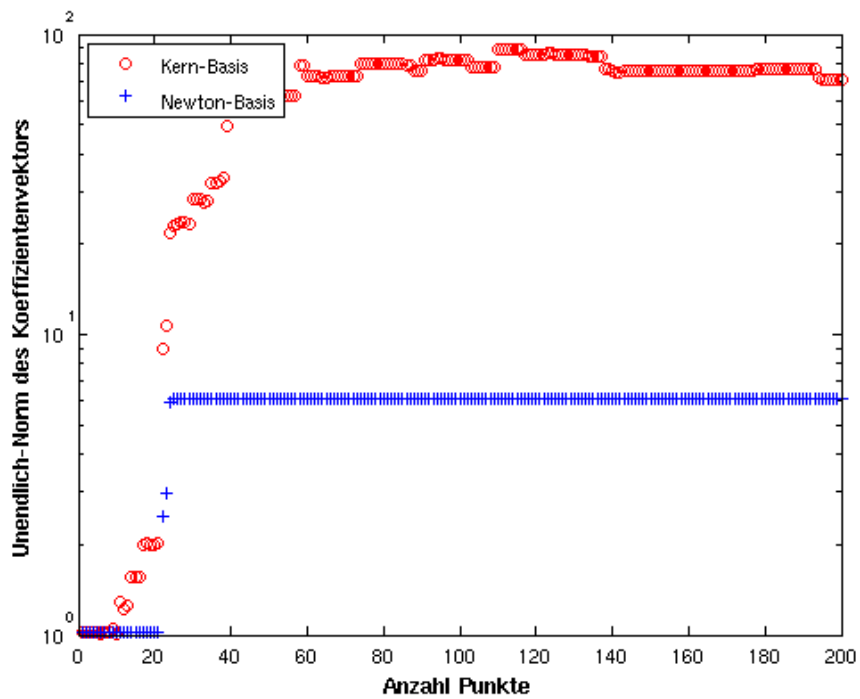


Abbildung 7.3: Koeffizienten bei Interpolation mit Wendlandfunktion

7.1: Stabilitätsverhalten

Die Konditionszahl der Interpolationsmatrix verhält sich ebenfalls deutlich besser für die Newton-Basis .

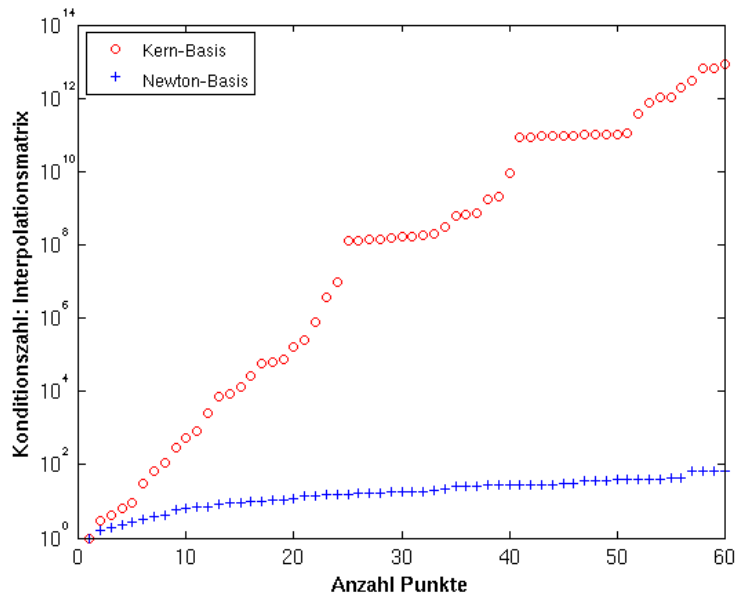


Abbildung 7.4: Kondition der Interpolationsmatrix bei Interpolation mit Gauss-Kern

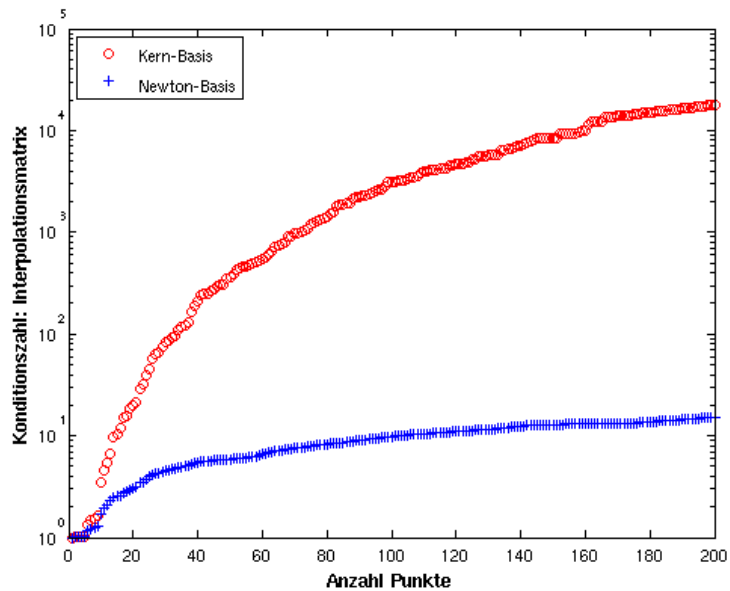


Abbildung 7.5: Kondition der Interpolationsmatrix bei Interpolation mit Wendland-funktion

7.1: Stabilitätsverhalten

Zuletzt betrachten wir noch gemäß Satz 4.3.8 $\sum_j |b_j| |B_j(x)|$, wobei B_j wie zu Beginn des Kapitels das j -te Element der verwendeten Basis mit zugehörigem Koeffizient b_j ist (Abb. 7.6, 7.7). Erwartungsgemäß verhält sich auch hier die Newton-Basis deutlich besser. Allerdings ist aus den Kurven nicht zu sehen, ob die Schranke

$$\sum_j |\lambda_j(f)| |v_j(x)| \leq C \sqrt{n} \|f\|_{\mathcal{N}_K} \quad (7.1)$$

aus Satz 4.3.8 scharf ist. Dazu müsste man nämlich die ungünstigste Sortierung der Punkte betrachten, was wie oben erwähnt wegen der schlechten Konditionierung nicht funktioniert hat. Bei den getesteten Beispielen schien die theoretische Schranke noch viel zu pessimistisch, allerdings ist aus den Plots nicht klar, wie das asymptotische Verhalten aussieht, da dafür die Anzahl der Punkte zu klein ist.

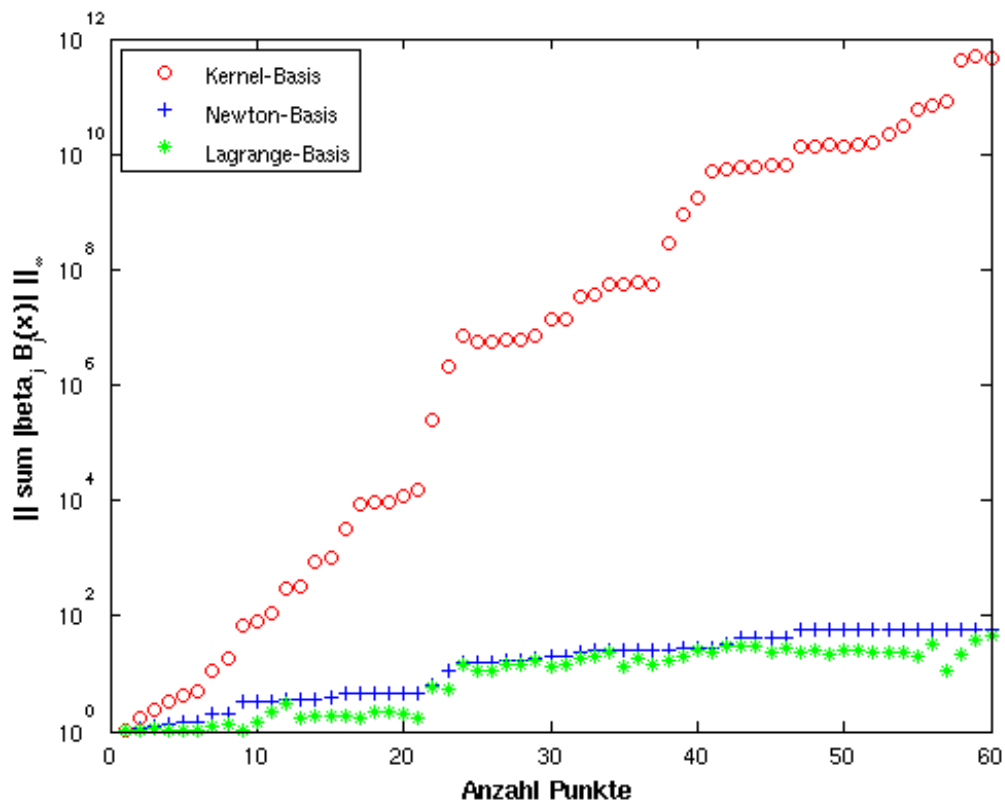


Abbildung 7.6: Stabilitätsvergleich der Basen, $K(x, y) = \exp(-\frac{\|x-y\|_2^2}{25})$

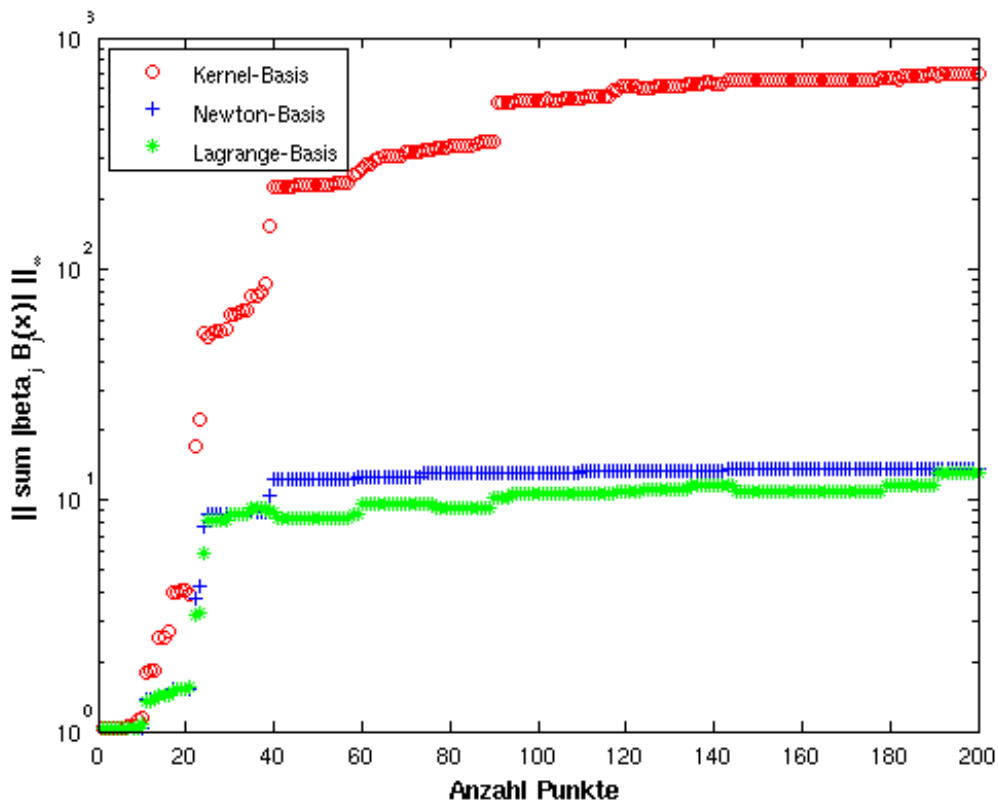


Abbildung 7.7: Stabilitätsvergleich für $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$

7.2 Die Newton-Basis kombiniert mit P -Greedy

Abschließend sollen hier noch typische Elemente der Newton-Basis bei Verwendung von P -Greedy grafisch veranschaulicht werden. Es wurden jeweils 60 Punkte aus einem Gitter mit Gitterweite 0,02 aus $[-1, 1]^2$ durch P -Greedy (Definition 3.1.1) mit verschiedenen Kernen ausgewählt. Verwendet wurden der Gauss-Kern, Inverse Multiquadric (IMQ) und zwei verschieden skalierte Wendlandfunktionen (W2, Interpolante in C^2). Auf der rechten Seite sieht man jeweils die Lage der Punkte, der letzte Punkt x_n ist als roter Kreis eingezeichnet. Dazu ist auf der linken Seite die letzte Basisfunktion v_n geplottet. Es gilt $v_n(x_n) = 1$, $\|v_n\|_{L_\infty([-1,1]^2)} \leq 1$, $v_n(x_i) = 0$, für $i < n$ (siehe Satz 5.2.1).

Die Punktverteilungen scheinen dabei quasi-uniform zu sein, aber kernabhängig. Es fällt auf, dass die Basisfunktion bei Verwendung des Gauss-Kerns sehr stark oszilliert, dagegen klingt sie für die Wendlandfunktionen sehr schnell auf Null ab. Es bestätigt sich, dass die Lagrangefunktionen zu Punkten im Inneren dort ein lokales Maximum haben und nirgendwo betragsmäßig den Wert 1 überschreiten.

7.2: Die Newton-Basis kombiniert mit P -Greedy

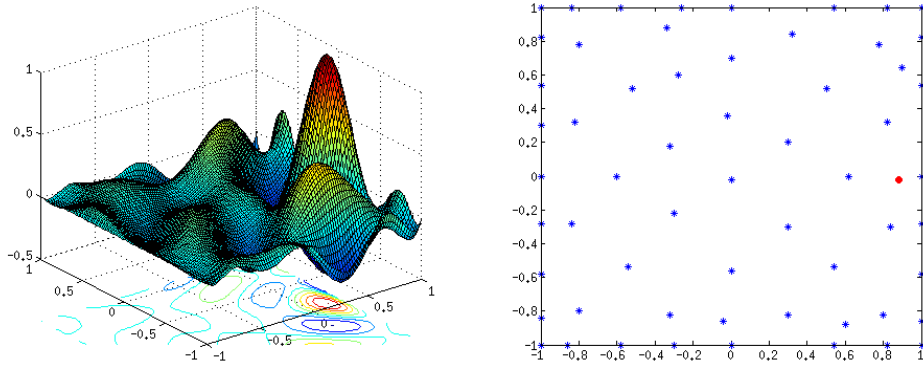


Abbildung 7.8: Newton-Basis für Gauss-Kern : $K(x, y) = e^{-\|x-y\|_2^2}$

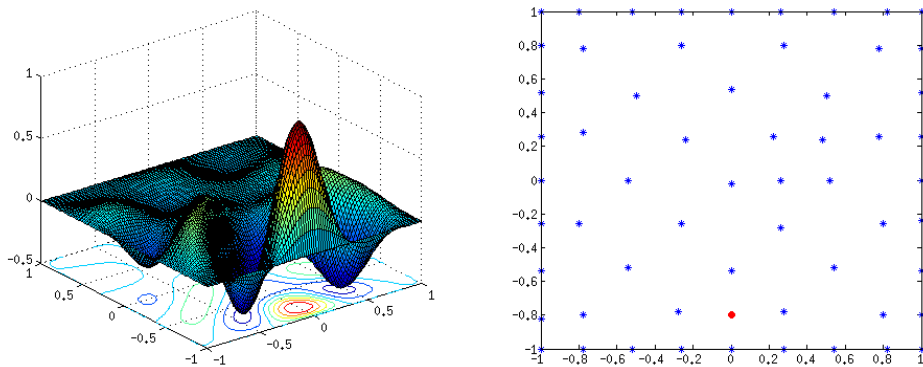


Abbildung 7.9: Newton-Basis für IMQ: $K(x, y) = 1/\sqrt{1 + \|x - y\|_2^2}$

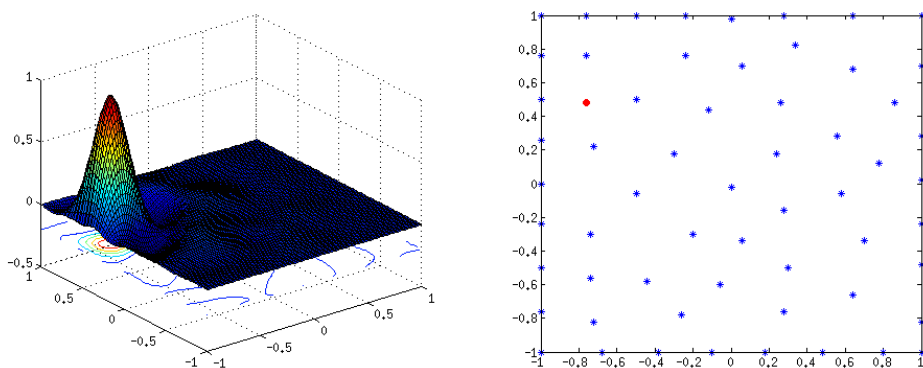


Abbildung 7.10: Newton-Basis für W2: $K(x, y) = (1 - \|x - y\|_2)_+^4 (4\|x - y\|_2 + 1)$

7.2: Die Newton-Basis kombiniert mit P -Greedy

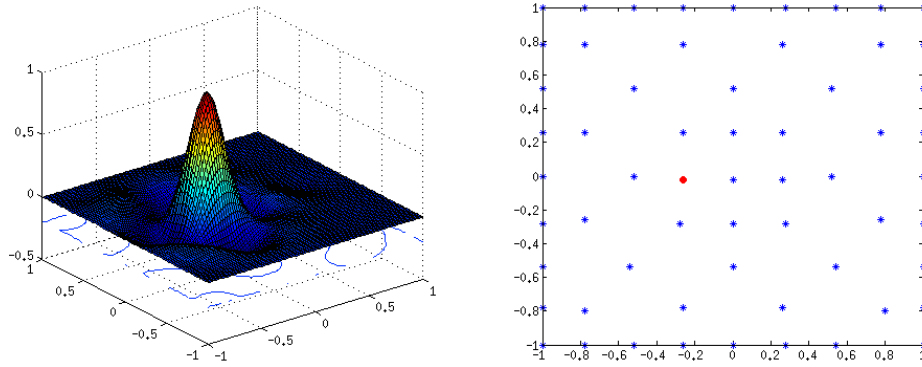


Abbildung 7.11: Newton-Basis für W_2 skaliert: $K(x, y) = (1 - \frac{\|x-y\|_2}{100})_+^4 (4 \frac{\|x-y\|_2+1}{100})$

Kapitel 8

Ergebnisse und Ausblick

Im Rahmen dieser Arbeit wurde bewiesen, dass bei der Punktwahl durch f -Greedy auf Intervallen für Teilfolgen mindestens die gleiche Konvergenzordnung wie bei äquidistanten Punkten erreicht wird (Korollar 3.4.5). Numerisch konnte ein wesentlich besseres Konvergenzverhalten als bei der Verwendung von Gitterpunkten gezeigt werden, da lokale Eigenschaften durch die Punktwahl berücksichtigt werden. Dadurch wird auch der sogenannte Randeffect umgangen, der ansonsten die Konvergenzgeschwindigkeit bei kernbasierter Interpolation stark verringert.

In höherer Dimension wurde für Teilfolgen zumindest lineare Konvergenz und im Inneren des Gebiets quadratische Konvergenz bewiesen (Korollar 3.3.8). Dafür wurde ausgenutzt, dass bei f -Greedy die Punkte nur dicht zusammenliegen können, wenn global der Fehler sehr klein ist. Außerdem wurde beispielhaft gezeigt, wie diese theoretischen Ergebnisse aus dem kontinuierlichen Fall bei hinreichend feinem Gitter auf den diskreten Fall übertragen werden können (Satz 3.3.18).

Zu dem Fehler- und Stabilitätsverhalten verschiedener Greedy-Verfahren wurden in Kapitel 6 numerische Tests durchgeführt. Dabei wurde deutlich, dass f -Greedy mit wenigen Punkten sehr kleine Fehler liefert, dagegen aber zu einem schlecht konditionierten Gleichungssystem führt. Dagegen sorgt die Punktauswahl durch P -Greedy für sehr gleichmäßige Punktverteilungen und dadurch mehr Stabilität, was allerdings durch ein schlechteres Fehlerverhalten „erkauft“ wird.

Zur Untersuchung der Stabilitätsprobleme bei der Interpolation mit Kernfunktionen wurde eine Newton-Basis für Kernfunktionen betrachtet und gezeigt, dass diese Basis sich stabil verhält (Satz 4.3.8). Das bedeutet, dass das Problem der Kerninterpolation an sich nicht schlecht konditioniert ist. Durch Wahl einer geeigneten Basis lassen sich Instabilitäten vermeiden, allerdings entstehen bei der Berechnung dieser Basis Probleme. Die dabei auftretende Ungenauigkeiten hängen aber zumindest nicht von den Fehlern in den Daten ab.

Darüber hinaus sind die Elemente der Newton-Basis im Native Space orthogonal zueinander. Mit dieser Orthogonalbasis ist es möglich unter schwachen Voraussetzun-

gen die Funktionen des Native Space in einer Reihe zu entwickeln (Satz 4.4.1). Eine spezielle Darstellung wurde dabei für den reproduzierenden Kern hergeleitet (Korollar 4.4.5). Schließlich konnte bewiesen werden, dass die Elemente der Newton-Basis ein globales Maximum an dem zugehörigen Interpolationspunkt besitzen und betragsmäßig durch 1 beschränkt sind, falls die verwendeten Punkte durch P -Greedy ausgewählt werden (Korollar 5.2.1). Das gute Stabilitätsverhalten der Newton-Basis wurde in Kapitel 7 zusätzlich durch numerische Tests belegt.

Ausblick

Für die Punktauswahl durch f -Greedy fehlen im Mehrdimensionalen bisher Konvergenzaussagen, die mit den besten Ergebnissen für äquidistante Punkte vergleichbar sind. Hierfür wäre es eventuell nützlich, statt der Füllichte ein anderes, von f abhängiges Maß für die Ausfüllung des Gebiets durch die Datenpunkte zu verwenden.

Desweiteren wäre es interessant, Greedy-Verfahren zu untersuchen, bei denen an den Stützstellen nicht exakt interpoliert wird. Besonders bei der Betrachtung verrauschter Daten könnte dies zu wesentlich besseren Rekonstruktionen führen.

Für die Newton-Basis stellt sich die Frage, ob es möglich ist, durch spezielle Algorithmen die Basiselemente effizienter zu berechnen und auszuwerten. Vielversprechend sehen dabei Ansätze aus, bei denen nur lokal Rekonstruktionsprobleme zu lösen sind, wodurch die Komplexität stark reduziert werden kann.

Anhang A

Xpress[©]-Programm

Das Programm zur Berechnung der Optimallösung aus Abschnitt 6.4 soll in diesem Abschnitt vorgestellt werden. Zunächst wird das umformulierte, lineare Programm gezeigt und anschließend wird der Quellcode aufgelistet.

Gesucht werden $z_j \in \{0, 1\}$, $\alpha_j \in [-M, M]$ für $j \in \{1, \dots, n\}$, die das folgende Minimierungsproblem zu Fehlerlevel $\varepsilon \in \mathbb{R}_+$ und Datenfunktion f auf den Punkten x_1, \dots, x_n lösen

$$\min_{z \in \{0,1\}^n} \sum_{j=1}^n z_j \quad \text{unter den Nebenbedingungen}$$

$$\left| \sum_{j=1}^n \alpha_j K(x_i, x_j) - f(x_i) \right| \leq \varepsilon, \quad 1 \leq i \leq n,$$
$$|\alpha_j| \leq z_j M.$$

```
model "kernelopt"
  uses "mmxprs"

  function kern(x,y: real):real      ! Kernfunktionen
    ! Gauss-Kern
    !returned:=exp(-((x-y))^2)
    ! Wendlandfunktion W0
    if 1-abs(x-y)>0
      then returned:=1-abs(x-y)
    else
      returned:=0
    end-if
    ! Wendlandfunktion W2
    !if 1-abs(x-y)>0
```

```

! then returned :=(1+4*abs(x-y))*(1-abs(x-y))^4;
!else
! returned:=0;
!end-if
end-function

function fun(x:real):real      ! Datenfunktionen
!returned := sin(x)
!returned:=1
!returned:=x^2-3*x
returned:=sin(5*x)
!returned:=abs(0.5-x)
end-function

declarations
status: string
N=80                ! Anzahl Datenpunkte
eps=0.01           ! Fehlertoleranz
M=100              ! Obere Schranke fuer Koeffizienten
X: array(1..N) of real ! Interpolationspunkte
data: array(1..N) of real ! Funktionswerte auf X
kernmatrix: array(1..N,1..N) of real
alpha: array(1..N) of mpvar ! Koeffizientenvektor
z: array(1..N) of mpvar ! boolsche Werte, die bestimmen, ob
! Koeffizient=0

end-declarations

forall (k in 1..N) do
alpha(k) is_free      ! alpha<0 wird zugelassen
end-do

! Festlegen der Datenpunkte, N Punkte in [0,1]
forall (j in 1..N) do
X(j):=(j-1)/(N-1)
end-do

! Aufbauen der Kernmatrix
forall (j in 1..N, k in 1..N | j<=k) do
kernmatrix(j,k):=kern(X(j),X(k))
kernmatrix(k,j):=kernmatrix(j,k)
end-do

! Berechnen der Datenpunkte

```

Kapitel A: Xpress[©]-Programm

```
forall (j in 1..N) do
  data(j):=fun(X(j))
end-do

! Nebenbedingungen
forall (j in 1..N) do
  M*z(j) >= alpha(j)      ! Beschraenktheit/Auswahl der Koeffs
  M*z(j) >= -alpha(j)
  ! Loesen der Approximation
  -data(j)+sum(k in 1..N) (kernmatrix(j,k)*alpha(k)) <= eps
  data(j)-sum(k in 1..N) (kernmatrix(j,k)*alpha(k)) <= eps
  z(j) is_binary        ! z(j) in {0,1}
end-do

! Loesen des Optimierungsproblems
minimize( sum (j in 1..N) (z(j)) )

writeln("Koeffizientenvektor")
forall (j in 1..N) do
  writeln(getsol(alpha(j)))
end-do

writeln("boolsche Variablen")
forall (j in 1..N) do
  write(getsol(z(j))," ")
end-do
writeln("")

writeln("Anzahl Koeffs ungleich Null")
!Ausgeben der Loesung
writeln(sum (j in 1..N)(getsol(z(j))) )
writeln("")

case getprobat of
XPRS_OPT: status:="Optimum found"
XPRS_UNF: status:="Unfinished"
XPRS_INF: status:="Infeasible"
XPRS_UNB: status:="Unbounded"
XPRS_OTH: status:="Failed"
else status:="???"
end-case
writeln("Problem status: ", status)
end-model
```

Symbolverzeichnis

- $[x_0, \dots, x_k]f$ k -te Dividierte Differenz von f an Punkten x_0, \dots, x_k , Seite 34
- δ_{ij} Kronecker-Delta, Seite 9
- $\Gamma(\cdot)$ Gammafunktion, Seite 4
- $\mathcal{N}_K(\Omega)$ Native Space von Ω bezüglich Kern K , Seite 7
- \mathcal{N}_K^* Dualraum zu \mathcal{N}_K , Seite 37
- $\text{span}\{X\}$ Menge aller Linearkombinationen von endlich vielen Elementen aus X , Seite 7
- $B_r(x)$ Kugel um Punkt x mit Radius r , Seite 3
- $\text{dist}(x, Y)$ Abstand von Punkt x zu Menge Y , Seite 3
- $h(\Omega, X)$ Füllabstand von X in Ω , Seite 3
- $L_{X,K}$ Lebesguekonstante zur Punktmenge X und Kern K , Seite 31
- $P_{K,X}$ Powerfunktion bezüglich Kern K und Menge X , Seite 9
- $q(X)$ Separationsabstand der Punktmenge X , Seite 3
- $s_{f,X}$ Interpolante zu f an Punkten aus X , Seite 7
- $U_{X_j,K}$ $\text{span}\{K(\cdot, x_\ell) : x_\ell \in X_j\}$, Seite 32
- $\text{vol}(\Omega)$ Volumen von Ω , Seite 5
- W0 Wendlandfunktion, Seite 61
- W2 Wendlandfunktion, Seite 61

Literaturverzeichnis

- [1] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. 2000.
- [2] St. De Marchi and R. Schaback. Stability of kernel-based interpolation. to appear in *Adv. in Comput. Math.*
- [3] St. De Marchi, R. Schaback, and H. Wendland. Near-optimal data-independent point locations for radial basis function interpolation. *Adv. Comput. Math.*, 23(3):317–330, 2005.
- [4] Y. C. Hon, R. Schaback, and X. Zhou. An adaptive greedy algorithm for solving large RBF collocation problems. *Numer. Algorithms*, 32(1):13–25, 2003.
- [5] H. Meschkowski. *Hilbertsche Räume mit Kernfunktionen*. Springer-Verlag, 1962.
- [6] M. Mouattamid. *Theory of Power Kernels (Dissertation)*. Göttingen, 2005.
- [7] M. Mouattamid and R. Schaback. Recursive Kernels. 2007. preprint.
- [8] St. Müller and R. Schaback. A Newton Basis for Kernel Spaces. to appear in *Journal of Approx. Theory*.
- [9] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. 2006.
- [10] R. Schaback. Creating surfaces from scattered data using radial basis functions. In *Mathematical methods for curves and surfaces (Ulvik, 1994)*, pages 477–496. Vanderbilt Univ. Press, Nashville, TN, 1995.
- [11] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Adv. Comput. Math.*, 3(3):251–264, 1995.
- [12] R. Schaback. Native Hilbert spaces for radial basis functions. I. In *New developments in approximation theory (Dortmund, 1998)*, volume 132 of *Internat. Ser. Numer. Math.*, pages 255–282. Birkhäuser, Basel, 1999.
- [13] R. Schaback. Remarks on meshless local construction of surfaces. In *The mathematics of surfaces, IX (Cambridge, 2000)*, pages 34–58. Springer, London, 2000.

- [14] R. Schaback. A unified theory of radial basis functions. Native Hilbert spaces for radial basis functions. II. *J. Comput. Appl. Math.*, 121(1-2):165–177, 2000. Numerical analysis in the 20th century, Vol. I, Approximation theory.
- [15] R. Schaback. Adaptive numerical solution of MFS systems. *preprint*, 2008.
- [16] R. Schaback. Reconstruction of multivariate functions from scattered data. monograph, 1997.
- [17] R. Schaback and H. Wendland. Adaptive greedy techniques for approximate solution of large RBF systems. *Numer. Algorithms*, 24(3):239–254, 2000.
- [18] R. Schaback and J. Werner. Linearly constrained reconstruction of functions by kernels, with applications to machine learning. *Adv. in Comp. Math.*, 25:237–258, 2006.
- [19] B. Schölkopf and A. J. Smola. *Learning with Kernels*. 2002.
- [20] The MathWorks, Inc. Matlab[©], 7.6.
- [21] V.N. Vapnik. *Statistical Learning Theory*. 1998.
- [22] H. Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.
- [23] J. Werner. *Numerische Mathematik 1*. Vieweg, 1991.
- [24] Xpress-Mosel 64-bit v2.4.0. (c) copyright fair isaac corporation 2008.

Lebenslauf

Name: Müller
Vorname: Stefan
Geburtsdatum: 24.03.1980
Geburtsort: Gießen
Staatsangehörigkeit: deutsch
Familienstand: ledig
Eltern: Marion Müller (geb. Rahyr) und Armin Müller

Schulische Ausbildung

1986-1990 Grundschule, Alten-Buseck
1990-1999 Landgraf-Ludwig-Schule, Gießen
22.06.1999 Abitur

Wissenschaftlicher Bildungsweg

2000-2005 Mathematikstudium an Justus-Liebig-Universität Gießen
16.10.2002 Vordiplom
19.12.2005 Diplom
seit 2006 Doktorand in Graduiertenprogramm der DFG