Diplomarbeit

Reduktion der effektiven Dimension und ihre Anwendung auf hochdimensionale Probleme

angefertigt am
Institut für Numerische Simulation

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Januar 2011

von
Jens Alexander Oettershagen
aus
Königswinter

Zusammenfassung und Danksagung

Zusammenfassung

Auf der *ANOVA-Zerlegung* aufbauend verallgemeinern wir den Begriff der *effektiven Dimension*, so dass konventionelle Definitionen als Spezialfall erhalten bleiben. Damit konstruieren wir ein Funktional auf dem Raum aller Diffeomorphismen, dessen Minimierung die Reduktion der effektiven Dimension reellwertiger Funktionen formalisiert.

Dieses Funktional minimieren wir vermöge eines *geometrischen CG-Verfahrens* über der Mannigfaltigkeit der speziellen orthogonlen Matrizen $\mathbf{SO}(d)$ und der Stiefelmannigfaltigkeit $\mathbf{St}(p, d)$, um die Superpositions- und Trunkationsdimension zu verringern. Wir stellen fest, dass die Lineare Transformation von Imai und Tan einen Spezialfall unserer Methode darstellt.

Vor dem Hintergrund der multivariaten Interpolation, Quadratur und Regression mit *Dünnen Gittern* überprüfen wir das Verfahren anhand von sowohl synthetischen, als auch aus dem Bereich der Finanzmathematik entnommenen Funktionen und Datensätzen.

Abstract

Based on the *ANOVA-Decompositon* we introduce a generalized definition of *effective dimension*, which contains the conventional one as a special case. We construct a functional defined on the space of diffeomorphisms, whose minimization formalizes the reduction of the effective dimension.

Using a geometric CG-Method, we minimize this functional on the manifold of orthogonal matrices $\mathbf{SO}(d)$ and on the Stiefel-manifold $\mathbf{St}(p, d)$ to reduce the superposition- and truncationdimension. We prove that the Linear Transformation introduced by Imai and Tan in 2007 is a special case of our approach.

We examine possible applications of our methods to multivariate quadrature, interpolation and regression with *sparse grids*, using synthetic functions and data sets as well as problems from computational finance and data-mining.

Danksagung

Allen voran danke ich meinen Eltern Evelyne und Konrad für mein Leben, meine Gesundheit und das langjährige Sponsoring meines Mathematikstudiums.

Des Weiteren danke ich Prof. Dr. Michael Griebel und dem Institut für Numerische Simulation der Universität Bonn für die Möglichkeit, mich ausführlich mit diesem interessanten Thema in einer freundlichen und bestens ausgestatteten Arbeitsumgebung auseinanderzusetzen.

Mein ganz besonderer Dank gilt Dr. Markus Holtz, Dr. Christian Feuersänger und Bastian Bohn ohne deren Codes, Kompetenz und Hilfsbereitschaft diese Arbeit niemals in diesem Umfang hätte entstehen können.

Zudem möchte ich mich bei Alexander Hullmann, Jutta Adelsberger, Alexander Rüttgers, Yanick Hamilton, Sandra Winterstein, Ralph Thesen, Christian Kuske, Dr. Jan Hamaekers, Frederik Heber und Ralf Wildenhues für gute Ratschläge, das Korrekturlesen und zahlreiche fruchtbare Diskussionen bedanken.

Notation

Symbol	Bedeutung
$\boldsymbol{a}, \boldsymbol{b}, \ldots, \boldsymbol{z}$	Vektoren
$a, b, \ldots, z$	Skalare
$\boldsymbol{A}, \boldsymbol{B}, \ldots$	Matrizen
$\boldsymbol{S}, \boldsymbol{T}$	schiefsymmetrische Matrizen
$\boldsymbol{Q}, \boldsymbol{R}$	orthogonale Matrizen
$\mathbf{Id}_p$	p-dimensionale Einheitsmatrix
$\boldsymbol{A}_{i,\cdot} / \boldsymbol{A}_{\cdot,i}$	i-te Zeile / Spalte von $\boldsymbol{A}$
P_n, Q_n	Homogenes Polynom vom Grad n
$\mathcal{D}$	die Menge $\{1, 2, \ldots, d\}$
$\mathbf{u}, \mathbf{v}, \mathbf{w}$	Teilmengen von $\mathcal{D}$
$\lvert\mathbf{u}\rvert$	Kardinalität einer Menge
$\mathbf{u} \setminus \mathbf{v}$	Mengendifferenz
$\mathbf{u}^c$	Komplement $\mathcal{D} \setminus \mathbf{u}$
Ω	zusammenhängendes Gebiet in $\mathbb{R}$
$V^{(d)}$	Hilbertraum $\mathcal{L}^2$, d-dimensionale Urbildmenge
$V^{\mathbf{u}}$	Hilbertraum $\mathcal{L}^2$, $\lvert\mathbf{u}\rvert$-dimensionale Urbildmenge
$f, g, F, G \ldots$	reellwertige Funktion
$\mu, \nu, \ldots$	normiertes Produktmaß
φ^d	d-dimensionale Dichtefunktion
η	d-dimensionales Gauß-Maß
λ	Lebesgue-Maß
$\int f \, d\mu$	Erwartungswert von f bezüglich μ
$P_{\mathbf{u}}$	Projektion $V^{(d)} \to V^{\mathbf{u}}$
$f_{\mathbf{u}}, \ldots$	ANOVA-Term von f zu den Richtungen in $\mathbf{u}$
$h_{\mathbf{u}}, \varphi_{\mathbf{u}}$	Test-Funktion aus $V^{\mathbf{u}}$
$Df(\boldsymbol{x})$	Differential einer Abbildung f am Punkt $\boldsymbol{x}$
$\nabla f(\boldsymbol{x})$	Gradient am Punkt $\boldsymbol{x}$
$\mathrm{Hess} f(\boldsymbol{x})$	Hessematrix am Punkt $\boldsymbol{x}$
$\mathbf{SO}(d)$	spezielle orthogonale Gruppe in d Dimensionen
$\mathbf{St}(p, d)$	Stiefelmannigfaltigkeit zum Parameter p in d Dimensionen

Inhaltsverzeichnis

Zusammenfassung und Danksagung iii

Notation v

1 Einleitung 1

2 Die ANOVA-Zerlegung 9
2.1 Dimensionsweise Zerlegungen . . . 10
2.1.1 Die allgemeine ANOVA-Zerlegung . . . 11
2.1.2 Niederdimensionale Bestapproximation . . . 16
2.1.3 Standard- und Anker-ANOVA . . . 20
2.2 Effektive Dimension . . . 24
2.2.1 Superpositions-, Trunkations- und Mittlere Dimension . . . 25
2.2.2 Ein allgemeinerer Begriff von effektiver Dimension . . . 28
2.2.3 Dimensionsweise Zerlegung des numerischen Fehlers . . . 31

3 Verfahren zur Reduktion der effektiven Dimension 33
3.1 Wahl des Koordinatensystems . . . 35
3.1.1 Beliebige Diffeomorphismen . . . 35
3.1.2 Orthogonale Transformationen . . . 40
3.1.3 Nichtlineare Transformationen . . . 42
3.2 Das allgemeine Minimierungsfunktional . . . 46
3.2.1 Vorbereitung . . . 46
3.2.2 Das äquivalente Maximierungsproblem . . . 47
3.2.3 Reduktion der Trunkationsdimension . . . 49
3.3 Optimierung auf der Mannigfaltigkeit $\mathrm{SO}(d)$ . . . 51
3.3.1 Minimierung in Vektorräumen . . . 52
3.3.2 Minimierung auf eingebetteten Untermannigfaltigkeiten . . . 53
3.3.3 Ein CG-Verfahren für $\mathbf{St}(p,d)$ . . . 58
3.3.4 Ein CG-Verfahren für $\mathbf{SO}(d)$ . . . 61
3.3.5 Anwendung auf Eigenwertprobleme . . . 65
3.4 Auswertung der Integrale . . . 68
3.4.1 Berechnung der Sensitivitätskoeffizienten . . . 68
3.4.2 Quasi-Monte Carlo- und Dünngitter Quadratur . . . 69
3.4.3 Verfahren I (Quadraturmethode) . . . 69
3.5 Diskretisierung durch globale Polynome . . . 72

3.5.1 Die homogene Polynombasis . . . 72
3.5.2 Lösung im linearen und quadratischen Fall . . . 75
3.6 Basiswechsel durch Differentiation . . . 76
3.6.1 Die Lineare Transformation nach Imai/Tan . . . 76
3.6.2 Eine Hauptachsentransformation . . . 78
3.7 Basiswechsel durch Projektion . . . 79
3.7.1 Darstellung als lineares Ausgleichsproblem . . . 80
3.7.2 Verfahren II (Polynom-Methode) . . . 80
3.8 Nichtlineare Transformationen . . . 81
3.8.1 Diskretisierung von stückweise-orthogonalen Abbildungen . . . 81

4 Numerische Ergebnisse **83**
4.1 Polynome . . . 83
4.2 Synthetische Testfunktionen . . . 86
4.3 Bewertung von Optionspreisen . . . 91

5 Anwendung auf hochdimensionale Probleme **99**
5.1 Interpolation mit Dünnen Gittern . . . 99
5.1.1 Hierarchische Basis und Dünne Gitter . . . 100
5.1.2 Numerische Versuche . . . 101
5.2 Hochdimensionale Integration . . . 108
5.2.1 Quasi-Monte Carlo Methoden . . . 108
5.2.2 Dünngitter Integration . . . 109
5.2.3 Numerische Versuche . . . 111
5.3 Regression mit Dünnen Gittern . . . 117
5.3.1 Multivariate Regression . . . 117
5.3.2 Numerische Versuche . . . 118

6 Zusammenfassung und Ausblick **123**

Literaturverzeichnis **129**

1 Einleitung

„Um klar zu sehen, genügt ein Wechsel der Blickrichtung." Dieses Zitat von Antoine de Saint-Exupéry[1] ist Ausdruck der in vielen Sprachen verbreiteten Metapher, dass die Komplexität eines Problems oft abhängig von der Perspektive ist, aus der man es betrachtet. Auch in weiten Teilen der Mathematik, Physik und Astronomie, den Ingenieurswissenschaften und anderen Bereichen findet sich dieses Konzept wieder, wenn man je nach Anwendung in andere Koordinatensysteme wechselt, wodurch sich Beschreibung und Lösung von Problemen deutlich vereinfachen können.

Das Auffinden eben dieser problemspezifischen Koordinatensysteme im Kontext dimensionsadaptiver Verfahren ist das zentrale Thema dieser Arbeit.

Im Folgenden geben wir einen kurzen Abriss über die Bedeutung hochdimensionaler Probleme und erläutern die Schwierigkeiten, die bei deren numerischer Behandlung auftreten. Daran anknüpfend folgt ein Überblick über vorhandene Lösungsansätze und die Beschreibung der relevanten Beiträge dieser Arbeit.

Hochdimensionale Probleme

Um unsere hochdimensionale Umwelt, ihre Akteure, deren Interaktionen und daraus resultierende Phänomene adäquat zu beschreiben, sind komplexe Modelle nötig, deren Dimensionalitäten die drei Raumdimensionen oft um ein Vielfaches übersteigen.

Damit aus solchen Modellen verwertbare Erkenntnisse über die Realität gewonnen werden können, muss man beispielsweise die von ihnen abgeleiteten (Integro-) Differentialgleichungen lösen oder Erwartungswerte berechnen. Ist dies analytisch nicht möglich, so geht man zu einem diskreten Modell über, welches als konsistent bezeichnet wird, falls die diskrete Lösung mit steigender Auflösung in Raum (und Zeit) gegen die kontinuierliche Lösung konvergiert.

So muss etwa für die Modellierung und Simulation eines zeitabhängigen Prozesses, wie der Bewegung eines geladenen Teilchens in einem Energiefeld oder der Entwicklung eines Aktienkurses, für jeden in das diskrete Modell eingehenden Zeitpunkt eine eigene Dimensionsrichtung hinzugenommen werden. Für eine gute Annäherung an das kontinuierliche Modell sind daher beispielsweise in der numerischen Finanzmathematik Dimensionszahlen der Größenordnung 1000 keine Seltenheit.

[1] Antoine de Saint-Exupéry (1900-1944) war ein französischer Flieger und Schriftsteller. „Le petit prince" („Der kleine Prinz", 1943) und „Terre des hommes" („Wind, Sand und Sterne", 1939) sind seine wohl bekanntesten Werke.

Problemstellung

Der *Fluch der Dimension* [Bel61] bezeichnet das Phänomen, dass die Kosten, um ein gegebenes numerisches Problem auf einem äquidistanten Gitter bis auf eine Genauigkeit ε zu lösen, exponentiell von der Dimension d abhängen. So begegnet man in der Regel Komplexitäten der Form $\mathcal{O}(\varepsilon^{-d/r})$, wobei $r > 0$ von der Glattheit der betrachteten Funktion und dem Polynomgrad der Ansatzfunktionen abhängt. Somit stößt man schon für moderate Dimensionen von 7 oder 8 an die Grenzen moderner Rechensysteme.

Dieses Problem lässt sich nicht beheben, ohne stärkere Anforderungen, wie etwa eine höhere Regularität, an die betrachtete Funktion zu stellen. Für $r \sim d$ wäre der Fluch der Dimension gebrochen, allerdings ist die Annahme einer isotropen Glattheit für viele interessante Anwendungen unrealistisch.

Betrachtet man jedoch die Quadratur oder Approximation einer d-dimensionalen Funktion $f(\boldsymbol{x}) = \sum_{k=1}^{d} f_k(x_k)$, welche sich aus einer Summe eindimensionaler Funktionen $f_k : \mathbb{R} \to \mathbb{R}$ zusammensetzt, so sieht man, dass sich intrinsisch hochdimensionale Probleme unter Umständen in eine Summe aus niedrigdimensionalen Teilproblemen aufspalten und getrennt behandeln lassen.

Eine weitere Vereinfachung ergibt sich, wenn die einzelnen Koordinatenrichtungen unterschiedlich viel zum Wert von f beitragen. Lassen sich die Dimensionen absteigend in ihrer Wichtigkeit anordnen, so kann es bei Inkaufnahme eines geringen Fehlers möglich sein, höhere Dimensionen zu vernachlässigen.

Dies führt unmittelbar auf das Konzept der *effektiven Dimension*, welches 1997 in [CMO97] eingeführt wurde und eine Quantifizierung des Effektes in obigen Beispielen liefert. Die effektive Dimension einer Funktion f ist also ein Maß dafür, wie gut sie sich als Summe niederdimensionaler Funktionen darstellen lässt und inwiefern man hohe Dimensionsrichtungen vernachlässigen kann.

Von diesem Effekt profitieren insbesondere moderne Verfahren, wie die *dimensionsadaptiven Dünnen Gitter*, da diese in der Lage sind, während des Approximationsprozesses relevante Richtungen zu erkennen und entsprechend ihrem Beitrag zum Gesamtwert zu verfeinern. Auch der Erfolg von *Quasi-Monte Carlo Methoden* zur Quadratur hochdimensionaler Funktionen aus dem Finanzbereich lässt sich durch eine geringe effektive Dimension der Integranden erklären[2]. Denn diese sind auf dem Pfad eines stochastischen Prozesses definiert, dessen Kovarianzen – in einer geeigneten Multiskalenbasis (wie etwa der Brownschen Brücke oder der Karhunen-Loève Konstruktion) diskretisiert – mit steigender Auflösung des Pfades exponentiell gegen 0 gehen.

In dieser Arbeit wollen wir uns mit der Frage auseinandersetzen, ob auch für Funktionen, die intrinsisch keine niedrige effektive Dimension besitzen, die Möglichkeit besteht, sie in numerisch günstigeren Koordinaten zu betrachten.

Einen ersten, eher heuristischen Ansatz haben wir bereits erwähnt. Handelt es sich um eine Funktion, die auf dem Pfad eines stochastischen Prozesses W defininiert ist (etwa die Auszah-

[2] Außerdem konnte in [GKS10] gezeigt werden, dass dieser niederdimensionale Anteil des Integranden oftmals glatter ist als die Funktion selbst, wovon Quasi-Monte Carlo ebenfalls deutlich profitiert.

lungsfunktion eines Aktienderivates oder das Potential eines geladenen Teilchens), so erzielt man durch eine geeignete hierarchische Diskretisierung von W bereits einen starken Abfall in der Wichtigkeit der Dimensionen. Dieser Zugang über die Struktur des zugrundeliegenden Prozesses W bringt für die meisten Funktionen von Interesse zwar durchaus eine Verringerung der effektiven Dimension mit sich, ist jedoch in den Bereich der Heuristik einzuordnen, da er die konkrete Struktur der Funktion selbst außer Acht lässt.

An dieser Stelle wollen wir nun mit dieser Diplomarbeit ansetzen und eine allgemeine Theorie zur Reduktion der effektiven Dimension reellwertiger Funktionen erarbeiten.

Lösungsansätze

Unser wichtigstes Werkzeug bei der Untersuchung hochdimensionaler Probleme ist die *ANOVA-Zerlegung (Analysis of Variance)*

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i=1}^{d} f_i(x_i) + \sum_{1\le i<j\le d} f_{ij}(x_i, x_j) + \ldots + f_{123\ldots d}(\boldsymbol{x}) \\ &= \sum_{\mathbf{u}\subseteq\{1,\ldots,d\}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}), \end{aligned}$$

welche wir daher im ersten Kapitel dieser Arbeit in einer allgemeinen Formulierung über Produktmaße einführen und ausführlich diskutieren werden. Wir zeigen, dass die abgeschnittene ANOVA-Reihe im Hilbertraum $\mathcal{L}^2$ die beste niederdimensionale Approximation darstellt, also den Anteil von f, der wirklich niederdimensional ist, auch vollständig erfasst und somit das geeignete Werkzeug zur Untersuchung verborgener Niederdimensionalitäten darstellt.

Damit können wir auch eine konsistente Verallgemeinerung der ANOVA-Zerlegung für Maße angeben, die nicht zwingend Produktstruktur besitzen müssen, indem wir die ANOVA-Terme durch eine Folge orthogonaler $\mathcal{L}^2$-Projektionen definieren, welche für Produktmaße wieder den üblichen Integralen entsprechen.

Um nun zu beschreiben, wie groß der niederdimensionale Anteil von f ist, dient der bereits erwähnte Begriff der effektiven Dimension. Wir diskutieren verschiedene auf der $\mathcal{L}^2$-Norm basierende Varianten aus der Literatur und verallgemeinern die Definition dann derart, dass auch die Fehlerabschätzungen diverser hochdimensionaler Verfahren, welchen komplexere Normen zugrunde liegen, als ein Maß für effektive Dimension interpretiert werden können.

Nachdem wir also eine Quantifizierung für die „Schwierigkeit“ multivariater Probleme gefunden haben, müssen wir uns fragen, welche Möglichkeiten es gibt, diese zu reduzieren. Dazu wollen wir – wie bereits eingangs erwähnt – in ein dem konkreten Problem angepasstes, möglichst günstiges Koordinatensystem wechseln, dessen Perspektive eine einfachere Lösung gestattet. Solche Koordinatensysteme werden durch Diffeomorphismen, also stetig-differenzierbare Bijektionen, beschrieben.

Auf $\mathbf{Diff}(\Omega^{(d)})$, der Menge aller Diffeomorphismen des zugrundeliegenden Gebietes $\Omega^{(d)}$ auf sich selbst, definieren wir dann ein Funktional $\mathfrak{M}_f : \mathbf{Diff}(\Omega^{(d)}) \to \mathbb{R}$

$$\mathfrak{M}_f(\phi) = \sum_{\mathbf{u} \subseteq \{1,\dots,d\}} \gamma_{\mathbf{u}} \, \|(f \circ \phi)_{\mathbf{u}}\|_*, \tag{1.1}$$

dessen Minimierung für geeignete Dimensionsgewichte $\gamma_{\mathbf{u}}$ und Normen $\|\cdot\|_*$ die Reduktion der effektiven Dimension von f formalisiert.

Dabei ergeben sich jedoch zwei wesentliche Schwierigkeiten:

- Die Auswertung des Funktionals $\mathfrak{M}$ erfordert die Berechnung aller 2^d ANOVA-Terme von $f \circ \phi$, was offensichtlich wieder auf den Fluch der Dimension führt.
- Die Diskretisierung eines beliebigen d-dimensionalen Diffeomorphismus beinhaltet ebenfalls wieder den Fluch der Dimension.

Das erste Problem lösen wir, indem wir ein Funktional herleiten, das unter gewissen Voraussetzungen die gleichen kritischen Punkte wie $\mathfrak{M}$ besitzt, aber bei Inkaufnahme eines geringen Fehlers das Vernachlässigen der ANOVA-Terme höherer Ordnung gestattet.

Die auftretenden Integrale bei der Berechung der Normen der ANOVA-Terme von $f \circ \phi$ berechnen wir effizient mit Quasi-Monte Carlo und Dünngitter-Verfahren.

Da sich das zweite Problem nicht ohne weiteres beheben lässt, ist es notwendig, sich auf solche Teilmengen $\Phi \subset \mathbf{Diff}(\Omega^{(d)})$ einzuschränken, die eine möglichst einfache Darstellung besitzen, deren Elemente in der Zahl der Freiheitsgrade also nicht exponentiell von d abhängig sind.

Geeignete Diffeomorphismen

Ein erster Ansatz besteht darin, lediglich komponentenweise abzubilden, wie es etwa bei der Transformation von Gauß-Integralen vom $\mathbb{R}^d$ auf den Einheitswürfel $[0,1]^d$ vermöge der inversen Normalverteilung üblich ist. Anhand von zwei Beispielen werden wir aufzeigen, dass dieser Ansatz durchaus Potential bietet, allerdings wollen wir den Schwerpunkt dieser Arbeit auf lineare Diffeomorphismen legen – insbesondere auf Drehungen des Koordinatensystems.

Diese werden gerade durch die Matrizen der *speziellen Orthogonalen Gruppe* $\mathbf{SO}(d)$ definiert, die sich mit der Struktur einer differenzierbaren Mannigfaltigkeit versehen lässt und als Teilmenge von $\mathbb{R}^{d\times d}$ lediglich $\mathcal{O}(d^2)$ Freiheitgrade besitzt.

Möchte man die Trunkationsdimension von f minimieren, ist es sogar möglich, sich auf die ersten p Spalten einer orthogonalen Matrix einzuschränken. Auch diese lassen sich als differenzierbare Mannigfaltigkeit auffassen, welche als *Stiefel-Mannigfaltigkeit* $\mathbf{St}(p,d)$ bezeichnet wird und in unserem Verfahren nur $\mathcal{O}(d)$ Freiheitsgrade benötigen wird.

Eine Verallgemeinerung des Konzeptes der orthogonalen Abbildungen stellen die *stückweise orthogonalen Transformationen* dar. Diese basieren auf der Idee, ein rotationssymmetrisches Gebiet in disjunkte Kreisringe zu unterteilen und auf jedem dieser Ringe eine eigene orthogonale Abbildung zu definieren. Lässt man den Radius dieser Kreisringe gegen Null gehen, so kann

man eine überall differenzierbare Bijektion konstruieren, welche im Gegensatz zu einer einzelnen orthogonalen Matrix nicht nur das gesamte Koordinatensystem dreht, sondern einzelne Achsen auch krümmen kann. Zudem erhalten diese Abbildungen die Produktstruktur des Gauß-Maßes, und ihre Komplexität wächst ebenfalls nur quadratisch in d.

Optimierung auf Mannigfaltigkeiten

Um nun das Funktional $\mathfrak{M}_f$ über $\mathbf{SO}(d)$ oder $\mathbf{St}(p,d)$ zu minimieren, existieren verschiedene Ansätze. Der konventionelle Zugang zu auf Mannigfaltigkeiten definierten Optimierungsproblemen besteht darin, die Mannigfaltigkeit durch Nebenbedingungen zu beschreiben und das zugehörige Lagrange-Problem zu lösen. Dies führt in unserem Falle jedoch auf ein d^2-dimensionales Problem mit nichtlinearen Nebenbedingungen.

Wir wollen eine andere Herangehensweise wählen und die Mannigfaltigkeit durch das Konzept der *Retraktion* lokal über ihrem Tangentialraum parametrisieren, welcher in unserem Falle nur $d(d-1)/2$-dimensional ist, und dadurch die Konzepte aus der Minimierung über Vektorräumen auf Mannigfaltigkeiten übertragen.

Dazu wählen wir den Zugang von [AMS08] und definieren geeignete Retraktionen zum einen durch die QR-Zerlegung und zum anderen über das Matrix-Exponential $\exp : \mathfrak{so}(d) \to \mathbf{SO}(d)$, wobei wir ausnutzen, dass $\mathbf{SO}(d)$ eine Lie-Guppe mit zugehöriger Lie-Algebra $\mathfrak{so}(d)$ ist. Da die Berechnung des Matrix-Exponentials im Allgemeinen numerisch sehr aufwändig ist, verwenden wir eine Padé-Approximation, also eine Darstellung als rationale Funktion auf $\mathfrak{so}(d)$.

Damit können wir nichtlineare CG-Verfahren für die Mannigfaltigkeiten $\mathbf{SO}(d)$ und $\mathbf{St}(p,d)$ angeben, welche für die Optimierung unseres Kostenfunktionals $\mathfrak{M}_f$ aus (1.1) geeignet sind, da sie nur wenige der teuren Funktionalauswertungen benötigen.

Diskretisierung durch homogene Polynome

Ein wesentliches Problem des bisher beschriebenen Verfahrens besteht darin, dass für jede Auswertung von $\mathfrak{M}_f$ an einem Punkt auf $\mathbf{SO}(d)$ die zugrundeliegende Funktion f erneut diskretisiert werden muss. Dies rührt daher, dass die Dünngitterbasis und die Quasi-Monte Carlo Punktfolgen, welche wir zur Diskretisierung der auftretenden Integrale verwenden, nicht rotationsinvariant sind.

Wir benötigen also Basisfunktionen, die nach einer Drehung des Koordinatensystems noch immer im Span dieser Basis liegen. Hier bietet sich die Basis der homogenen Polynome zum Grad n an

$$\mathcal{P}_n := \{\boldsymbol{x}^{\boldsymbol{\alpha}} : |\boldsymbol{\alpha}|_1 \leq n\},$$

welche man durch ein an das Prinzip der dünnen Gitter angelehntes Tensorprodukt der eindimensionalen Monombasen erhält.

In dieser Basis können wir die analytischen Lösungen der bei jeder Punktauswertung von $\mathfrak{M}_f$ auftretenden Integrale herleiten, wodurch sich das Verfahren deutlich beschleunigen lässt.

Vollzieht man die Projektion von $\mathcal{L}^2$ in den Raum $\mathcal{P}_n$ durch die Taylorreihe (also durch Differentiation) der Funktion, so erhält man für den Spezialfall $n = 1$ die Lineare Transformation (LT) [IT06] und für $n = 2$ die Diagonal Methode (DM) [Mor98].

Da die Taylorreihe jedoch im Allgemeinen eher schlechte Approximationseigenschaften besitzt – insbesondere dann, wenn f nicht hinreichend oft differenzierbar ist – wollen wir die orthogonale Projektion des Hilbertraumes $\mathcal{L}^2$ verwenden, also das Minimierungsproblem

$$\arg\min_{P_n \in \mathcal{P}_n} \int_{\Omega^{(d)}} (f(\boldsymbol{x}) - P_n(\boldsymbol{x}))^2 \, d\boldsymbol{x} \tag{1.2}$$

lösen.

Dazu diskretisieren wir das Integral in (1.2) durch dünne Gitter und leiten daraus ein gewichtetes lineares Least-Squares Problem her, das wir mit Hilfe der QR-Zerlegung effizient lösen können.

Anwendung auf hochdimensionale Probleme und numerische Ergebnisse

Anhand von verschiedenen, sowohl synthetischen als auch aus dem Bereich der Finanzmathematik, entnommenen Modellfunktionen weisen wir die Relevanz der von uns entwickelten Methoden nach. Dabei beschränken wir uns im Fall der Quadratur der Einfachheit halber auf Anwendungen, denen das Black-Scholes Modell, bzw. das Gauß-Maß zugrunde liegt. Wir weisen jedoch darauf hin, dass sich die Resultate genau so auf Modelle, die auf allgemeinen Lévy-Prozessen basieren, übertragen lassen [IT].

Erstmalig wenden wir die in dieser Arbeit vorgestellen Konzepte auch auf den Bereich der Interpolation an und zeigen anhand verschiedener Modellfunktionen, dass das orts- und dimensionsadaptive Dünngitter Verfahren aus [Feu10] substanziell von einer reduzierten effektiven Dimension profitieren kann.

Auch für die multivariate Regression demonstrieren wir den Einfluss der effektiven Dimension des Datensatzes auf die Erkennungsrate des in [Gar04, Boh10] beschriebenen Verfahrens.

Eigene Beiträge

Die wesentlichen Beiträge dieser Arbeit wollen wir an dieser Stelle kurz zusammenfassen:

- Beweis der Charakterisierung der ANOVA-Zerlegung durch eine niederdimensionale Best-Approximations-Eigenschaft in der Theorie des Hilbertraumes $\mathcal{L}^2$, welche letztlich auch ANOVA-Zerlegungen ohne Produkt-Maße ermöglicht.
- Entwicklung eines allgemeinen Begriffes von effektiver Dimension, der Superpositions-, Trunkations- und Mittlere Dimension als Spezialfall enthält und einen direkten Zusammenhang zum numerischen Fehler hochdimensionaler Verfahren, wie der Quasi-Monte Carlo Integration ermöglicht.

- Formalisierung der Reduktion von effektiver Dimension durch ein Funktional auf der Gruppe der Diffeomorphismen.
- Einführung *stückweiser orthogonaler Transformationen* – einer Klasse von nichtlinearen Diffeomorphismen, deren Freiheitgrade nur quadratisch von der Dimension abhängen.
- *Reduktion der Trunkationsdimension* reellwertiger Funktionen vermöge eines aus der Literatur entnommenen nichtlinearen CG-Verfahrens auf $\mathbf{St}(p, d)$.
- Entwicklung eines eigenen Verfahrens zur Minimierung von Funktionalen auf der Mannigfaltigkeit $\mathbf{SO}(d)$ und seine Verwendung, um erstmalig die *Superpositionsdimension* reellwertiger Funktionen zu reduzieren.
- Vereinfachte Auswertung des Minimierungsfunktionals $\mathfrak{M}$ durch die Verwendung homogener Polynome und die Realisierung des dazu nötigen Projektionsoperators durch dünne Gitter.
- Einbettung vorhandener Theorie in diese Arbeit: Die *Lineare Transformation* (LT) und die *Diagonalmethode* (DM) stellen einen Spezialfall unseres Ansatzes dar. Ferner beweisen wir neuartige Eigenschaften dieser beiden Verfahren.

Aufbau der Arbeit

Kapitel 2 liefert das wichtigste Werkzeug für unsere weiteren Betrachtungen – die ANOVA-Zerlegung. Wir führen diese in einer allgemeinen Form ein und konstruieren daraus die verschiedenen Begriffe der effektiven Dimension.

Kapitel 3 beschäftigt sich mit der Reduktion der effektiven Dimension. Dies wird durch ein Funktional $\mathfrak{M}$ auf den Mannigfaltigkeiten $\mathbf{SO}(d)$ und $\mathbf{St}(p, d)$ formalisiert. Für die Minimierung von $\mathfrak{M}$ wird ein geometrisches CG-Verfahren entwickelt.

Kapitel 4 untersucht sowohl synthetische als auch aus der Praxis entnommene Modellfunktionen hinsichtlich ihrer effektiven Dimension und der Möglichkeit, diese zu reduzieren.

Kapitel 5 demonstriert, dass hochdimensionale Verfahren von unseren Methoden profitieren können. Speziell betrachten wir die Interpolation mit Dünnen Gittern, die Quadratur mit Quasi-Monte Carlo Verfahren und dünnen Gittern sowie Dünngitter-Regressionsverfahren.

Kapitel 6 fasst die wesentlichen Ergebnisse dieser Arbeit zusammen und gibt einen Ausblick auf Erweiterungs- und Vertiefungsmöglichkeiten des Themas.

2 Die ANOVA-Zerlegung

In diesem Kapitel werden wir die *ANOVA-Zerlegung* reellwertiger Funktionen und die daraus konstruierten Begriffe von *effektiver Dimension* einführen, diskutieren und erweitern. Wir werden einen Bezug zu Fehlerabschätzungen von *Dünngitter-* und *Quasi-Monte Carlo*-Methoden herstellen und somit den theoretischen Grundstein für die weiteren Kapitel dieser Arbeit legen.

Bei der ANOVA-Zerlegung handelt es sich um eine Zerlegung einer auf einem Gebiet $\Omega^{(d)} \subset \mathbb{R}^d$ definierten quadratintegrierbaren Funktion $f : \Omega^d \to \mathbb{R}$ in eine Summe von insgesamt 2^d niederdimensionalen Funktionen $f_{\mathbf{u}} : \Omega^{\mathbf{u}} \to \mathbb{R}$ – den so genannten ANOVA-Termen zu Richtungen $\mathbf{u} \subseteq \{1, \ldots, d\}$. Bezogen auf Modelle der realen Welt kann man sich die ANOVA-Terme als die Interaktion der verschiedenen Parameter des Modelles miteinander vorstellen.

Das Konzept von effektiver Dimension wird nun durch die Beobachtung motiviert, dass reale Vorgänge durch eine Vielzahl dieser Parametern modelliert werden müssen, um die intrinsisch hochdimensionale Welt in der wir leben hinreichend genau zu approximieren, jedoch längst nicht alle dieser Parameter gleichermaßen stark miteinander interagieren. Dies spiegelt sich auch in der in diesen Modellen verwendeten Mathematik wieder, denn in vielen Anwendungen lassen sich die auftretenden Modellfunktionen bereits durch eine geringe Zahl der oben erwähnten ANOVA-Terme hinreichend genau approximieren. Ein Überblick verschiedener Disziplinen findet sich in [Gri06].

Die effektive Dimension sollte also in möglichst kompakter und einfacher Form wiedergeben, welche ANOVA-Terme relevant sind und welche man bei Inkaufnahme eines geringen Fehlers vernachlässigen kann – sie gibt die „effektive Niederdimensionalität“ des Problemes wieder.

Der weitere Aufbau dieses Kapitels gestaltet sich nun folgendermaßen: Im ersten Teil werden wir die ANOVA-Zerlegung in einem allgemein gehaltenen Kontext einführen und ausführlich diskutieren, da es sich bei ihr um das Schlüsselwerkzeug beim Verständnis aller weiteren Überlegungen handelt.

Als Beispiele für konkrete ANOVA-Zerlegungen betrachten wir die Standard- und die Anker-ANOVA und führen einige ihrer wichtigen Eigenschaften auf. Dazu gehört unter anderem ein neuartiger Beweis, dass die abgeschnittene Summe der Standard-ANOVA bezüglich der $\mathcal{L}^2$-Norm die beste niederdimensionale Approximation an f darstellt, also wirklich alle niederdimensionalen Aspekte von f auch tatsächlich erfasst. Damit lässt sich die Definition der ANOVA-Zerlegung für Produktmaße konsistent auf beliebige normierte Maße verallgemeinern.

Im zweiten Abschnitt dieses Kapitels werden wir dann den Begriff der *effektiven Dimension* einführen. Dazu diskutieren wir zunächst die konventionellen Begriffe von Superpositions-, Trunkations- und Mittlerer Dimension und stellen neuartige Zusammenhänge zwischen ihnen

her. Danach werden wir einen verallgemeinerten Begriff definieren, welcher sowohl die verschiedenen herkömmlichen Definitionen, also auch die Fehlerabschätzung für die Quasi-Monte Carlo Integration als Spezialfall enthält und damit einen direkten Bezug zur numerischen Praxis gestattet.

2.1 Dimensionsweise Zerlegungen

In diesem Abschnitt werden wir eine Klasse von Zerlegungen einführen, welche eine gegebene Funktion $f \in \mathcal{L}^2(\Omega^{(d)})$ in insgesamt 2^d niederdimensionale Funktionen $f_{\mathbf{u}}$, welche nur noch auf den $|\mathbf{u}|$-dimensionalen Mengen $\Omega^{\mathbf{u}}$ definiert sind, zerlegt. Diese Zerlegungen werden wir durch eine Formulierung als orthogonale Projektion auf den Unterraum $\mathcal{L}(\Omega^{\mathbf{u}})$ definieren und zeigen, dass diese Definition für Produktmaße mit der in der Literatur gebräuchlichen Definition (etwa [Hol08, Gri06, Sob01]) übereinstimmt.

Zur Definition einiger grundlegenden Begriffe wollen wir uns am Vorgehen von [Hol08] orientieren.

Für alle $j = 1, 2, \ldots, d$ seien zusammenhängende Mengen $\Omega_j \subseteq \mathbb{R}$ gegeben. Für alle Teilmengen $\mathbf{u} \subseteq \mathcal{D} := \{1, 2, \ldots, d\}$ definieren wir damit

$$\Omega^{\mathbf{u}} := \bigotimes_{j \in \mathbf{u}} \Omega_j \ \subseteq \mathbb{R}^{|\mathbf{u}|}$$

als die $|\mathbf{u}|$-dimensionale Produktmenge der Ω_j. Zur Vereinfachung der Notation schreiben wir $\Omega^{(d)} := \Omega^{\{1..d\}}$.

Sind Wahrscheinlichkeitsmaße μ_j (d.h. $\mu_j(\Omega_j) = 1$) auf den Borelmengen von Ω_j gegeben, so erhalten wir mit

$$d\mu = \prod_{j=1}^{d} d\mu_j(x_j)$$

ein d-dimensionales Produktmaß auf $\Omega^{(d)}$ für das $\mu(\Omega^{\mathbf{u}}) = 1$ gilt.

$V^{(d)} := \mathcal{L}^2(\Omega^{(d)}, \mu)$ sei der Hilbertraum aller bezüglich μ quadratintegrierbaren Funktionen $f : \Omega^{(d)} \to \mathbb{R}$, mit dem Skalarprodukt

$$(f, g)_\mu := \int_{\Omega^{(d)}} f(\boldsymbol{x}) g(\boldsymbol{x}) \, d\mu(\boldsymbol{x})$$

und der durch dieses induzierten Norm

$$\|f\|_{2,\mu} = \sqrt{(f, f)_\mu} = \sqrt{\int_{\Omega^{(d)}} f(\boldsymbol{x})^2 \, d\mu(\boldsymbol{x})}.$$

Entsprechend seien für $\mathbf{u} \subset \mathcal{D}$ die Räume $V^{\mathbf{u}} := \mathcal{L}^2(\Omega^{\mathbf{u}}, \mu)$ definiert, welche wir als Unterraum von $V^{(d)}$ auffassen wollen, indem wir die Elemente von $V^{\mathbf{u}}$ als d-dimensionale Funktionen

betrachten, die nur von den Variablen in $\mathbf{u}$ abhängen.
(Formal: $V^{\mathbf{u}} := \{f \in V^{(d)} : f(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{y}_{\mathbf{u}^c}) = f(\boldsymbol{x}_{\mathbf{u}}, \tilde{\boldsymbol{y}}_{\mathbf{u}^c}) \text{ für alle } \boldsymbol{y}_{\mathbf{u}^c}, \tilde{\boldsymbol{y}}_{\mathbf{u}^c} \in \Omega^{\mathbf{u}^c}\}$)

2.1.1 Die allgemeine ANOVA-Zerlegung

Wir führen die ANOVA-Zerlegung für normierte Produktmaße ein. Ein noch allgemeinerer Zugang findet sich in [KSWW08].

Projektion durch Integration

Die Maße $\mu_{\mathbf{u}}$ auf $\Omega^{\mathbf{u}}$ definieren nun Projektionen $P_{\mathbf{u}} : V^{(d)} \to V^{\mathbf{u}}$

$$P^{\mu}_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) := \int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x})\, d\mu_{\mathbf{u}^c}(\boldsymbol{x}) \qquad \text{für } \mathbf{u} \subsetneq \mathcal{D} \tag{2.1}$$

$$P^{\mu}_{\mathbf{u}}(f)(\boldsymbol{x}) := f(\boldsymbol{x}) \qquad \text{für } \mathbf{u} = \mathcal{D}, \tag{2.2}$$

wobei wir mit $\boldsymbol{x}_{\mathbf{u}}$ den $|\mathbf{u}|$-dimensionalen Vektor bezeichnen, der gerade die Komponenten von $\boldsymbol{x}$ enthält, deren Indizes in $\mathbf{u}$ liegen. Entsprechend ist dann $d\,\mu_{\mathbf{u}^c}(\boldsymbol{x}) := \prod_{j \notin \mathbf{u}} d\,\mu_j(x_j)$.

Diese Projektion wollen wir durch eine gewisse „Minimaleigenschaft“ motivieren: Wir suchen eine niederdimensionale Approximation an f durch eine Funktion $\varphi_{\mathbf{u}} : \Omega^{\mathbf{u}} \to \mathbb{R}$, mit $|\mathbf{u}| < d$, welche unter allen Funktionen aus $V^{\mathbf{u}}$ den $\mathcal{L}^2$-Abstand zu f über $\Omega^{(d)}$ minimiert. Da $V^{(d)}$ ein Hilbertraum und $V^{\mathbf{u}}$ ein Unterraum von diesem ist, ist die Lösung des Optimierungsproblemes

$$\arg\min_{\varphi_{\mathbf{u}} \in V^{\mathbf{u}}} \|f - \varphi_{\mathbf{u}}\|_{2,\mu}$$

äquivalent zur Bestimmung der orthogonalen Projektion von f auf $V^{\mathbf{u}}$, was wir im folgenden Lemma beweisen wollen.

Lemma 2.1. ($\mathcal{L}^2$ - Optimalität orthogonaler Projektionen)

Für einen Hilbertraum $V^{(d)}$ und eine Projektion $P : V^{(d)} \to V^{\mathbf{u}}$ in einen Unterraum $V^{\mathbf{u}} \subset V^{(d)}$ gilt:

Ist P orthogonal, d.h.

$$(f - P(f), h)_{\mu} = 0 \text{ für alle } h \in V^{\mathbf{u}}, \tag{2.3}$$

so minimiert P den Abstand zu f bezüglich der durch das Skalarprodukt $(\cdot,\cdot)_{\mu}$ induzierten $\mathcal{L}^2$-Norm, also

$$\arg\min_{\varphi \in V^{\mathbf{u}}} \|f - \varphi\|_{2,\mu} = P(f)$$

Beweis. Wir zeigen dass es sich bei $P(f)$ tatsächlich um ein Minimum handelt, indem wir für beliebige Testfunktionen $h \in V^{\mathbf{u}}$ beweisen, dass

$$\|f - P(f) + h\|_{2,\mu} \geq \|f - P(f)\|_{2,\mu}.$$

gilt.

$$\begin{aligned}\|f - P(f) + h\|_{2,\mu}^2 &= (f - P(f) + h, f - P(f) + h)_\mu \\ &= (f - P(f), f - P(f))_\mu + (h, h)_\mu + 2\underbrace{(f - P(f), h)_\mu}_{=0\,(2.3)} \\ &= \|f - P(f)\|_{2,\mu}^2 + \|h\|_{2,\mu} \\ &\geq \|f - P(f)\|_{2,\mu}^2\end{aligned}$$

□

Damit können wir nun die folgende Aussage beweisen.

Satz 2.2.

Für $\mathbf{u} \subseteq \mathcal{D}$ gilt:

$$\arg\min_{\varphi_\mathbf{u} \in V^\mathbf{u}} \int_{\Omega^{(d)}} (f(\boldsymbol{x}) - \varphi_\mathbf{u}(\boldsymbol{x}_\mathbf{u}))^2 \, d\mu(\boldsymbol{x}) = P_\mathbf{u}^\mu(f)(\boldsymbol{x}_\mathbf{u}),$$

Beweis. Nach Lemma 2.1 genügt es zu zeigen, dass $(f - P_\mathbf{u}^\mu(f), h)_\mu = 0$ für alle $h \in V^\mathbf{u}$ gilt.

$$\begin{aligned}&\int_{\Omega^{(d)}} (f(\boldsymbol{x}) - P_\mathbf{u}^\mu(f)(\boldsymbol{x}_\mathbf{u}))\, h(\boldsymbol{x}_\mathbf{u})\, d\mu(\boldsymbol{x}) \\ = &\int_{\Omega^\mathbf{u}} \left(\int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x})\, d\mu_{\mathbf{u}^c}(\boldsymbol{x}) \right) h(\boldsymbol{x}_\mathbf{u}) - P_\mathbf{u}^\mu(f)(\boldsymbol{x}_\mathbf{u})\; h(\boldsymbol{x}_\mathbf{u})\, d\mu_\mathbf{u}(\boldsymbol{x}) \\ = &\int_{\Omega^\mathbf{u}} P_\mathbf{u}^\mu(f)(\boldsymbol{x}_\mathbf{u})\, h(\boldsymbol{x}_\mathbf{u}) - P_\mathbf{u}^\mu(f)(\boldsymbol{x}_\mathbf{u})\, h(\boldsymbol{x}_\mathbf{u})\, d\mu_\mathbf{u}(\boldsymbol{x}) \\ = &\; 0.\end{aligned}$$

□

Dimensionweise Zerlegung

Wir wollen nun eine Funktion $f \in \mathcal{L}^2(\Omega^{(d)})$ in eine Summe niederdimensionaler Funktionen $f_\mathbf{u} \in \mathcal{L}^2(\Omega^\mathbf{u})$ zerlegen, indem wir sie vermöge der oben definierten Projektionen $P_\mathbf{u}^\mu$ in die Funktionenräume $V^\mathbf{u}$ projizieren und dort nach dem Teleskopsummen-Prinzip niederdimensionale Anteile wieder abziehen.

Der folgende Satz liefert zwei in der Literatur gebräuchliche zueinander äquivalente Charakterisierungen dieser Zerlegung.

Satz 2.3. (Zerlegung multivariater Funktionen)

Für alle Teilmengen $\mathbf{u} \subseteq \mathcal{D} := \{1, \ldots, d\}$ seien Funktionen $f_\mathbf{u} : \Omega^\mathbf{u} \to \mathbb{R}$ gegeben. Dann ist für jedes normierte Produktmaß $\mu = \otimes \mu_i$ äquivalent:

i) Die $f_{\mathbf{u}}$ erfüllen

$$\sum_{\mathbf{u}\subseteq\mathcal{D}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = f(\boldsymbol{x}) \text{ für alle } \boldsymbol{x} \in \Omega^{(d)}, \tag{2.4}$$

d.h. sie stellen eine *additive Zerlegung* von f über $\Omega^{(d)}$ dar
und
für jedes $f_{\mathbf{u}}$ mit $|\mathbf{u}| \geq 1$ gilt

$$\int_{\Omega_i} f_{\mathbf{u}}(\boldsymbol{x})\, d\mu_{\{i\}} = 0 \text{ für alle } i \in \mathbf{u}. \tag{2.5}$$

ii)

$$f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = P^{\mu}_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}) \text{ für alle } \boldsymbol{x}_{\mathbf{u}} \in \Omega^{\mathbf{u}} \tag{2.6}$$

Beweis. $(i) \Rightarrow (ii)$: Es gilt

$$\begin{aligned}
f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) &= \int_{\Omega^{\mathbf{u}^c}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})\, d\mu_{\mathbf{u}^c}(\boldsymbol{x}) \\
&\overset{(2.4)}{=} \int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x}) - \sum_{\mathbf{v}\neq\mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}})\, d\mu_{\mathbf{u}^c}(\boldsymbol{x}) \\
&= P_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}) - \sum_{\mathbf{v}\not\subseteq\mathbf{u}} \underbrace{\int_{\Omega^{\mathbf{u}^c}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}})\, d\mu_{\mathbf{u}^c}(\boldsymbol{x})}_{=0 \ (2.5)} \\
&= P_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}).
\end{aligned}$$

$(ii) \Rightarrow (i)$: Wir führen Induktion über die Kardinalität von $\mathbf{u}$.
Sei $|\mathbf{u}| = 1$, etwa $\mathbf{u} = \{k\}$. Dann ist nach (ii) $f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = P_{\{k\}}(f)(x_k) - f_{\emptyset}$. Damit folgt

$$\begin{aligned}
\int_{\Omega_k} f_{\mathbf{u}}(\boldsymbol{x})\, d\mu_{\{k\}}(\boldsymbol{x}) &= \int_{\Omega_k} P_{\{k\}}(f)(x_k) - f_{\emptyset}\, d\mu_{\{k\}}(\boldsymbol{x}) \\
&= \int_{\Omega_k} \int_{\Omega^{\{k\}^c}} f(\boldsymbol{x})\, d\mu_{\{k\}^c}(\boldsymbol{x})\, d\mu_{\{k\}}(\boldsymbol{x}) - f_{\emptyset} \\
&= f_{\emptyset} - f_{\emptyset} \\
&= 0
\end{aligned}$$

Sei die Behauptung nun für alle $\mathbf{v} \subsetneq \mathbf{u}$ bewiesen. Dann gilt für alle $i \in \mathbf{u}$:

$$\int_{\Omega_i} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})\, d\mu_{\{i\}}(\boldsymbol{x}) = \int_{\Omega_i} P_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}})\, d\mu_{\{i\}}(\boldsymbol{x})$$

$$
\begin{aligned}
&= \int_{\Omega_i} P_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}})\, d\mu_{\{i\}}(\boldsymbol{x}) - \sum_{\mathbf{v}\subseteq\mathbf{u}-\{i\}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}) - \underbrace{\int_{\Omega_i} \sum_{\substack{\mathbf{v}\subseteq\mathbf{u}\\ i\in\mathbf{u}}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}})\, d\mu_{\{i\}}(\boldsymbol{x})}_{=0} \\
&= \int_{\Omega_i} P_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}})\, d\mu_{\{i\}}(\boldsymbol{x}) - \sum_{\mathbf{v}\subseteq\mathbf{u}\setminus\{i\}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}) \\
&= P_{\mathbf{u}\setminus\{i\}}(f)(\boldsymbol{x}) - \sum_{\mathbf{v}\subseteq\mathbf{u}\setminus\{i\}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}) \\
&= P_{\mathbf{u}\setminus\{i\}}(f)(\boldsymbol{x}) - P_{\mathbf{u}-\{i\}}(f)(\boldsymbol{x}) \\
&= 0
\end{aligned}
$$

Die Eigenschaft der additiven Zerlegung (2.4) ergibt sich direkt aus der Definition der Projektion $P_{\mathcal{D}}$ in (2.1).

□

Definition 2.4. (Die ANOVA-Zerlegung)

Zu einer Funktion $f : \Omega^{(d)} \to \mathbb{R}$ seien für jedes $\mathbf{u} \subseteq \mathcal{D}$ Funktionen $f_{\mathbf{u}} : V^{\mathbf{u}} \to \mathbb{R}$ gegeben. Erfüllen diese eine der äquivalenten Bedingungen (i) oder (ii), so heißen die $f^{\mu}_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})$ *ANOVA-Terme* von f bezüglich des Maßes μ über $\Omega^{(d)}$.

Die Menge aller 2^d ANOVA-Terme heißt *ANOVA-Zerlegung* von f.

Eine in der Praxis nützliche Darstellung der rekursiven Formel (2.6) liefert der folgende Satz aus [KSWW08]. Dort findet sich auch der Beweis.

Satz 2.5. (Direkte Darstellung der ANOVA-Terme)

Die rekursiv definierten ANOVA-Terme lassen sich mittels

$$f^{\mu}_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = \sum_{\mathbf{v}\subseteq\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} P^{\mu}_{\mathbf{v}}(f)(\boldsymbol{x}_{\mathbf{v}})$$

berechnen.

Orthogonalität

Die ANOVA-Zerlegung besitzt die praktische Eigenschaft, dass sie bezüglich des $\mathcal{L}^2(\Omega^{(d)}, \mu)$ Skalarproduktes im folgenden Sinne *orthogonal* ist:

Satz 2.6. (Orthogonalität der ANOVA-Zerlegung)

Für alle ANOVA-Terme $f_{\mathbf{u}}$ und $f_{\mathbf{v}}$ mit $\mathbf{u} \neq \mathbf{v}$ gilt:

$$(f^{\mu}_{\mathbf{u}}, f^{\mu}_{\mathbf{v}})^{\mu} = \int_{\Omega^d} f^{\mu}_{\mathbf{u}}(\boldsymbol{x})\; f^{\mu}_{\mathbf{v}}(\boldsymbol{x})\, d\mu(\boldsymbol{x}) = 0. \tag{2.7}$$

Beweis. Die Behauptung folgt direkt aus (2.5) in Satz 2.3, denn für $\mathbf{u} \neq \mathbf{v}$ gibt es immer mindestens ein $k \in \mathbf{u} \cup \mathbf{v}$, so dass $k \notin \mathbf{u} \cap \mathbf{v}$. □

Zerlegung der Varianz

Genau wie sich die Funktion f aus den ANOVA-Termen $f_{\mathbf{u}}$ additiv zusammensetzt, lässt sich, wie wir in Satz 2.9 sehen werden, auch die Varianz $\sigma^2(f)$ durch die Summe der Varianzen der einzelnen ANOVA-Terme $\sigma^2_{\mathbf{u}}(f)$ berechnen.

Definition 2.7. (Varianz)

Die *Varianz* einer Funktion ist definiert als die Summe der quadrierten Abweichungen von ihrem Mittelwert,

$$\sigma^2_\mu(f) := \int_{\Omega^{(d)}} \left(f(\boldsymbol{x}) - \int_{\Omega^{(d)}} f(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \right)^2 d\mu(\boldsymbol{x}).$$

Nach dem Verschiebungssatz läßt sich die Varianz auch durch

$$\sigma^2_\mu(f) = \int_{\Omega^{(d)}} (f(\boldsymbol{x}))^2 \, d\mu(\boldsymbol{x}) - \left(\int_{\Omega^{(d)}} f(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \right)^2$$

berechnen.

Definition 2.8. (Varianz eines ANOVA-Termes)

Analog zur Definition der Varianz von f sei für $\mathbf{u} \subseteq \mathcal{D}$ die *Varianz eines ANOVA-Terms* von f als

$$\begin{aligned} \sigma^2_{\mathbf{u},\mu}(f) &:= \int_{\Omega^{\mathbf{u}}} f^\mu_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})^2 \, d\mu_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) - \underbrace{\left(\int_{\Omega^{\mathbf{u}}} f^\mu_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})\, d\mu_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) \right)^2}_{=0 \ (2.5)} \\ &= \int_{\Omega^{\mathbf{u}}} f^\mu_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})^2 \, d\mu_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) \end{aligned}$$

definiert.

Aufgrund der Orthogonalität der ANOVA-Zerlegung lässt sich die Varianz einer Funktion f bezüglich des Maßes μ durch die Summe der Varianzen ihrer ANOVA-Terme darstellen.

Satz 2.9. (Zerlegung der Varianz)

Es gilt:

$$\sigma^2_\mu(f) = \sum_{|\mathbf{u}|>0} \sigma^2_{\mathbf{u},\mu}. \tag{2.8}$$

Beweis.

$$\begin{aligned}
\sigma_\mu^2(f) &= \int_{\Omega^d} f(\boldsymbol{x})^2 \, d\mu(\boldsymbol{x}) - \left(\int_{\Omega^d} f(\boldsymbol{x}) \, d\mu(\boldsymbol{x}) \right)^2 \\
&= \int_{\Omega^d} \left(\sum_{|\mathbf{u}|\geq 0} f_{\mathbf{u}}^\mu(\boldsymbol{x}_\mathbf{u}) \right) \left(\sum_{|\mathbf{u}|\geq 0} f_{\mathbf{u}}^\mu(\boldsymbol{x}_\mathbf{u}) \right) d\mu(\boldsymbol{x}) - f_\emptyset^2 \\
&\overset{(2.7)}{=} \sum_{|\mathbf{u}|>0} \int_{\Omega^d} f_{\mathbf{u}}^\mu(\boldsymbol{x}_\mathbf{u})^2 \, d\boldsymbol{x} \\
&= \sum_{|\mathbf{u}|>0} \sigma_{\mathbf{u},\mu}^2
\end{aligned}$$

□

2.1.2 Niederdimensionale Bestapproximation

In diesem Abschnitt werden wir nun darlegen, weshalb die ANOVA-Zerlegung das geeignete Werkzeug ist, um die verborgene Niederdimensionalität hochdimensionaler Funktionen zu untersuchen.

Dabei stellt der folgende Satz sicher, dass die ANVOA-Zerlegung zu einem *beliebigen* Maß μ niemals „hochdimensionaler“ ist, als die ursprüngliche Funktion selbst – eine effektiv niederdimensionale Funktion bleibt also unabhängig vom gewählten Maß niederdimensional.

Satz 2.10. (Minimaleigenschaft der ANOVA-Zerlegung)

Eine Funktion $f : \Omega^{(d)} \to \mathbb{R}$ lasse sich durch eine Summe niederdimensionaler Funktionen $g_\mathbf{u}$ mit $|\mathbf{u}| \leq k$ darstellen, d.h. es existiert eine additive Zerlegung $f = \sum_{\mathbf{u}\in\mathcal{D}} g_\mathbf{u}$ mit $g_\mathbf{u} \equiv 0$ für alle $\mathbf{u}$ mit $|\mathbf{u}| > k$. Dann gilt auch für jede ANOVA-Zerlegung zu beliebigem Wahrscheinlichkeits-Maß μ: $f_\mathbf{u}^\mu \equiv 0$ für alle $|\mathbf{u}| > k$.

Beweis. Dies ist gerade Theorem 6 aus [KSWW08] □

Nun stellt sich die Frage, welche ANOVA-Zerlegung die bestmögliche niederdimensionale Approximation (im $\mathcal{L}^2$-Sinne) darstellt, also den größtmöglichen Varianzanteil bereits mit den Termen kleiner Ordnung erfasst.

Optimale niederdimensionale Approximation in $\mathcal{L}^2$

Im Folgenden werden wir eine $\mathcal{L}^2$-Optimalitätseigenschaft zeigen, welcher die Aussage von Satz 2.2 verallgemeinert. Sei dazu $\mathcal{C} \subseteq \mathcal{P}(\mathcal{D})$ eine Teilmenge der Potenzmenge von $\{1, \ldots, d\}$, welche die Zulässigkeitsbedingung

$$\mathbf{u} \in \mathcal{C} \text{ und } \mathbf{v} \subseteq \mathbf{u} \Rightarrow \mathbf{v} \in \mathcal{C} \tag{2.9}$$

erfüllt. Wir definieren

$$V^{\mathcal{C}} := \{f \in V^{(d)} : f = \sum_{\mathbf{u}\in\mathcal{C}} \varphi_{\mathbf{u}}, \text{ mit } \varphi_{\mathbf{u}} \in V^{\mathbf{u}}\}$$

als den Raum aller Funktionen f, die sich aus einer Summe niederdimensionaler Funktionen $\varphi_{\mathbf{u}} \in V^{\mathbf{u}}$, mit Mengenindex $\mathbf{u} \in \mathcal{C}$, zusammensetzen.

Im nachfolgenden Satz werden wir zeigen, dass die ANOVA-Zerlegung die optimale Projektion auf $V^{\mathcal{C}}$ darstellt. Ähnliche Resultate finden sich in [Hoo07] und [RA99].

Doch zuvor wollen wir uns die Aussage noch anhand eines motivierenden Beispiels vor Augen führen:

Beispiel 2.1: (Approximation durch die Summe eindimensionaler Funktionen)

Sei etwa $\mathcal{C} = \{\emptyset, \{1\}, \{2\}, \ldots, \{d\}\}$. Dann ist nach dem nachfolgenden Satz die Funktion

$$f_C(\boldsymbol{x}) := \sum_{|\mathbf{u}|\leq 1} f^{\mu}_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}).$$

die Bestapproximation an f durch eine Summe eindimensionaler Funktionen bezüglich der durch μ induzierten $\mathcal{L}^2$-Norm.

Satz 2.11. (Niederdimensionale Bestapproximation im $\mathcal{L}^2$-Sinne)

Sei $\sum_{\mathbf{u}} f^{\mu}_{\mathbf{u}} = f$ die ANOVA-Zerlegung einer Funktion $f \in \mathcal{L}^2(\Omega^{(d)}, \mu)$. Unter allen Funktionen der Form $\varphi(\boldsymbol{x}) := \sum_{\mathbf{u}\in\mathcal{C}} \varphi_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) \in V^{\mathcal{C}}$, minimiert

$$f_{\mathcal{C}}(\boldsymbol{x}) := \sum_{\mathbf{u}\in\mathcal{C}} f^{\mu}_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})$$

den $\mathcal{L}^2(\Omega^{(d)}, \mu)$-Abstand zu f über dem Gebiet $\Omega^{(d)}$, d.h. die Lösung von

$$\arg\min_{\varphi\in V^{\mathcal{C}}} \|f - \varphi\|_{2,\mu} \tag{2.10}$$

ist gerade $f_{\mathcal{C}}$.

Beweis. Wir gehen ähnlich vor, wie beim Beweis von Satz 2.2, das heißt wir zeigen, dass $(f - f_{\mathcal{C}}, h)_{\mu} = 0$ für alle $h \in V^{\mathcal{C}}$ gilt, woraus dann mit Lemma 2.1 die Bestapproximation im Raum $\mathcal{L}^2(\Omega^{(d)}, \mu)$ folgt.

Dazu betrachten wir eine Teilmenge der Potenzmenge $\mathcal{C} \subset \mathcal{P}(\{1, \ldots, d\})$, welche (2.9) erfüllt, anstatt Testfunktionen $h \in V^{\mathcal{C}}$ aus Gründen der Lesbarkeit jedoch nur $h_{\mathbf{w}} \in V^{\mathbf{w}}$, mit $\mathbf{w} \in \mathcal{C}$. Aufgrund der Linearität des Skalarproduktes folgt dann die eigentliche Behauptung.

Es gilt

$$\begin{aligned}
(f - f_{\mathcal{C}}, h_{\mathbf{w}})_\mu &= \int_{\Omega^{(d)}} (f(\boldsymbol{x}) - f_{\mathcal{C}}(\boldsymbol{x}))\ h_{\mathbf{w}}(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \\
&= \int_{\Omega^{(d)}} f(\boldsymbol{x}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) - \sum_{\mathbf{u}\in\mathcal{C}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}})\ d\mu(\boldsymbol{x}) \\
&= \int_{\Omega^{(d)}} f(\boldsymbol{x}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) - f_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) - \sum_{\mathbf{u}\in\mathcal{C}\setminus\mathbf{w}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}})\ d\mu(\boldsymbol{x}) \\
&\overset{(2.6)}{=} \int_{\Omega^{(d)}} f(\boldsymbol{x}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) - P_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) - \sum_{\substack{\mathbf{u}\in\mathcal{C}\setminus\mathbf{w} \\ \mathbf{u}\not\subseteq\mathbf{w}}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}})\, d\mu(\boldsymbol{x}) \\
&= \int_{\Omega^{\mathbf{w}}} P_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) - P_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}})\ d\mu_{\mathbf{w}}(\boldsymbol{x}) \\
&\quad - \int_{\Omega^{(d)}} \sum_{\substack{\mathbf{u}\in\mathcal{C}\setminus\mathbf{w} \\ \mathbf{u}\not\subseteq\mathbf{w}}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}})\, d\mu(\boldsymbol{x}) \\
&= 0 - \int_{\Omega^{(d)}} \sum_{\substack{\mathbf{u}\in\mathcal{C}\setminus\mathbf{w} \\ \mathbf{u}\not\subseteq\mathbf{w}}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}})\ d\mu(\boldsymbol{x}).
\end{aligned}$$

Aus $\mathbf{u} \not\subseteq \mathbf{w}$ folgt nun $\mathbf{u} \cap \mathbf{w}^c \neq \emptyset$ und damit aufgrund von (2.5) aus Satz 2.3

$$\int_{\Omega^{(d)}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) h_{\mathbf{w}}(\boldsymbol{x}_{\mathbf{w}})\, d\mu(\boldsymbol{x}) = 0 \text{ für alle } \mathbf{u} \in (\mathcal{C} \setminus \mathbf{w}) \cap \{\mathbf{v} : \mathbf{v} \not\subseteq \mathbf{w}\}.$$

□

Ist μ absolut-stetig bezüglich des Lebesgue-Maßes λ^d, so gilt auch die Umkehrung, was wir im folgenden Korollar festhalten wollen.

Korollar 2.12. (Verschiedene Charakterisierungen der ANOVA-Zerlegung)

Für zum Lebesgue-Maß absolutstetige, normierte Produktmaße $\mu = \otimes\mu_i$ sind die folgenden Aussagen äquivalent:

i) Die $f_{\mathbf{u}}$ erfüllen

$$\sum_{\mathbf{u}\subseteq\{1..d\}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = f(\boldsymbol{x}) \text{ für alle } \boldsymbol{x} \in \Omega^{(d)}, \tag{2.11}$$

und
für jedes $f_{\mathbf{u}}$ mit $|\mathbf{u}| \geq 1$ gilt

$$\int_{\Omega_i} f_{\mathbf{u}}(\boldsymbol{x})\, d\mu_{\{i\}} = 0 \text{ für alle } i \in \mathbf{u} \tag{2.12}$$

ii) Es gilt

$$f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = P^{\mu}_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}) \text{ für alle } \boldsymbol{x}_{\mathbf{u}} \in \Omega^{\mathbf{u}} \tag{2.13}$$

iii) Es gilt

$$f_{\mathbf{u}} = \operatorname*{arg\,min}_{\varphi_{\mathbf{u}}\in V^{\mathbf{u}}} \left\| \left(f - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}} \right) - \varphi_{\mathbf{u}} \right\|_{2,\mu} \quad \text{für alle } \mathbf{u} \subseteq \mathcal{D}, \tag{2.14}$$

wobei die Definition dieses Optimierungsproblemes rekursiv zu verstehen ist.

Beweis. $(i) \Leftrightarrow (ii)$ ist gerade Satz 2.3 und $(ii) \Rightarrow (iii)$ ist die Aussage von Satz 2.11. Somit ist nur noch $(iii) \Rightarrow (i)$ zu zeigen.

Die Eigenschaft der additiven Zerlegung (2.11) ist offensichtlich, da die Minimierung (2.14) für $\mathbf{u} = \mathcal{D}$ im Raum $V^{\{1..d\}}$ stattfindet – unabhängig von $f_{\mathbf{u}}$ mit $|\mathbf{u}| < d$ wird im letzten Schritt also (2.11) sichergestellt.

Die Identität (2.12) wollen wir mittels Induktion zeigen, wozu wir zunächst für $\varphi_{\mathbf{u}} \in V^{\mathbf{u}}$, $\mathbf{u} \subseteq \mathcal{D}$ die Folge der Funktionale

$$\begin{aligned} F_{\mathbf{u}}(\varphi_{\mathbf{u}}) &:= \|((f - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}) - \varphi_{\mathbf{u}})^2\|^2_{2,\mu} \\ &= \int_{\Omega^{(d)}} ((f - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}) - \varphi_{\mathbf{u}})^2 \, d\mu \end{aligned} \tag{2.15}$$

definieren, wobei die $f_{\mathbf{v}}$ (rekursiv) für $\mathbf{v} \subsetneq \mathbf{u}$ die Lösung von $\operatorname{argmin}_{\varphi_{\mathbf{v}}} F_{\mathbf{v}}(\varphi_{\mathbf{v}})$ bezeichnen.

An einem kritischen Punkt von $F_{\mathbf{u}}$ muss nun das Gateaux-Differential identisch Null sein, also

$$\frac{\delta F_{\mathbf{u}}}{\delta \varphi_{\mathbf{u}}} = \int_{\Omega^{(d)}} -2(f - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}) + 2\varphi_{\mathbf{u}} \, d\mu = 0.$$

Damit ergibt sich als notwendige Bedingung für ein Minimum von $F(\varphi_{\mathbf{u}})$

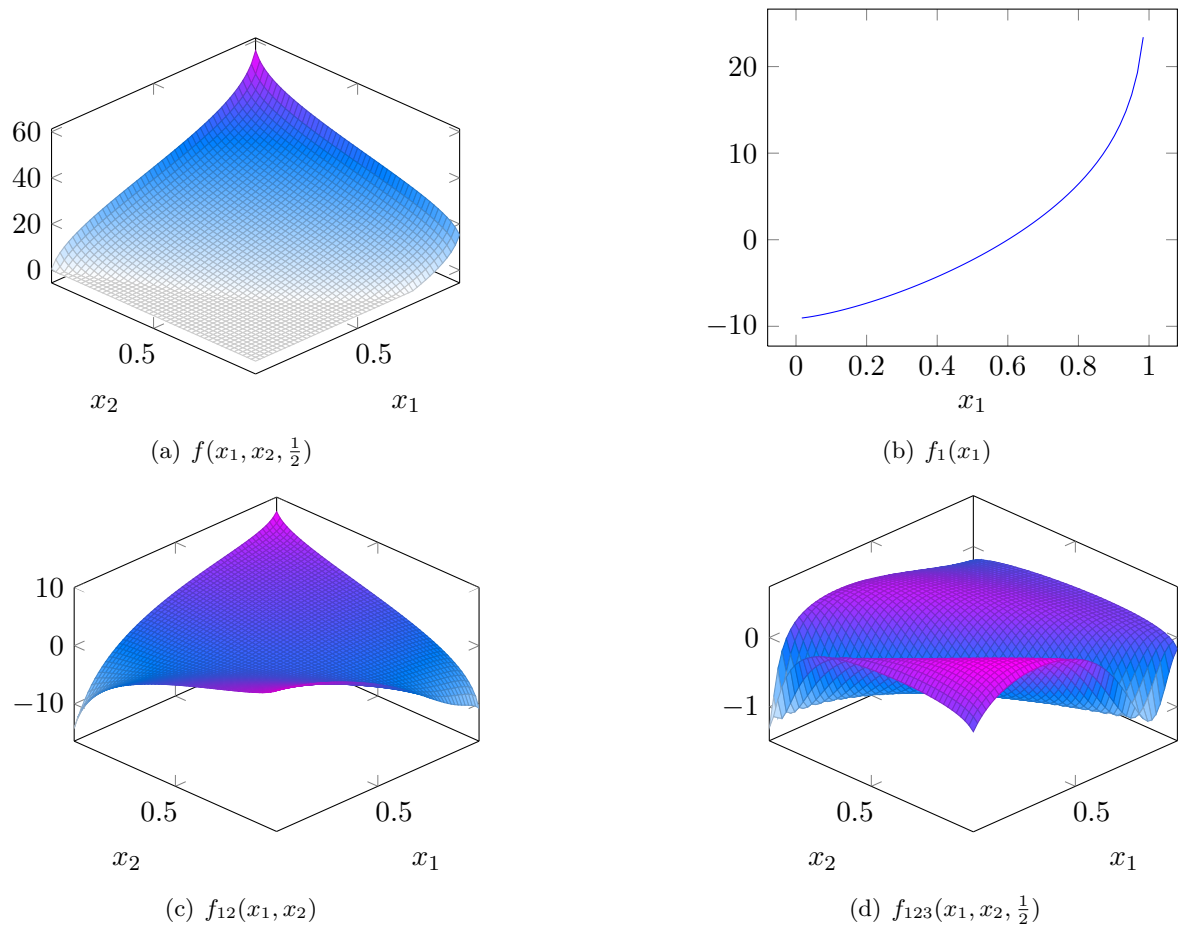
$$\begin{aligned} & \int_{\Omega^{(d)}} f - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}} \, d\mu && = \int_{\Omega^{(d)}} \varphi_{\mathbf{u}} \, d\mu \\ \Leftrightarrow \quad & \int_{\Omega^{(d)}} f \, d\mu - \sum_{\substack{\mathbf{v}\subsetneq\mathbf{u} \\ |\mathbf{u}|>0}} \int_{\Omega^{\mathbf{v}}} f_{\mathbf{v}} \, d\mu_{\mathbf{v}} - f_{\emptyset} && = \int_{\Omega^{\mathbf{u}}} \varphi_{\mathbf{u}} \, d\mu_{\mathbf{u}}. \end{aligned}$$

Wegen

$$\operatorname*{arg\,min}_{\varphi_{\emptyset}\in\mathbb{R}} F_{\emptyset}(\varphi_{\emptyset}) = \int_{\Omega^{(d)}} f \, d\mu = f_{\emptyset}$$

ergibt sich nun (2.12) für alle $|\mathbf{u}| > 0$ durch Induktion über die Kardinalität von $\mathbf{u}$.

□



(a) $f(x_1, x_2, \frac{1}{2})$ (b) $f_1(x_1)$

(c) $f_{12}(x_1, x_2)$ (d) $f_{123}(x_1, x_2, \frac{1}{2})$

Abb. 2.1: Ausgewählte Lebesgue-ANOVA Terme einer dreidimensionalen Asiatischen Option

Bemerkung: (ANOVA-Zerlegung ohne Produktmaß)

Die Definition der ANOVA-Zerlegung nach (2.14) ist unabhängig davon, ob μ ein Produktmaß ist. Nach Korollar 2.12 stimmt diese Definition also für Produktmaße mit der konventionellen Definition überein.

2.1.3 Standard- und Anker-ANOVA

Als konkrete Beispiele für die ANOVA-Zerlegung mit allgemeinem Wahrscheinlichkeitsmaß μ als Parameter werden wir in diesem Abschnitt nun zwei prominente Zerlegungen, die Standard- und die Anker-ANOVA Zerlegung, betrachten.

Den Unterschied zwischen beiden werden wir anhand einer Beispielfunktion aus $C^0([0,1]^3)$ veranschaulichen und dabei feststellen, dass die Anker-ANOVA bezüglich der Standard-$\mathcal{L}^2$-Norm zum einen schlechtere Approximationseigenschaften aufweist (was nach Satz 2.11 ja auch zu erwarten ist) und dass ihre ANOVA-Terme weniger glatt sind als die der Standard-ANOVA.

Die Standard-ANOVA Zerlegung

Die *Standard-ANOVA*-Zerlegung (oder auch *Lebesgue-ANOVA)* erhält man, wenn als Maß μ für die Projektionen (2.1), bzw. für die Minimierung (2.10), das Lebesgue-Maß (oder eines das zu diesem absolutstetig ist) gewählt wird.

Mit $\mu = \lambda^d$ erhalten wir den Projektor

$$P^{\lambda}_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) = \int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x})\, d\boldsymbol{x}_{\mathbf{u}^c}, \tag{2.16}$$

bzw. für absolut-stetiges $\mu_i = \varphi_i \lambda$, $\mu_{\mathbf{u}} = \varphi^{\mathbf{u}} \lambda^{\mathbf{u}}$

$$P^{\mu}_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) = \int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x})\ \varphi^{\mathbf{u}^c}(\boldsymbol{x})\, d\boldsymbol{x}_{\mathbf{u}^c}. \tag{2.17}$$

Kennt man die inverse Verteilungsfunktion von μ, so kann man aufgrund des Transformationssatzes die Berechnung der ANOVA-Terme stets auf ein Integrationsproblem im Einheitswürfel $[0,1]^d$ zurückführen. (Siehe Satz 3.3).

Wie wir in Satz 2.11 gezeigt haben, stellt diese ANOVA-Zerlegung bezüglich der Standard-$\mathcal{L}^2$-Norm $\|f\|_2 = \sqrt{\int_{\Omega^{(d)}} f(\boldsymbol{x})^2\, d\boldsymbol{x}}$ die bestmögliche niederdimensionale additive Approximation dar, ist dafür jedoch auch ausgesprochen teuer zu berechnen.

Eine andere interessante Eigenschaft der Standard-ANOVA ist, dass ihre niederdimensionalen Terme oftmals glatter sind, als f selbst. Deutlich ist dies in Abbildung 2.1 zu sehen. Denn während die Auszahlungsfunktion einer dreidimensionalen arithmetischen asiatischen Option (mehr dazu in Kapitel 4) entlang einer $(d-1)$-dimensionalen Hyperebene nicht differenzierbar ist („einen Knick hat“), sind alle ein- und zweidimensionalen ANOVA-Terme von f mindestens einmal stetig-differenzierbar, da durch das Anwenden des Projektors $P^{\lambda}_{\mathbf{u}}$ die Regularität erhöht wird. Mehr zum Glättungseffekt der ANOVA, seinen Vorraussetzungen und Details findet sich in [GKS10].

Die Anker-ANOVA Zerlegung

Wählt man für die Projektionen (2.1) μ als das Diracmaß $\delta_{\boldsymbol{a}}$ an einem Punkt $\boldsymbol{a} \in \Omega^{(d)}$, so erhält man die so genannte *Anker-ANOVA* Zerlegung von f. Diese spielt im Gegensatz zur relativ teuren Lebesgue-ANOVA (2.16) auch in der numerischen Praxis eine Rolle, da sie wesentlich einfacher und mit geringen Kosten zu berechnen ist. Auch die *Dünnen Gitter*, mit welchen wir uns später noch eingehender beschäftigen werden, sind der Spezialfall einer *diskretisierten Anker-ANOVA Zerlegung*. (Mehr dazu findet sich in [Feu10, GH10b, Hol08].)

Ausserdem wollen wir an dieser Stelle darauf hinweisen, das wir Experimente (welche nicht Teil dieser Arbeit sind) durchgeführt haben, die nahelegen, dass die Approximationsgüte der Anker-ANOVA Zerlegung hochgradig abhängig von der Wahl ihrer Verankerung im Koordinatensystem ist. Eine Auseinandersetzung mit dieser Thematik im Zummenhang mit der

Dünngitter-Integration findet sich in [GH10a] und im Kontext von High-Dimensional-Modell-Representations (HDMR) in [Wan10].

Wir definieren den $|\mathbf{u}|$-dimensionalen Projektor also als

$$\begin{aligned} P_{\mathbf{u}}^{\delta_{\boldsymbol{a}}}(f)(\boldsymbol{x}_{\mathbf{u}}) &= \int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x})\, d\delta_{\mathbf{u}}^{\boldsymbol{a}}(\boldsymbol{x}) \\ &= f\left(\boldsymbol{x}\right)_{|\boldsymbol{x}_{\mathbf{u}^c}=\boldsymbol{a}_{\mathbf{u}^c}}, \end{aligned} \tag{2.18}$$

wobei wir mit $f\left(\boldsymbol{x}\right)_{|\boldsymbol{x}_{\mathbf{u}^c}=\boldsymbol{a}_{\mathbf{u}^c}}$ die Auswertung am Punkt $\boldsymbol{x}$, dessen in $\mathbf{u}^c$ enthaltenen Komponenten durch $\boldsymbol{a}$ ersetzt wurden, bezeichnen. Zwischen der Anker- und der Standard-ANOVA gibt es den Zusammenhang

$$f_{\mathbf{u}}^{\lambda}(\boldsymbol{x}_{\mathbf{u}}) = \int_{\Omega^{\mathbf{u}}} f_{\mathbf{u}}^{\delta_{\boldsymbol{a}}}(\boldsymbol{x}_{\mathbf{u}})\, d\boldsymbol{a}_{\mathbf{u}^c},$$

d.h. die Standard-ANOVA stellt eine Mittelung über alle möglichen Anker-ANOVA Zerlegungen dar. Die Identität (2.5) wird für das Dirac-Maß zu

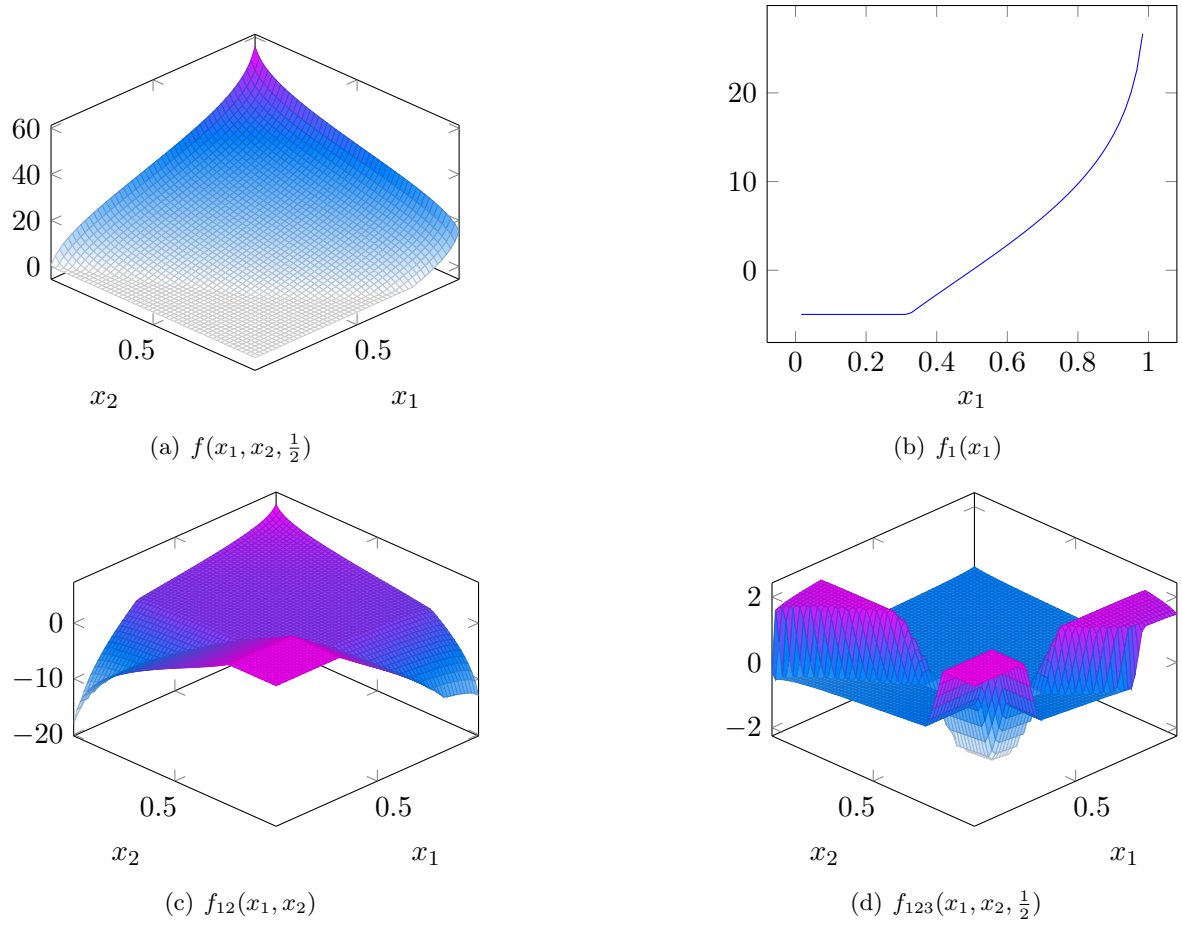
$$f_{\mathbf{u}}^{\delta_{\boldsymbol{a}}}(\boldsymbol{x}_{\mathbf{u}})_{x_i=a_i} = 0 \text{ für } i \in \mathbf{u} \tag{2.19}$$

woraus sich ergibt, dass in allen affinen Unterräumen der Dimension k, welche $\boldsymbol{a}$ enthalten,

$$f_k^{\delta}(\boldsymbol{x}) := \sum_{|\mathbf{u}|\leq k} f_{\mathbf{u}}^{\delta_{\boldsymbol{a}}}(\boldsymbol{x}_{\mathbf{u}}) = f(\boldsymbol{x})$$

gilt. Das heißt, f_k^{δ} ist zwar nicht die beste niederdimensionale Approximation im $\mathcal{L}^2$-Sinne, dafür aber in besagten k-dimensionalen affinen Unterräumen interpolierend.

In Abbildung 2.2 haben wir einige Terme der Anker-ANOVA einer arithmetischen Asiatischen Option mit $\boldsymbol{a} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ geplottet. Es ist deutlich zu sehen, dass im Gegensatz zur Standard-ANOVA die höheren Terme sogar weniger Regularität besitzen als f selbst.



(a) $f(x_1, x_2, \frac{1}{2})$

(b) $f_1(x_1)$

(c) $f_{12}(x_1, x_2)$

(d) $f_{123}(x_1, x_2, \frac{1}{2})$

Abb. 2.2: Ausgewählte Anker-ANOVA Terme einer dreidimensionalen Asiatischen Option

2.2 Effektive Dimension

Der Begriff der *effektiven Dimension* ist eine Sammlung von Konzepten, welche unseres Wissens erstmals in [CMO97], [Owe03] und [Hol08] erscheinen. Sie basieren auf der grundlegenden Erkenntnis, dass für viele Funkionen die ANOVA-Terme hoher Kardinalität deutlich weniger zur Gesamtvarianz beitragen als die niedrigen Terme – und dass Methoden zur Lösung hochdimensionaler Probleme von diesem Effekt profitieren können. Eine Liste mit Beispielen aus verschiedenen Bereichen findet sich etwa in [Gri06].

Die Motivation hinter den verschiedenen Definitionen zur effektiven Dimension ist es nun, diese Approximationsgüte der niederdimensionalen Terme in einer einzigen Zahl zusammenzufassen. In [CMO97] wird zur Definition der *Superpositionsdimension* und der *Trunkationsdimension* die Menge aller Terme zusammengefasst, die benötigt werden, um f bis auf einen gegebenen Fehler in der $\mathcal{L}^2$-Norm anzunähren. In [WF03] wird nachgewiesen, dass eine geringe effektive Dimension die Konvergenzeigeschaften von Quasi-Monte Carlo Methoden substanziell verbessern kann.

In [Owe03] werden zur Definition der *Mittleren Dimension* die Varianzen $\sigma^2_{\mathbf{u}}$ der einzelnen ANOVA-Terme mit ihrer Kardinalität $|\mathbf{u}|$ multipliziert - hohe Terme also mit großem Gewicht und niedrige Terme mit einem kleinen. Auch hierfür lässt sich zeigen, dass eine geringe mittlere Dimension mit einer guten Rate der Quasi-Monte Carlo Methoden korreliert.

[GH10b] ist der erste uns bekannte Versuch, einen Begriff von effektiver Dimension zu schaffen, der nicht auf der $\mathcal{L}^2$-Norm aufbaut, sondern die $\mathcal{L}^1$-Norm zugrunde legt, wodurch ein direkter Zusammenhang zum Integrationsfehler der *Dünnen Gitter* und der *Dimension-wise Integration* [Hol08, GH10b] geschaffen wird.

Aus Gründen der Vollständigkeit sei noch ein anderer Zugang zum Integrationsfehler erwähnt, welcher in dieser Arbeit jedoch nicht weiter verfolgt werden soll. [SW97] und [SWW04] betrachten *Reproduzierende Kern-Hilberträume* und geben Kerne an, die einen Abfall in der Gewichtung der Dimensionen definieren, wodurch sich für Funktionen aus diesen Räumen spezielle Quasi-Monte Carlo Verfahren konstruieren lassen, die stets optimale Konvergenz $\mathcal{O}(N^{-1})$ besitzen. Ein ähnlicher Ansatz findet sich in [Hol08] und [Feu10] wo für vorgegebene Funktionenräume Dünne Gitter konstruiert werden, die ein optimales Kosten-Nutzen-Verhältnis aufweisen.

In diesem Abschnitt werden wir nun zunächst die bereits erwähnten Begriffe von *effektiver Dimension* aus der Litaratur [CMO97, Owe03, Hol08] einführen und diese dann später auf sinnvolle Art und Weise verallgemeinern.

Dabei wollen wir die prinzipielle Abhängigkeit der ANOVA-Zerlegung vom Maß μ im Hinterkopf behalten, aber aus Gründen der Lesbarkeit fortan nur noch bei Bedarf erwähnen und ansonsten einfach

$$f(\boldsymbol{x}) = \sum_{\mathbf{u} \subseteq \mathcal{D}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}),$$

beziehungsweise

$$f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = P_{\mathbf{u}} f(\boldsymbol{x}_{\mathbf{u}}) - \sum_{\mathbf{v} \subsetneq \mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}})$$

schreiben.

Sofern nicht anders angegeben, meinen wir damit stets die Zerlegung bezüglich des Lebesgue-Maßes oder einem, das absolut-stetig zu diesem ist (zum Beispiel das Gauß-Maß auf dem $\mathbb{R}^d$).

Definition 2.13. (Sensitivitätskoeffizienten)

Aufgrund der Zerlegung (2.8) ist nach [Sob01] die Definition der sogenannten *Sensitivitäts-Koeffizienten*

$$s_{\mathbf{u}} := \frac{\sigma_{\mathbf{u}}^2}{\sigma^2}$$

sinnvoll.

Sie beschreiben den relativen Anteil der Varianz von $f_{\mathbf{u}}$ an der Gesamtvarianz von f, denn wegen (2.8) gilt

$$\sum_{|\mathbf{u}| \geq 1} s_{\mathbf{u}} = 1.$$

In den insgesamt 2^d Sensitivitätskoeffizienten einer Funktion $f : \Omega^{(d)} \to \mathbb{R}$ ist somit die Information über die dimensionale Struktur von f , also die Relevanz verschiedener Koordinatenrichtungen und deren Interaktion miteinander, kodiert. Wie bereits erwähnt lässt sich bei vielen in der Praxis relevanten Funktionen ein Abfall der $s_{\mathbf{u}}$ für große $|\mathbf{u}|$ beobachten.

Ziel der folgenden Begriffe von effektiver Dimension ist es nun, den relevanten Teil der Information über die Wichtigkeit verschiedener Dimensionen in einer einzigen Zahl zusammenzufassen.

2.2.1 Superpositions-, Trunkations- und Mittlere Dimension

Trunkations- und Superpositionsdimension

Ein erster Begriff von Effektiver Dimension stammt von [CMO97] und gibt an, wieviele der ersten k Eingangsvariablen, bzw. welcher maximale Grad ihrer Interaktionen, einer Funktion diese bei Inkaufnahme eines geringen relativen Fehlers hinreichend genau beschreiben. In der Regel wird als Fehlermaß die $\mathcal{L}^2$-Norm verwendet (etwa in [CMO97], [Owe03], [IT04]), doch es ist durchaus auch möglich, andere (Semi-)Normen zu verwenden, wie es beispielsweise in [GH10b] und [Hol08] zur Definition eines auf der Anker-ANOVA Zerlegung basierenden Integrationsverfahren beschrieben wird.

Wir wollen uns jedoch zunächst an der Standard-Literatur orientieren und benutzen die $\mathcal{L}^2$-Norm, respektive die Varianz, um den Beitrag relevanter ANOVA-Terme zu messen.

Definition 2.14. (Trunkationsdimension)

Als *Trunkationsdimension* bezeichen wir die kleinste natürliche Zahl $d_T(\alpha)$, so dass für ein $\alpha \in (0, 1]$

$$\sum_{\mathbf{u} \subseteq \{1,\ldots,d_T(\alpha)\}} \sigma_{\mathbf{u}}^2(f) \geq \alpha \; \sigma^2(f)$$

gilt. Üblicherweise wählt man $\alpha \geq 0.99$.

Die Trunkationsdimension gibt also an, wieviele der ersten d_T Variablen tatsächlich relevant sind.

Definition 2.15. (Superpositionsdimension)

Als *Superpositionsdimension* bezeichen wir die kleinste natürliche Zahl $d_S(\alpha)$, so dass für ein $\alpha \in (0, 1]$

$$\sum_{|\mathbf{u}| \leq d_S(\alpha)} \sigma_{\mathbf{u}}^2(f) \geq \alpha \; \sigma^2(f)$$

gilt. Auch hier wählt man üblicherweise $\alpha \geq 0.99$.

Die Superpositionsdimension ist ein Maß für den maximalen Grad der Interaktion zwischen den Dimensionen. $d_S = 1$ impliziert, dass die Varianz der Funktion bis auf einen kleinen Teil allein durch eindimensionale Funktionen bestimmt ist, $d_S = 2$ bedeutet, dass sie durch eine Summe von ein- und zweidimensionalen Funktionen bestimmt ist, und so weiter. Angelehnt an dieses Konzept bezeichnen wir mit

$$\bar{d_S}(k) := \sum_{|\mathbf{u}| \leq k} s_{\mathbf{u}} \tag{2.20}$$

und

$$\bar{d_T}(k) := \sum_{\mathbf{u} \subseteq \{1..k\}} s_{\mathbf{u}} \tag{2.21}$$

den Anteil der Varianz, welcher sich in den ersten k Superpositions-, bzw. Trunkationsdimensionen befindet. Aus $d_S(\alpha) = k$ folgt also $\bar{d_S}(k) \geq \alpha$.

Satz 2.16. (Fehlerabschätzung für die abgeschnittene ANOVA-Reihe)

Es gelten

$$\|f - \sum_{|\mathbf{u}| \leq d_S(\alpha)} f_{\mathbf{u}}\|_{\mathcal{L}^2}^2 \leq (1 - \alpha) \; \sigma^2(f) \tag{2.22}$$

und

$$\|f - \sum_{\mathbf{u} \subseteq \{1..d_T(\alpha)\}} f_{\mathbf{u}}\|_{\mathcal{L}^2}^2 \leq (1 - \alpha) \; \sigma^2(f). \tag{2.23}$$

Beweis. Wir zeigen die Behauptung für die Superpositionsdimension:
Es gilt:

$$\|f - \sum_{|\mathbf{u}| \leq d_S(\alpha)} f_{\mathbf{u}}\|^2_{\mathcal{L}^2} = \| \sum_{|\mathbf{u}| > d_S(\alpha)} f_{\mathbf{u}}\|^2_{\mathcal{L}^2} = \sum_{|\mathbf{u}| > d_S(\alpha)} \sigma^2_{\mathbf{u}}(f) \leq (1-\alpha)\sigma^2(f).$$

Die zweite Behauptung folgt durch analoges Vorgehen. □

Allgemein beinhaltet das Konzept von Superpositions- und Trunkationsdimension jedoch die Schwierigkeit, dass eine Aussage wie „die Superpositionsdimension von f zum Parameter $\alpha = 0,99$ ist 5“ keinerlei Informationen darüber enthält, wie die Varianzmasse der $f_{\mathbf{u}}$ innerhalb der Blöcke $\{|\mathbf{u}| \leq 5\}$ und $\{|\mathbf{u}| > 5\}$ verteilt ist. Um ein genaueres Bild zu erhalten, müsste man $d_S(\alpha)$ für verschiede α oder $\bar{d_S}(k)$ für verschieden große k kennen.

Mittlere Dimension

Beim Konzept der *mittleren Dimension* (orig. *Mean Dimension*) nach [Owe03] betrachtet man eine Zufallsvariable U auf den Teilmengen $\mathbf{u} \subseteq \mathcal{D}$ mit den Wahrscheinlichkeiten

$$P[U = \mathbf{u}] = \frac{\sigma^2_{\mathbf{u}}}{\sigma^2}.$$

Misst man die Größe dieser Zufallsvariablen U nun durch die Kardinalität $|\mathbf{u}|$, so kommt man auf das Konzept der *mittleren Dimension im Superpositionssinne*, betrachtet man hingegen das größte Element $\max \mathbf{u}$, so kommt man auf die *mittlere Dimension im Trunkationssinne.*

Definition 2.17. (Mittlere Dimension im Superpositionssinne)

Als *mittlere Dimension im Superpositionssinne* definieren wir den Erwartungswert der Zufallsvariable $|U|$

$$d_{M_S} := \sum_{|\mathbf{u}| \geq 1} s_{\mathbf{u}} \, |\mathbf{u}| \tag{2.24}$$

$$= \sum_{k=1}^{d} k \cdot \sum_{|\mathbf{u}|=k} \frac{\sigma^2_{\mathbf{u}}}{\sigma^2} \tag{2.25}$$

Definition 2.18. (Mittlere Dimension im Trunkationssinne)

Als *mittlere Dimension im Trunkationssinne* definieren wir den Erwartungswert der Zufallsvariable $\max(U)$, also

$$d_{M_T} := \sum_{|\mathbf{u}| \geq 1} s_{\mathbf{u}} \, \max\{j : j \in \mathbf{u}\} \tag{2.26}$$

$$= \sum_{k=1}^{d} k \cdot \sum_{\max(\mathbf{u})=k} \frac{\sigma^2_{\mathbf{u}}}{\sigma^2}. \tag{2.27}$$

Satz 2.19. (Abschätzung der mittleren Dimension)
Es gilt

$$d_{M_S} \leq d_S + (1-\alpha)d$$

und

$$d_{M_T} \leq d_T + (1-\alpha)d.$$

Beweis. Wir zeigen die Behauptung für die mittlere Dimension im Superpositionssinne:

$$\begin{aligned} d_{M_S} &= \sum_{k=1}^{d} k \cdot \left(\sum_{|\mathbf{u}|=k} \frac{\sigma_{\mathbf{u}}^2}{\sigma^2} \right) \\ &= \sum_{k=1}^{d_S} k \cdot \left(\sum_{|\mathbf{u}|=k} \frac{\sigma_{\mathbf{u}}^2}{\sigma^2} \right) + \sum_{k=d_S+1}^{d} k \cdot \left(\sum_{|\mathbf{u}|=k} \frac{\sigma_{\mathbf{u}}^2}{\sigma^2} \right) \\ &\leq d_S \cdot \sum_{k=1}^{d_S} \left(\sum_{|\mathbf{u}|=k} \frac{\sigma_{\mathbf{u}}^2}{\sigma^2} \right) + d \cdot \sum_{k=d_S+1}^{d} \left(\sum_{|\mathbf{u}|=k} \frac{\sigma_{\mathbf{u}}^2}{\sigma^2} \right) \\ &\leq d_S \cdot \alpha + d(1-\alpha) \\ &\leq d_S + d(1-\alpha) \end{aligned}$$

Die zweite Behauptung folgt durch analoges Vorgehen. □

2.2.2 Ein allgemeinerer Begriff von effektiver Dimension

Nachdem wir uns jetzt einige Konzepte von effektiver Dimension angesehen haben, stellt sich die Frage, inwiefern sich daraus Erkenntnisse für die numererische Praxis ergeben. Im Abschnitt 2.2.1 haben wir bereits gesehen, dass die Angabe der Superpositions- oder Trunkationsdimension zu einem festen Parameter α zwar eine Fehlerabschätzung für die abgeschnittene ANOVA-Reihe (siehe Satz 2.16) zulässt, dies aber keinerlei Aussage darüber beinhaltet, mit welchen tatsächlichen Kosten die Lösung des hochdimensionalen Problemes verbunden ist, wenn man einen Fehler erreichen möchte, der kleiner als $(1-\alpha)$ ist.

Eine gewisse Abhilfe schafft hier die Mittlere Dimension, denn in ihre Berechnung fließen sämtliche ANOVA-Terme nach Gewichtung mit einem dimensionsabhängigen Faktor ein. Dennoch lässt sich auch hier kein direkter Zusammenhang zum numerischen Fehler – und damit der Praxis – herstellen.

Daher wollen wir in diesem Abschnitt nun das Konzept der mittleren Dimension verallgemeinern, indem wir

- beliebige Maße μ zur Definition der ANOVA-Zerlegung $f = \sum_{\mathbf{u}} f_{\mathbf{u}}^{\mu}$,
- beliebe (Semi-)Normen $\|\cdot\|_*$ zur Messung der Beiträge von $f_{\mathbf{u}}^{\mu}$ und
- beliebige $\gamma_{\mathbf{u}} \in \mathbb{R}^+$ zur Gewichtung der $\|f_{\mathbf{u}}^{\mu}\|_*$

zulassen, womit sich als *verallgemeinerte effektive Dimension* die Zahl

$$d_A := \frac{1}{\sum_{\mathbf{u}} \|f_{\mathbf{u}}^{\mu}\|_*} \sum_{\mathbf{u}\subseteq\mathcal{D}} \gamma_{\mathbf{u}} \|f_{\mathbf{u}}^{\mu}\|_*. \tag{2.28}$$

ergibt.

Die sinnvolle Wahl von Maß, Norm und Gewichtung ist abhängig von der betrachteten Anwendung und deren Hintergrund.

Beispiel 2.2: (Spezialfälle von d_A)

Wir geben einige Beispiele aus der Literatur hochdimensionaler Probleme an, die Spezialfälle der Definition (2.28) darstellen.

1. Mittlere Dimension d_{M_S}:
   - $\mu = \lambda^d$ sei das Lebesgue-Maß
   - $\|\cdot\|_* = \|\cdot\|_2$, die $\mathcal{L}^2$-Norm
   - $\gamma_{\mathbf{u}} = |\mathbf{u}|$.
2. Relative Superpositionsdimension $\bar{d_S}(k)$:
   - $\mu = \lambda^d$
   - $\|\cdot\|_* = \|\cdot\|_2$
   - $\gamma_{\mathbf{u}} = 1$ für $|\mathbf{u}| \leq k$, ansonsten $\gamma_{\mathbf{u}} = 0$.
3. *Anchored Effective Dimension* $T_{\mathbf{u}}$ aus [Hol08]:
   - $\mu = \delta^{\boldsymbol{a}}$ sei das Dirac-Maß im Punkte $\boldsymbol{a}$ $\Rightarrow$ Anker-ANOVA Zerlegung
   - $\|g\|_* = |\int g \, d\lambda|$, der Betrag des Erwartungswert
   - $\gamma_{\mathbf{u}}$ wie in 2.
4. *Weighted Spaces* aus [Feu10]:
   - $\mu = \delta^{\mathbf{0}}$ sei das Dirac-Maß im Punkte Null $\Rightarrow$ Anker-ANOVA Zerlegung
   - $\|g\|_* = |g|_{2,2} = \|D^{\mathbf{2}} g\|_2$ die Mix-Seminorm [1].
   - $\gamma_{\mathbf{u}}$ Dimensionsgewichte zur Definition des zugrundeliegenden *Weighted Space*.
5. Quasi-Monte Carlo Quadraturfehler (siehe Abschnitt 2.2.3):
   - $\mu = \lambda^d$, Lebesgue-Maß
   - $\|\cdot\|_* = \|\cdot\|_{\mathcal{V}}$ sei die Variation nach Hardy und Krause
   - $\gamma_{\mathbf{u}} = \frac{\log(N)^{|\mathbf{u}|}}{N}$, für ein gegebenes $N \in \mathbb{N}$

Obwohl natürlich nicht alle Kombinationen Sinn machen, wollen wir die Definition weiterhin in dieser Allgemeinheit verwenden, da die Entscheidung, welche abstrakten Einschränkungen gelten, nicht trivial ist. In Abschnitt 3.2.2 werden wir dann jedoch einige, durch eine günstige

[1] mit $D^{\boldsymbol{\alpha}} g := \frac{\partial^{|\boldsymbol{\alpha}|_1}}{\prod_{m=1}^d \partial x_m^{\alpha_m}}$ bezeichnen wir die α_m-fache Ableitung in jede Richtung $m = 1, \ldots, d$

Numerik begründete, Einschränkungen treffen, welche wir im folgenden Beispiel schon einmal vorwegnehmen.

Beispiel 2.3: (Verallgemeinerte effektive Dimension zum Maß μ)
Wählt man als Norm $\|\cdot\|_*$ die durch μ induzierte $\mathcal{L}^2$-Norm, so ergibt sich

$$d_A = \frac{1}{\sigma_\mu^2} \sum_{\mathbf{u}\subseteq\mathcal{D}} \gamma_{\mathbf{u}}\, \sigma^2_{\mathbf{u},\mu} = \sum_{\mathbf{u}\subseteq\mathcal{D}} \gamma_{\mathbf{u}}\, s_{\mathbf{u},\mu}$$

Durch die Orthogonalität der ANOVA-Terme lässt sich die Summe also auf einfache Weise normieren. Später werden wir sehen, dass dies einige Vorteile bietet.

Im nächsten Unterabschnitt werden wir einige allgemeine d_A betreffende Aussagen beweisen und anschließend darlegen, wieso die Fehlerabschätzung für Quasi-Monte Carlo Quadratur und die Approximation mit Dünnen Gittern einen Speziallfall von (2.28) darstellen.

Gewichtete Räume

Unsere neu definierte Kennzahl d_A hat eine große Ähnlichkeit zum Konzept der *Gewichteten Räume*, welches in [Feu10] ausführlich diskutiert wird, um daraus a-priori optimierte Dünne Gitter zu konstruieren. Wir wollen dieses Konzept mit wenigen Änderungen in der Notation hier übernehmen und es exemplarisch auf die mittlere Dimension im Superpositions- und Trunkationssinne anwenden.

Definition 2.20. (Gewichtete Räume)
Für alle $\mathbf{u} \subseteq \mathcal{D}$ seien positive Gewichte $\gamma_{\mathbf{u}} \in [0,\infty]$ gegeben. (Gegebenenfalls kann man auch die Menge aller Gewichte bezüglich ihres größten Elementes oder ihrer Summe $\sum_{\mathbf{u}} \gamma_{\mathbf{u}}$ normieren).
Dann ist für jede (Semi-)Norm $\|\cdot\|_*$ eines Hilbertraumes $V^{(d)}$ und jedes normierte Produktmaß μ auch

$$\|f\|_{\boldsymbol{\gamma}} := \sum_{\mathbf{u}\subseteq\mathcal{D}} \gamma_{\mathbf{u}}\ \|f^\mu_{\mathbf{u}}\|^2_*$$

eine (Semi-)Norm auf

$$H^{\boldsymbol{\gamma}}_{\mu,*} := \{f \in V^{(d)} : \|f\|_{\boldsymbol{\gamma}} < \infty\}. \tag{2.29}$$

Der Fall $\gamma_{\mathbf{u}} = \infty$ sei mit der Definition $\infty \cdot 0 = 0$ explizit zugelassen.

Wegen

$$\|f^\mu_{\mathbf{u}}\|_* = \frac{1}{\gamma_{\mathbf{u}}}(\|f\|_{\boldsymbol{\gamma}} - \sum_{\mathbf{v}\neq\mathbf{u}} \gamma_{\mathbf{v}}\|f^\mu_{\mathbf{v}}\|_*) \tag{2.30}$$

gilt nun die Abschätzung

$$\|f^\mu_{\mathbf{u}}\|_* \leq \frac{1}{\gamma_{\mathbf{u}}}\|f\|_{\boldsymbol{\gamma}}. \tag{2.31}$$

Die Gewichte $\gamma_{\mathbf{u}}$ lassen also eine (relativ grobe) Abschätzung für den Anteil der zu $\mathbf{u}$ gehörigen ANOVA-Komponente $\|f^\mu_{\mathbf{u}}\|_*$ an der gewichteten Norm $\|f\|_{\boldsymbol{\gamma}}$ zu.

Wendet man diese Abschätzung nun beispielsweise auf die in (2.24) definierte mittlere Dimension im Superpositionssinne d_{M_S} an, so erhält man für Funktionen mit normierter Varianz durch (2.31) die Ungleichung

$$\|f_{\mathbf{u}}\|_2 \leq \frac{1}{|\mathbf{u}|} d_{M_S}.$$

Für die mittlere Dimension im Trunkationssinne aus (2.26) ergibt sich

$$\|f_{\mathbf{u}}\|_2 \leq \frac{1}{\max \mathbf{u}} d_{M_T}.$$

Für den höchsten ANOVA-Term $f_{\mathcal{D}}$ ergeben sich somit die oberen Schranken

$$\|f_{\mathcal{D}}\|_2 \leq \frac{1}{d}\, d_{M_S} \quad \text{und} \quad \|f_{\mathcal{D}}\|_2 \leq \frac{1}{d}\, d_{M_T}, \tag{2.32}$$

wobei jedoch für viele praktisch relevante Funktionen ein deutlich stärkerer Abfall in der Relevanz der Dimensionen festgestellt wurde [WS03].

2.2.3 Dimensionsweise Zerlegung des numerischen Fehlers

Der Bezug der konventionellen Definitionen von effektiver Dimension, welche sich der $\mathcal{L}^2$-Norm bedienen (um die Orthogonalität der ANOVA-Terme auszunutzen) sind ungeeignet um tatsächliche Bezüge zu Fehlerabschätzungen oder der ε-Komplexität mehrdimensionaler numerischer Methoden herzustellen, da hierzu andere Normen, wie etwa die Variation nach Hardy und Krause [Nie92] oder Mix-(Semi-)Normen [BG04, Gri06] benötigt werden. Die von uns vorgeschlagene Verallgemeinerung lässt derartige Normen explizit zu – auch wenn dadurch die Orthogonalität und damit die Vorraussetzungen für eine einfache Numerik wegfallen.

Obwohl die im nächsten Kapitel von uns angegebenen Methoden zur Reduktion der effektiven Dimension diese Orthogonalität benötigen und wir daher wieder auf der $\mathcal{L}^2$-Norm aufbauen werden, wollen wir uns anhand des Quadraturfehlers von Quasi-Monte Carlo Methoden anschauen, weshalb die Verallgemeinerung (2.28) sinnvoll ist und Gegenstand zukünftiger Untersuchungen sein sollte.

Das grundsätzliche Prinzip dabei ist es, den Fehler hochdimensionaler Verfahren in seine ANOVA-Bestandteile aufzuspalten und diese dann getrennt abzuschätzen. Für die stückweise-lineare Interpolation auf Dünnen Gittern wird dies in [Feu10] durch ein recht kompliziertes Drei-Term-Splitting in konstanten und linearen Anteil plus Rest erreicht.

Für Quasi-Monte Carlo Methoden auf dem Einheitswürfel lässt sich dies etwas leichter bewerkstelligen, weshalb wir dieses Beispiel hier vorrechnen werden. Der erste Teil der folgenden Abschätzung stammt aus [WS03], der zweite aus [Nie92].

Mit $Q^N(f) := \sum_{i=1}^N f(\boldsymbol{x}^{(i)})$ bezeichnen wir die Quasi-Monte Carlo Nährung an $\int_{[0,1]^d} f(\boldsymbol{x})\, d\boldsymbol{x}$.

Für den absoluten Fehler gilt dann

$$\begin{aligned}
\left| \int_{[0,1]^d} f(\boldsymbol{x})\, d\boldsymbol{x} - Q^N(f) \right| &= \left| \int_{[0,1]^d} \sum_{\mathbf{u}\subseteq\mathcal{D}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})\, d\boldsymbol{x} - Q^N\left(\sum_{\mathbf{u}\subseteq\mathcal{D}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) \right) \right| \\
&\leq \sum_{\mathbf{u}\subseteq\mathcal{D}} \left| \int_{[0,1]^d} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})\, d\boldsymbol{x} - Q^N\left(f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) \right) \right| \\
&\leq \sum_{\mathbf{u}\subseteq\mathcal{D}} \|f_{\mathbf{u}}\|_{\mathcal{V}}\ \mathfrak{D}_N^*(\boldsymbol{x}_{\mathbf{u}}) \\
&\leq \sum_{\mathbf{u}\subseteq\mathcal{D}} \|f_{\mathbf{u}}\|_{\mathcal{V}}\ C\frac{\log(N)^{|\mathbf{u}|}}{N}.
\end{aligned}$$

wobei $\|f_{\mathbf{u}}\|_{\mathcal{V}}$ die Variation von $f_{\mathbf{u}}$ nach Hardy und Krause und $\mathfrak{D}_N^*(\boldsymbol{x}_{\mathbf{u}})$ die $*$-Diskrepanz der Punktfolge $\boldsymbol{x}^{(i)}$ bezeichne. Eine genaue Erläuterung dieser Begriffe findet sich in [Nie92] und im Anhang von [Hol08].

Die Abschätzung für die $*$-Diskrepanz

$$\mathfrak{D}_N^*(\boldsymbol{x}_{\mathbf{u}}) \leq C\frac{\log(N)^{|\mathbf{u}|}}{N}$$

findet sich ebenfalls in [Nie92] und gilt beispielsweise für die Halton-Folge [Hol08].

3 Verfahren zur Reduktion der effektiven Dimension

In diesem Kapitel werden wir uns nun mit der Frage befassen, mit welchen Methoden und unter welchen Vorraussetzungen man die effektive Dimension einer d-dimensionalen, reellwertigen Funktionen f verringern kann. Dazu betrachten wir f in einem anderen Koordinatensystem, also an Stelle von f die Funktion $\hat{f} := f \circ \psi$, wobei ψ einen *Diffeomorphismus* bezeichne.

Die elementaren Bausteine der Dimensionsreduktionsverfahren, die wir im Laufe dieses Kapitels entwickeln, werden wir zwar für beliebige Koordinatensysteme/Diffeomorphismen einführen und diskutieren, jedoch müssen wir dabei stets im Hinterkopf behalten, dass eine Diskretisierung allgemeiner Bijektionen d-dimensionaler Gebiete in sich selbst unweigerlich wieder auf den Fluch der Dimension führt. Ziel muss es also sein, Klassen von Diffeomorphismen zu konstruieren, deren Diskretisierung nicht exponentiell von der Dimension abhängt.

Geeignete Diffeomorphismen Anhand einiger Beispiele werden wir uns kurz mit Diffeomorphismen beschäftigen, die nur komponentenweise agieren (und damit nicht dem Fluch der Dimension unterworfen sind) – das Hauptaugenmerk dieser Arbeit soll jedoch auf der *speziellen Orthogonale Gruppe* $\mathbf{SO}(d)$ liegen. Dabei handelt es sich um die Menge aller orthogonalen Matrizen $\boldsymbol{Q}$ mit Determinante $\det \boldsymbol{Q} = 1$, was gerade den Drehungen des d-dimensionalen Koordinatensystemes um $\frac{d(d-1)}{2}$ verschiedene Achsen entspricht – die Zahl ihrer Freiheitsgrade wächst also nur quadratisch mit der Dimension d.

Diese Matrizen stellen auf jedem rotationssymmetrischen Gebiet Ω eine Teilmenge der Diffeomorphismen von Ω auf sich selbst dar – insbesondere also auf dem Ganzraum $\mathbb{R}^d$. Sie besitzen den Vorteil, dass sie im Rahmen der Linearen Algebra sehr gut verstanden sind, die Produktstruktur rotationsinvarianter Maße (insbesondere des Gauß-Maßes) erhalten und als Lie-Gruppe zudem mit der Struktur einer differenzierbaren Mannigfaltigkeit versehen werden können, womit später die Anwendung von *geometrischen Optimierungsmethoden* möglich ist.

Obwohl orthogonale Abbildungen „nur“ linear sind, lassen sich mit ihnen bereits erstaunlich gute Resultate in der Reduktion der effektiven Dimension erzielen, wie wir im nächsten Kapitel sehen werden. Dafür werden wir bereits in diesem Kapitel einige analytische, aber für den allgemeinen Fall durchaus instruktive Beispiele angeben.

Eine Verallgemeinerung des Konzeptes der orthogonalen Abbildungen stellen die von uns entwickelten *stückweise orthogonalen Transformationen* dar. Diese basieren auf der Idee, ein rotationssymmetrisches Gebiet in disjunkte Kreisringe zu unterteilen und auf jedem dieser Ringe eine eigene orthogonale Abbildung zu definieren.

Lässt man den Radius dieser Kreisringe gegen Null gehen, so kann man eine überall differenzierbare Bijektion konstruieren, welche im Gegensatz zu einer einzelnen orthogonalen Matrix nicht nur das gesamte Koordinatensystem dreht, sondern einzelne Achsen auch krümmen kann. Zudem erhalten diese Abbildungen ebenfalls die Produktstruktur des Gauß-Maßes und ihre Komplexität hängt nur quadratisch von der Dimension ab.

Formalisierung der Dimensionsreduktion Die Reduktion der effektiven Dimension reellwertiger Funktionen formalisieren wir durch die Minimierung eines Funktionals $\mathfrak{M}$ über der Menge der speziellen orthogonalen Matrizen $\mathbf{SO}(d)$. Dieses ist per se numerisch nicht zu handhaben, allerdings lässt sich unter geeigneten Vorraussetzungen ein äquivalentes Funktional angeben, das die selben kritischen Punkte besitzt, aber deutlich günstiger auszuwerten ist. Die dabei auftretenden Integrale diskretisieren wir entweder für jede Auswertung des Funktionals durch Dünne Gitter oder Quasi-Monte Carlo Quadraturformeln, oder wir projizieren f einmalig in den Raum der *homogenen Polynome* n-ten Grades $\mathcal{P}_n$. Dieser ist abgeschlossen unter allen Transformationen aus $\mathbf{SO}(d)$, wodurch man die Quadratur nun analytisch vollziehen kann, was das Verfahren substanziell beschleunigt.

Vollzieht man diese Projektion auf $\mathcal{P}_n$ über die abgeschnittene Taylorreihe, so erhält man für lineare Polynome die Lineare Transformation (LT) [IT06] und für den quadratischen fall die Diagonal Methode (DM) [Mor98].

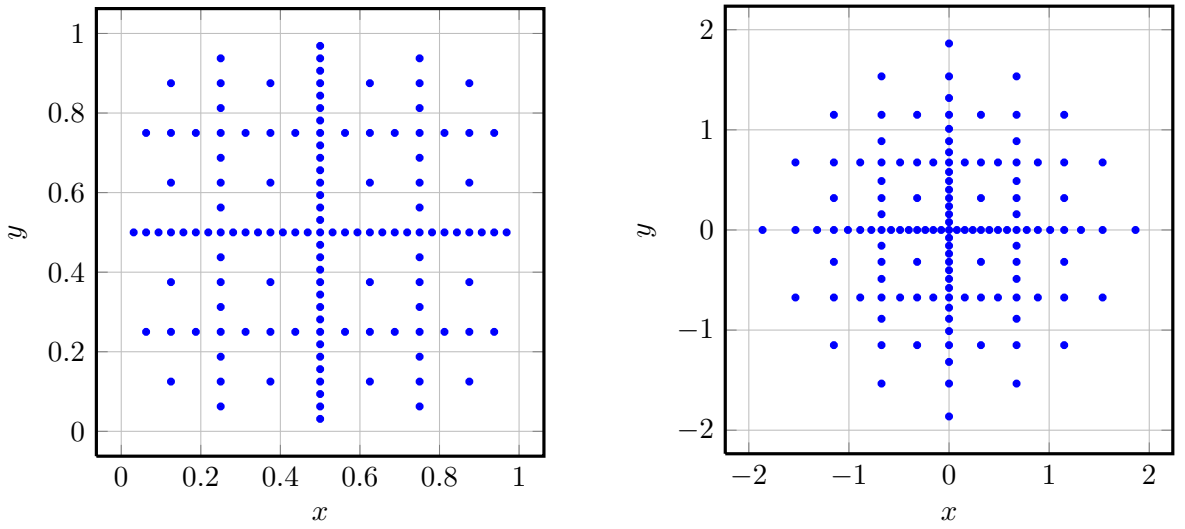
Allerdings besitzt die Taylorreihe im Allgemeinen schlechtere Approximationseigenschaften als andere Interpolationsmethoden, weshalb wir die orthogonale $\mathcal{L}^2$-Projektion auf den Untervektorraum der homogenen Polynome verwenden wollen. Den dazu notwendigen Projektionsoperator realisieren wir durch eine Dünngitter-Diskretisierung zu einem kleinen Level und der anschliessenden Lösung eines gewichteten Least-Squares Problems.

Eine weitere Vereinfachung ergibt sich, wenn man die dimensionale Struktur von f nicht im Superpositionssinne, sondern nur im Trunkationssinne optimieren möchte. Dann kann man, angelehnt an das Konzept der *Linearen Transformation* nach Imai und Tan, die Suche auf jene Drehungen aus $\mathbf{SO}(d)$ einschränken, die nur die ersten p Spalten der Matrix verändern. Diese lassen sich durch die sogenannte Stiefelmannigfaltigkeit $\mathbf{St}(p, d)$ beschreiben, deren Komplexität bei kleinem p sogar nur noch linear von der Dimension abhängt.

Optimierung auf Mannigfaltigkeiten Um $\mathfrak{M}$ auf $\mathbf{St}(p, d)$ (Trunkationsdimension) zu minimieren verwenden wir ein *geometrisches CG-Verfahren* aus [AMS08], welches im Tangentialraum der Mannigfaltigkeit operiert und vermöge von *Retraktionen* die lokale Struktur der Mannigfaltigkeit ausnutzt.

Wegen $\mathbf{SO}(d) = \mathbf{St}(d, d)$ könnte man dieses Verfahren prinzipiell auch für die Superpositionsdimension benutzen, jedoch werden wir hier ein eigenes Verfahren basierend auf der *Riemannschen Exponentialabbildung* konstruieren, welches besser für das Funktional $\mathfrak{M}$ geeignet ist, da es weniger Funktionsauswertungen benötigt.

Dabei nutzen wir, dass $\mathbf{SO}(d)$ eine Lie-Gruppe mit zugehöriger Lie-Algebra $\mathfrak{so}(d)$ ist, auf welcher wir die Exponentialabbildung mit einer Pade-Approximation siebten Grades darstellen.



(a) Kartesische Koordinaten auf $(0,1)^2$

(b) Inverse Normalverteilung auf $\mathbb{R}^2$

Abb. 3.1: Ein zweidimensionales dünnes Gitter vom Level 5 in verschiedenen Koordinantesystemen

3.1 Wahl des Koordinatensystems

Wie bereits angedeutet, wollen wir, um die effektive Dimension von $f : \Omega^{(d)} \to \mathbb{R}$ zu verringern, in ein anderes Koordinatensystem von $\Omega^{(d)} \subseteq \mathbb{R}^d$ wechseln. Ein solcher Wechsel von kartesischen Koordinaten des $\mathbb{R}^d$ in ein anderes Koordinatensystem wird durch einen *Diffeomorphismus* definiert. In diesem Abschnitt werden wir exemplarisch anhand der Approximation und der Quadratur die grundsätzliche Bedeutung und Anwendung von Diffeomorphismen auf hochdimensionale Probleme deutlich machen und Klassen von Diffeomorphismen identifizieren, deren Diskretisierung nicht dem Fluch der Dimension unterworfen ist – deren Freiheitsgrade also nicht exponentiell von der Dimension abhängig sind.

3.1.1 Beliebige Diffeomorphismen

Zunächst führen wir Diffeomorphismen und ihre möglichen Anwendungen auf hochdimensionale Probleme anhand einiger ausgewählter Beispiele in größtmöglicher Allgemeinheit ein.

Definition 3.1. (Diffeomorphismus)

Ein *Diffeomorphismus* ψ zwischen zwei Mengen $\hat{\Omega} \subseteq \mathbb{R}^d$ und $\Omega^{(d)} \subseteq \mathbb{R}^d$ ist eine überall differenzierbare, bijektive Abbildung, d.h. ihr Differential $D\psi(\boldsymbol{x})$ ist für alle $\boldsymbol{x} \in \hat{\Omega}$ invertierbar.

$$\mathrm{Diff}(\hat{\Omega}, \Omega^{(d)}) := \{\psi : \hat{\Omega} \to \Omega^{(d)} : |\det\, D\psi(\boldsymbol{x})| > 0 \text{ für alle } \boldsymbol{x} \in \hat{\Omega}\} \tag{3.1}$$

Der Transformationssatz für die Integration

Der *Transformationssatz* ist ein unverzichtbares Werkzeug der Maß- und Integrationstheorie, da er es unter gewissen Vorraussetzungen ermöglicht, die Quadratur einer Funktion f auf die Quadratur von $f \circ \psi$ zurückzuführen.

Satz 3.2. (Transformationssatz)
Für jeden Diffeomorphismus $\psi : \hat{\Omega} \to \Omega^{(d)}$, $\hat{\Omega} \subseteq \mathbb{R}^d$ gilt die Identität

$$\int_{\Omega^{(d)}} f(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\hat{\Omega}} f \circ \psi(\boldsymbol{y})\ |\det\ D\psi(\boldsymbol{y})|\, d\boldsymbol{y}. \tag{3.2}$$

Integriert man bezüglich eines zum Lebesgue-Maß absolut-stetigen Maßes μ mit zugehöriger Dichtefunktion φ, so ist zu beachten, dass man auch die Dichtefunktion transformieren muss:

$$\int_{\Omega^{(d)}} f(\boldsymbol{x})\, d\mu(\boldsymbol{x}) = \int_{\Omega^{(d)}} f(\boldsymbol{x})\, \varphi(\boldsymbol{x})\, d\boldsymbol{x} \tag{3.3}$$

$$= \int_{\hat{\Omega}} f \circ \psi(\boldsymbol{y})\ |\det\ D\psi(\boldsymbol{y})|\, \varphi(\psi(\boldsymbol{y}))\, d\boldsymbol{y}. \tag{3.4}$$

Beweis. Der Beweis findet sich in den meisten Einführungen in die Analysis oder Maß-Theorie – etwa [Kön09]. □

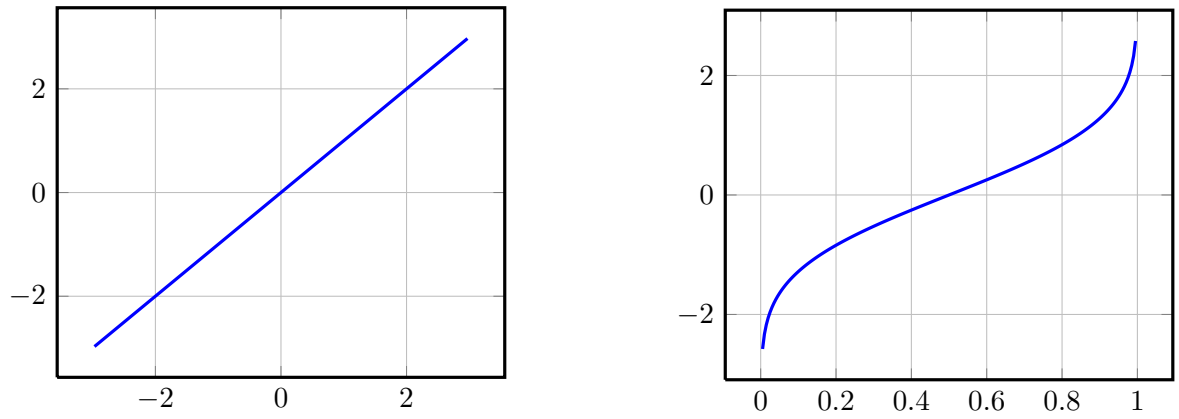
Bemerkung: Der zweite Teil von Satz 3.2 ist besonders für Produktmaße $\mu = \otimes\mu_i$ relevant. Denn es ist durchaus möglich, dass die Anwendung eines Diffeomorphismus dessen Produktstruktur zerstört. Dies hätte die Konsequenz, dass sich die ANOVA-Terme nicht mehr durch Satz 2.3 charakterisieren und berechnen lassen.

Ein einfaches, aber praxisrelevantes Beispiel ist die Transformation von Gauß-Integralen vom Ganzraum $\mathbb{R}$ auf das Einheitsintervall $[0, 1]$.

Beispiel 3.1: (Inverse Normalverteilung)

$\eta(x)$ bezeichne das eindimensionale Gauß-Maß auf $\mathbb{R}$. Seine Dichtefunktion ist $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2})$ und seine Verteilungsfunktion $\phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt$. Es gilt $\phi'(x) = \varphi(x)$ und $(\phi^{-1})'(y) = \frac{1}{\phi'(\phi^{-1}(y))}$. Damit folgt durch den Transformationssatz 3.2

$$\begin{aligned}
\int_{\mathbb{R}} f(x)\, d\eta(x) &= \int_{\mathbb{R}} f(x)\ \varphi(x)\, dx \\
&= \int_{\phi(\mathbb{R})} f(\phi^{-1}(y))\ |(\phi^{-1})'(y)|\ \varphi(\phi^{-1}(y))\, dy \\
&= \int_0^1 f(\phi^{-1}(y)) \frac{1}{\phi'(\phi^{-1}(y))}\ \varphi(\phi^{-1}(y))\, dy \\
&= \int_0^1 f(\phi^{-1}(y))\, dy.
\end{aligned}$$



(a) f (b) $f \circ \phi^{-1}$

Abb. 3.2: $f(x) = x$ auf $\mathbb{R}$ und nach Transformation mit der inversen Normalverteilung ϕ^{-1} auf den Einheitswürfel

Kennt man keine Quadraturformel für das unbeschränkte Gebiet $\mathbb{R}$, wie beispielsweise die Hermite-Integration (siehe [Hol08]), oder kann diese aufgrund mangelnder Regularität des Integranden nicht verwenden, bildet man den Ganzraum also diffeomorph auf $(0, 1)$ ab und bedient sich hier nun der zahlreichen Standard-Methoden für beschränkte Gebiete, wie etwa Monte Carlo-Verfahren, Trapezregel oder Gauß-Quadratur.

Man beachte, dass Funktionen, die für $\lim_{x\to\pm\infty}$ divergent sind durch die Transformation auf das Einheitsintervall an den Rändern Singularitäten entstehen, wie in Abbildung 3.2 anhand der linearen Funktion $f(x) = x$ zu sehen ist.

Allgemeiner ist der folgende Satz, der die einfache numerische Berechnung der ANOVA-Zerlegung zu Maßen, deren inverse Verteilungsfunktion bekannt ist, erlaubt.

Satz 3.3. (Transformationssatz für die ANOVA-Zerlegung)

Ist μ ein zum d-dimensionalen Lebesgue-Maß λ^d absolut-stetiges Wahrscheinlichkeitsmaß auf $\Omega^{(d)}$ mit Verteilungsfunktion $\phi : \Omega^{(d)} \to [0, 1]$ und Dichtefunktion $\phi' =: \varphi$, so läßt sich die ANOVA-Zerlegung mit der inversen Verteilungsfunktion ϕ^{-1} über $[0, 1]^d$ berechnen, denn es gilt:

$$P^{\mu}_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) = P^{\lambda}_{\mathbf{u}}(f \circ \phi^{-1})(\phi(\boldsymbol{x}_{\mathbf{u}})), \tag{3.5}$$

womit man offensichtlich

$$f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}) = (f \circ \phi^{-1})_{\mathbf{u}}(\phi(\boldsymbol{x}_{\mathbf{u}})) \tag{3.6}$$

erhält.

Beweis. Wir beweisen (3.5) mit Hilfe des Transformationssatzes 3.2.

$$P_{\mathbf{u}}(f)(\boldsymbol{x}_{\mathbf{u}}) = \int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x})\, d\mu_{\mathbf{u}^c}(\boldsymbol{x})$$

$$
\begin{aligned}
&= \int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x})\varphi_{\mathbf{u}^c}(\boldsymbol{x})\, d\boldsymbol{x}_{\mathbf{u}^c} \\
&= \int_{[0,1]^{\mathbf{u}^c}} f(x_{\mathbf{u}}, \phi_{\mathbf{u}^c}^{-1}(\boldsymbol{x}_{\mathbf{u}^c}))\, d\boldsymbol{x}_{\mathbf{u}^c} \\
&= \int_{[0,1]^{\mathbf{u}^c}} f \circ \phi^{-1}(\phi_{\mathbf{u}}(x_{\mathbf{u}}), (\boldsymbol{x}_{\mathbf{u}^c}))\, d\boldsymbol{x}_{\mathbf{u}^c}
\end{aligned}
$$

□

Interpolation / Approximation

Bei der Approximation und der Interpolation von Funktionen geht es darum, aus einer endlichen Zahl von bekannten Stützstellen eine möglichst genaue Annäherung von f an unbekannten Punkten zu erzielen.

Ist $\psi : \hat{\Omega} \to \Omega^{(d)}$ ein gegebener Diffeomorphismus, so ist es möglich, anstatt eine Funktion $f : \Omega^{(d)} \to \mathbb{R}$ in der kanonischen Basis zu interpolieren, dies in den durch ψ definierten Koodinaten von $\Omega^{(d)}$ zu tun

$$f_h \approx f \circ \psi : \hat{\Omega} \to \mathbb{R},$$

womit dann

$$f \approx f_h \circ \psi^{-1}$$

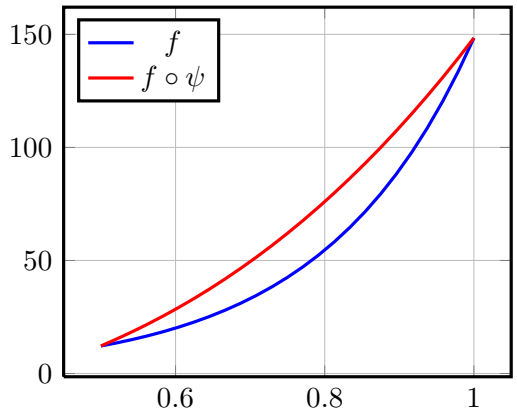
gilt. Dies kann dann sinnvoll sein, wenn $f \circ \psi$ gutartiger ist, als f selbst. Zum besseren Verständnis betrachten wir zwei Beispiele.

Beispiel 3.2: (Gittergradierungen)

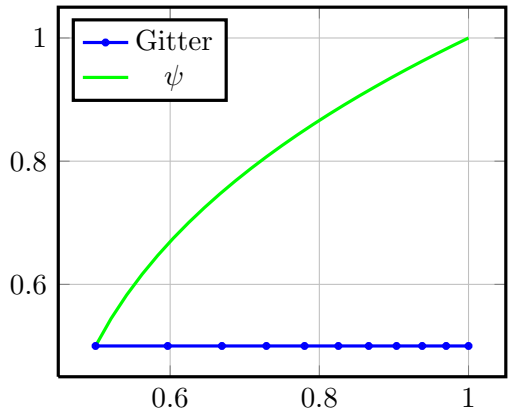
Auf dem Intervall $[\frac{1}{2}, 1]$ sei $f(x) = \exp(5x)$ durch eine stückweise lineare Funktion f_h auf einem äquidistanten Gitter der Maschenweite h zu interpolieren. Für den $\mathcal{L}^2$-Fehler auf $[\frac{1}{2}, 1]$ gilt dann nach [Feu10] die Abschätzung

$$
\begin{aligned}
E_h(f) &:= \int_{\frac{1}{2}}^{1} (f(x) - f_h(x))^2\, dx \\
&\leq \frac{h^2}{9} \int_{\frac{1}{2}}^{1} f''(x)^2\, dx \\
&= \frac{h^2}{9} \int_{\frac{1}{2}}^{1} 25 \exp(x)^2\, dx \\
&\approx 1.151 \cdot 10^5\, h^2.
\end{aligned}
$$

Betrachtet man statt dessen jedoch f in den durch $\phi(x) := \sqrt[3]{x}$ gradierten Koordinaten, so interpoliert man $\hat{f} := f \circ \psi = \exp(5\sqrt[3]{x})$ auf $\hat{\Omega} = \psi^{-1}([\frac{1}{2}, 1]) = [\frac{1}{8}, 1]$, wobei $\psi^{-1}(x) = x^3$ gilt.



(a) f und $f \circ \psi$



(b) $\psi(x) = \sqrt[3]{1.75x - 0.75}$ und das dadurch definierte Gitter

Abb. 3.3: Gittergradierung von $f(x) := \exp(5x)$ auf $[\frac{1}{2}, 1]$

Somit ergibt sich für den Interpolationsfehler der deutlich kleinere Wert

$$E_h(\hat{f}) \leq \frac{h^2}{9} \int_{\frac{1}{8}}^{1} \hat{f}''(x)^2 \, dx \approx 0.036 \cdot 10^5 \, h^2.$$

Der gradierte Interpolationsfehler $E_h(\hat{f})$ konvergiert damit ungefähr um den Faktor 40 schneller als der nicht gradierte Fehler $E_h(f)$. Die Funktion f, die Transformation ψ und die transformierte Funktion $f \circ \psi$ (linear auf $[\frac{1}{2}, 1]$ zurückskaliert) sind in Abb. 3.3 dargestellt. Die geringere Krümmung von $f \circ \psi$ ist deutlich zu erkennen.

Gradiert man mit der exakten Inversen $\phi = f^{-1}$ von f (sofern diese existiert), so ist $\hat{f}$ sogar linear. Allerdings ist das Bestimmen der Inversen nur mit unverhältnismäßigen Kosten möglich. Wie man jedoch an diesem Beispiel sehen konnte, ist es bereits mit sehr geringem Aufwand möglich eine Verbesserung in der Konstanten zu erzielen, indem man die Punktdichte des Gitters an Stellen erhöht, an denen a-priori große Variation zu erwarten ist.

Wir bemerken, dass sich lokale Adaptivität (im Ort) als eine derartige Gittergradierung interpretieren lässt.

Beispiel 3.3: (Interpolation auf unendlichen Intervallen)

Anstatt eine Funktion $f : \mathbb{R} \to \mathbb{R}$ auf $\Omega = \mathbb{R}$ (unendliches Intervall) zu interpolieren, wählt man als Transformation die Inverse der kumulativen Normalverteilung $\phi^{-1} : \mathbb{R} \to (0, 1)$ und interpoliert

$$\hat{f}(x) := f \circ \phi^{-1}(x)$$

auf $\hat{\Omega} := (0, 1)$.

3.1.2 Orthogonale Transformationen

Bei den *orthogonalen Abbildungen* handelt es sich um eine Teilmenge von $\mathrm{Diff}(\mathbb{R}^d)$ – also den differenzierbaren Bijektionen des Ganzraumes $\mathbb{R}^d$ auf sich selbst, mit der wir uns in dieser Arbeit besonders auseinandersetzen werden.

Definition 3.4. (Orthogonale Matrizen)
Eine quadratische Matrix $\boldsymbol{Q} \in \mathbb{R}^{d\times d}$ heißt *orthonormal*, wenn ihre Spalten-, bzw. Zeilenvektoren eine Orthonormalbasis des $\mathbb{R}^d$ bilden, also

$$\boldsymbol{Q}^t\,\boldsymbol{Q} = \boldsymbol{I}.$$

Dies impliziert, dass alle Spalten von $\boldsymbol{Q}$ paarweise orthonormal zueinander sein müssen:

$$\|\boldsymbol{Q}_{\cdot i}\|_2 = 1 \text{ und } \langle \boldsymbol{Q}_{\cdot i}, \boldsymbol{Q}_{\cdot j}\rangle = 0 \text{ für alle } i \neq j, \tag{3.7}$$

wobei $\boldsymbol{Q}_{\cdot i}$ die i-te Spalte von $\boldsymbol{Q}$ bezeichne.

Die Menge aller d-dimensionalen orthonormalen Matrizen $\boldsymbol{Q}$ mit positiver Determinante bezeichnen wir als die *spezielle orthogonale Gruppe*

$$\mathbf{SO}(d) := \{\boldsymbol{Q} \in \mathbb{R}^{d\times d} : \boldsymbol{Q}^t\,\boldsymbol{Q} = \boldsymbol{I} \text{ und } \det \boldsymbol{Q} = 1\}.$$

Diese ist zusammenhängend innerhalb der Menge aller Matrizen und bildet mit der Matrix-Multiplikation als Verknüpfung eine Untergruppe von $\mathbf{Gl}(d)$. Ihre Elemente beschreiben die Drehungen um insgesamt $\frac{d(d-1)}{2}$ verschiedene Achsen – das sind gerade die ortientierungserhaltenden Elemente der orthogonalen Gruppe $\mathbf{O}(d)$ (siehe etwa [Fis02]).

Wie wir in Abschnitt 3.3 noch weiter ausführen werden, handelt es sich bei der Gruppe $\mathbf{SO}(d)$ um eine Mannigfaltigkeit der Dimension $\frac{d(d-1)}{2}$, d.h. sie lässt sich lokal über einer Teilmenge des $\mathbb{R}^{\frac{d(d-1)}{2}}$ parametrisieren. Somit wird für steigende Dimension d der Aufwand des Optimierungsproblemes deutlich größer, allerdings nur quadratisch und nicht exponentiell, was in gewisser Weise einen Bruch des Fluches der Dimension darstellen würde, falls es gelingt, durch eine Drehung des Koordinatensystemes die exponentielle Abhängigkeit hochdimensionaler Verfahren von der Dimension d zu umgehen.

Wie wir im folgenden Beispiel sehen werden, besitzen die orthogonalen Matrizen zudem die für viele Anwendungen praktische Eigenschaft, dass sie die Produktstruktur des Gauß-Maßes erhalten – genauer gesagt ist das d-dimensionale Gauß-Maß sogar invariant unter jeder orthogonalen Transformation.

Beispiel 3.4: (Orthogonale Transformation eines Gauß-Integrales im $\mathbb{R}^d$)

Die Dichtefunktion $\varphi^d(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}}\ \exp(-\frac{\boldsymbol{x}^t\boldsymbol{x}}{2})$ der multivariaten Standard-Normalverteilung ist wegen

$$(\boldsymbol{Q}\boldsymbol{x})^t(\boldsymbol{Q}\boldsymbol{x}) = \boldsymbol{x}^t\boldsymbol{Q}^t\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}^t\boldsymbol{x} \tag{3.8}$$

invariant unter jeder orthogonalen Transformation $\boldsymbol{Q}$, womit für jedes Integral

$$\int_{\mathbb{R}^d} f(\boldsymbol{x})\,\varphi^d(\boldsymbol{x})\,d\boldsymbol{x} = \int_{\mathbb{R}^d} f(\boldsymbol{Q}\boldsymbol{x})\,\varphi^d(\boldsymbol{x})\,d\boldsymbol{x}$$

gilt.

Das nachfolgende Beispiel soll uns einen ersten Eindruck von der Mächtigkeit orthogonaler Transformationen vermitteln. Denn obwohl diese „nur“ linear sind, ist es damit möglich intrinsisch hochdimensionale Funktionen vermöge oben definierter Drehungen auf eine lediglich eindimensionale Funktion zu reduzieren.

Beispiel 3.5: (Ridge-Funktionen auf rotationssymmetrischen Gebieten)

$K_r := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \leq r\}$ bezeichne die Kugel vom Radius r mit dem Mittelpunkt $\mathbf{0}$. Wegen (3.8) gilt $\|\boldsymbol{Q}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2$, womit alle Elemente $\boldsymbol{Q} \in \mathbf{SO}(d)$ Diffeomorphismen von K_r auf sich selbst sind.

Für $d = 2$ sei etwa $f(\boldsymbol{x}) := \sin(x_1 + x_2)$ eine 2-dimensionale Funktion von K_r nach $\mathbb{R}$. Mit

$$\boldsymbol{Q} := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

ist

$$\begin{aligned} f(\boldsymbol{Q}\boldsymbol{x}) &= \sin(\frac{1}{\sqrt{2}}((x_1 - x_2) + (x_1 + x_2))) \\ &= \sin(\sqrt{2}x_1) \end{aligned}$$

dann nur noch eindimensional.

In Abschnitt 3.6.1 werden wir sehen, dass sich auf rotationssymmetrischen Gebieten sogar alle Ridge-Funktionen der Form $f(\boldsymbol{x}) = g(\sum_{i=1}^d b_i x_i)$ auf eindimensionale Funktionen reduzieren lassen.

Beispiel 3.6: (Interpolation)

Für $f : \mathbb{R}^d \to \mathbb{R}$ sei $f \circ \phi^{-1}$ (siehe Beispiel 3.2) auf $[0,1]^d$ zu interpolieren. Da ϕ^{-1} den Einheitswürfel $[0,1]^d$ diffeomorph auf den $\mathbb{R}^d$ abbildet und $\boldsymbol{Q}$ ein Isomorphismus ist, kann man

$$\hat{f}(\boldsymbol{x}) := f \circ \boldsymbol{Q} \circ \phi^{-1}$$

interpolieren und erhält f durch

$$f(\boldsymbol{x}) = \hat{f}(\boldsymbol{x}) \circ \phi \circ \boldsymbol{Q}^t$$

zurück. Wir werden zeigen, dass dies in vielen Fällen günstiger ist, als $f \circ \phi^{-1}$ oder f direkt zu interpolieren.

3.1.3 Nichtlineare Transformationen

Komponentenweise Abbildungen

In Beispiel 3.2 haben wir bereits eine eindimensionale Koordinatentransformation vorgestellt, die nichtlinear ist. Dieses Prinzip lässt sich auch im Mehrdimensionalen auf jede Achse separat anwenden.

Abbildung 3.4 stellt das bereits in Beispiel 3.2 vorgestellte Konzept einer komponentenweisen Gittergradierung für ein zweidimensionales dünnes Gitter auf $[0.5, 1]^2$ dar. Abgebildet ist ein berandetes dünnes Gitter zum Level 5, das komponentenweise mit $\psi(x) = \sqrt[3]{x}$ gradiert wurde.

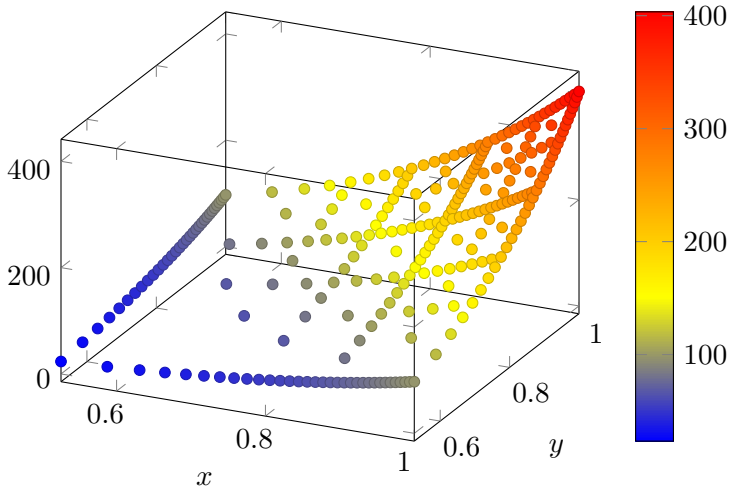


Abb. 3.4: Zweidimensionales, gradiertes dünnes Gitter für die Funktion $\exp(3x + 3y)$.

Wie bereits eingangs erwähnt, sind allgemeinere Diffeomorphismen jedoch schwierig zu diskretisieren, da sie wieder den Fluch der Dimension beinhalten. Dennoch wollen wir im nächsten Abschnitt eine Form von nichtlinearen Transformationen einführen, die dem Konzept der radialen Basen ähnelt.

Stückweise orthogonale Abbildungen

Diesen Ansatz, der auf orthogonalen Abbildungen aufbaut, bezeichen wir als *Stückweise Orthogonale Transformation*. Wie wir in Beispiel 3.5 dargelegt haben, ist eine Matrix $\boldsymbol{Q} \in \mathbf{SO}(d)$ für jedes rotationssymmetrische Gebiet $K_r := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \leq r\}$ stets ein linearer Diffeomorphismus von K_r auf sich selbst.

Die Idee hinter stückweise orthogonale Transformationen ist es nun, ein Gebiet in disjunkte Kreisscheiben $K_{r_1,r_2} := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \in [r_1, r_2)\}$ zu zerlegen

$$\Omega = \bigcup_{i=1}^{N} K_{r_{i-1},r_i}, \text{ mit } r_0 = 0 \tag{3.9}$$

und auf jeder dieser Scheiben eine andere Drehung $\boldsymbol{Q}^{(i)}$ anzuwenden.

Als *stückweise orthogonale Transformation* bezeichnen wir die daraus resultierende Abbildung

$$\hat{\boldsymbol{Q}}_N(\boldsymbol{x}) := \sum_{i=1}^{N} I_{K_{r_{i-1},r_i}}(\boldsymbol{x})\ \boldsymbol{Q}^{(i)}\boldsymbol{x}, \tag{3.10}$$

wobei $I_{K_{r_{i-1},r_i}}$ die Indikatorfunktion zum Kreisring K_{r_{i-1},r_i} bezeichne. Sie ist bijektiv und im Inneren jedes K_{r_{i-1},r_i} differenzierbar. Die Menge der Ränder $R := \bigcup_{i=1}^{N} \partial K_{r_{i-1},r_i}$ ist offensichtlich eine Nullmenge bezüglich des Lebesgue-Maßes.

Da eine orthogonale Matrix $\boldsymbol{Q}$ über $\frac{d(d-1)}{2}$ Freiheitsgrade verfügt, kann man eine stückweise orthogonale Abbildung $\hat{\boldsymbol{Q}}$ mit $N \cdot \frac{d(d-1)}{2}$ Freiheitsgraden darstellen. Für moderate N geht also auch hier die Dimension d nur quadratisch und nicht exponentiell in die Kosten ein.

Satz 3.5. Für alle $\boldsymbol{x} \in \mathbb{R}^d \setminus R$ gilt

$$\det D\hat{\boldsymbol{Q}}_N(\boldsymbol{x}) = 1.$$

Beweis. Für alle $\boldsymbol{x} \in \mathbb{R}^d \setminus R$ gibt es ein $\boldsymbol{Q}_{\boldsymbol{x}} \in \mathbf{SO}(d)$, so dass $\hat{\boldsymbol{Q}}_N(\boldsymbol{x}) = \boldsymbol{Q}_{\boldsymbol{x}}\boldsymbol{x}$ gilt. □

Somit gilt für jedes rotationssymmetrische Gebiet K_r

$$\int_{K_r} f(\boldsymbol{x})\, d\boldsymbol{x} = \int_{K_r} f \circ \hat{\boldsymbol{Q}}(\boldsymbol{x})\, d\boldsymbol{x}$$

und insbesondere für $r = \infty$

$$\int_{\mathbb{R}^d} f(\boldsymbol{x})\, \varphi^d(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbb{R}^d} f \circ \hat{\boldsymbol{Q}}(\boldsymbol{x})\, \varphi^d(\boldsymbol{x})\, d\boldsymbol{x},$$

denn das Gauß-Maß ist invariant unter allen Transformationen, die

$$\|\hat{\boldsymbol{Q}}(\boldsymbol{x})\|_2 = \|\boldsymbol{x}\|_2$$

erfüllen.

Übergang zu kontinuierlich gekrümmten Koordinaten

Die Wahl einer Rotationsmatrix $\boldsymbol{Q}^{(i)}$ auf jedem Kreisring K_{r_{i-1},r_i} lässt sich auch als eine stückweise konstante Abbildung $\boldsymbol{q}_N : \mathbb{R}^+ \to \mathbf{SO}(d)$ auffassen.

Diesen Ansatz wollen wir verallgemeinern, indem wir annehmen, dass eine differenzierbare Abbildung $\boldsymbol{q} : \mathbb{R}^+ \to \mathbf{SO}(d)$ gegeben sei. Damit definieren wir eine Abbildung $\hat{\boldsymbol{Q}} : \mathbb{R}^d \to \mathbb{R}^d$ mit

$$\hat{\boldsymbol{Q}}(\boldsymbol{x}) := \underbrace{\boldsymbol{q}(\|\boldsymbol{x}\|_2^2)}_{\in \mathbb{R}^{d\times d}}\boldsymbol{x}$$

Den nachfolgenden Satz können wir nur für $d = 2, 3$ beweisen – wir vermuten jedoch, dass er auch für $d \geq 4$ gilt.

Satz 3.6. Für alle $\boldsymbol{x} \in \mathbb{R}^2$ gilt

$$\det D\hat{\boldsymbol{Q}}(\boldsymbol{x}) = 1.$$

Beweis. Wir beweisen die Behauptung für $d = 2$. Für $d = 3$ lässt sich unter Verwendung der Euler-Winkel zur Darstellung der Elemente von $\mathbf{SO}(3)$ ein ähnlicher Beweis angeben.

Wir wechseln in Kugelkoordinaten. Der zugehörige Diffeomorphismus $\psi : \Omega \to \mathbb{R}^2$ mit $\Omega := \mathbb{R}^+ \times (0, 2\pi)$ ist durch

$$\psi(r, \theta) := (r\, \cos\theta, r\, \sin\theta)$$

definiert. Die Funktionaldeterminante von ψ ist

$$\det D\psi(r, \theta) = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

Um zu zeigen, dass $\det D(\hat{\Omega} \circ \psi) = r$ gilt (woraus die Behauptung folgt), ist festzustellen, dass für jedes $\boldsymbol{Q} \in \mathbf{SO}(2)$ ein $s \in [0, 2\pi)$ existiert, so dass

$$\boldsymbol{Q} = \begin{pmatrix} \cos s & -\sin s \\ \sin s & \cos s \end{pmatrix}$$

gilt. Damit besitzt $\hat{\boldsymbol{Q}}$ die Darstellung

$$\hat{\boldsymbol{Q}}(\boldsymbol{x}) = \boldsymbol{q}(r^2)\boldsymbol{x} = \boldsymbol{Q} = \begin{pmatrix} \cos g(r^2) & -\sin g(r^2) \\ \sin g(r^2) & \cos g(r^2) \end{pmatrix}, \tag{3.11}$$

wobei $g : \mathbb{R}^+ \to [0, 2\pi)$ eine beliebige differenziere Funktion sei.

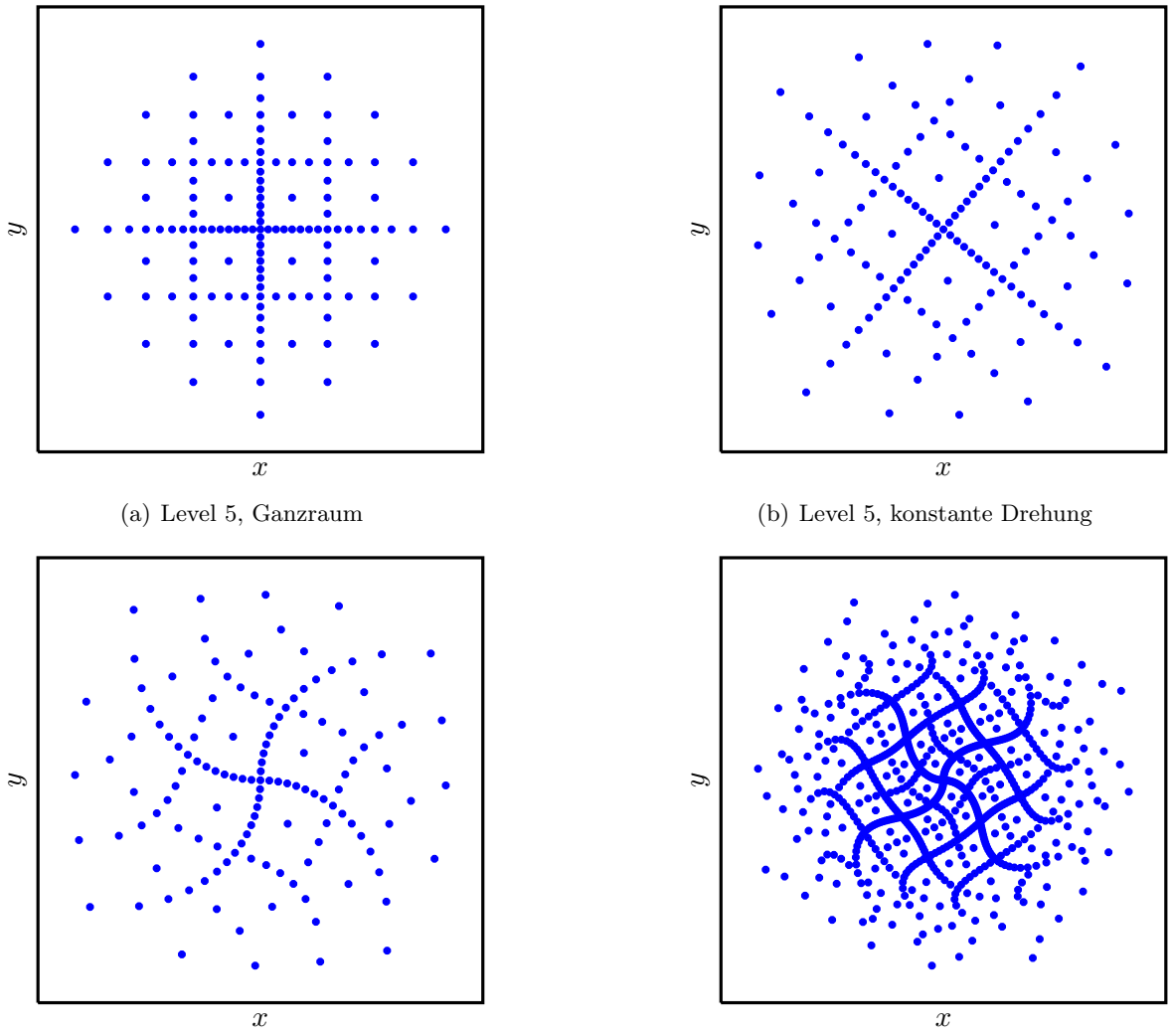
Wir rechnen nun unter Anwendung der Additionstheoreme nach

$$\begin{aligned} \hat{\boldsymbol{Q}} \circ \psi(r, \theta) &= \boldsymbol{q}(r^2) \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos g(r^2) & -\sin g(r^2) \\ \sin g(r^2) & \cos g(r^2) \end{pmatrix} \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} \\ &= \begin{pmatrix} r\cos(\theta + g(r^2)) \\ r\sin(\theta + g(r^2)) \end{pmatrix} \end{aligned}$$

Damit ist

$$\begin{aligned}\det D\hat{\boldsymbol{Q}} \circ \psi(r,\theta) &= \begin{vmatrix}\cos(\theta + g(r^2)) - 2r^2 \sin(\theta + g(r^2))g'(r^2) & -r\sin(\theta + g(r^2)) \\ \sin(\theta + g(r^2)) - 2r^2 \cos(\theta + g(r^2))g'(r^2) & -r\cos(\theta + g(r^2))\end{vmatrix} \\ &= r(\cos^2(\theta + g(r^2)) + \sin^2(\theta + g(r^2))) \\ &= r.\end{aligned}$$

□



(a) Level 5, Ganzraum

(b) Level 5, konstante Drehung

(c) Level 5, stetige Drehung

(d) Level 7, stetige Drehung

Abb. 3.5: Stückweise orthogonale Transformationen eines dünnen Gitters im Ganzraum.

3.2 Das allgemeine Minimierungsfunktional

In diesem Abschnitt werden wir ein Funktional $\mathfrak{M} : \Phi \to \mathbb{R}$ konstruieren, dessen Minimierung über einer Menge $\Phi \subset \mathrm{Diff}(\Omega^{(d)})$ die Reduktion der effektiven Dimension von $f \circ \phi$ formalisiert. Dazu erinnern wir uns an die allgemeine Definition von effektiver Dimension aus Abschnitt 2.2.2

$$d_A = \|f\|_\gamma = \sum_{\mathbf{u} \subseteq \mathcal{D}} \gamma_{\mathbf{u}} \|f_{\mathbf{u}}\|_*.$$

Sei nun $\Phi \subset \mathrm{Diff}(\Omega^{(d)} \to \Omega^{(d)})$ eine Teilmenge der Diffeomorphismen von $\Omega^{(d)}$ auf sich selbst. Der Übersicht halber definieren wir $f^\phi_{\mathbf{u}}(\boldsymbol{x})$ als die ANOVA-Zerlegung von $f^\phi := (f(\phi(\boldsymbol{x}))$ und damit das Funktional

$$\begin{aligned} \mathfrak{M}_f : &\quad \Phi \to \mathbb{R}^+ \\ \mathfrak{M}_f(\phi) &:= \sum_{\mathbf{u} \subseteq \mathcal{D}} \gamma_{\mathbf{u}} \|f^\phi_{\mathbf{u}}\|_*, \end{aligned} \tag{3.12}$$

welches noch von der Norm $\|\cdot\|_*$, dem Produkt-Maß für die ANOVA-Zerlegung μ, den Dimensionsgewichten $\boldsymbol{\gamma}_{\mathbf{u}}$ und der Menge der zulässigen Diffeomorphismen Φ, über der optimiert werden soll, abhängt.

3.2.1 Vorbereitung

Bevor wir uns mit der genauen Wahl dieser Parameter befassen, wollen wir uns die Probleme, die bei der Auswertung eines Funktionales wie $\mathfrak{M}$ auftreten in einem allgemein gehaltenen Rahmen anschauen. Später werden wir uns uns jedoch auf Spezialfälle einschränken.

Dazu betrachten wir zunächst sinnvolle Belegungen der Gewichte $\gamma_{\mathbf{u}}$ und legen die sich daraus ergebenden Schwierigkeiten dar.

Gewichtung der Dimensionen

Von der Wahl der Dimensionsgewichte $\gamma_{\mathbf{u}}$ hängt ab, welche Art von *effektiver Dimension* erfasst und minimiert werden kann.

Wählt man $\boldsymbol{\gamma}$ derart, dass

$$|\mathbf{u}| = |\mathbf{v}| \Rightarrow \gamma_{\mathbf{u}} = \gamma_{\mathbf{v}}$$

so erfasst man die dimensionale Struktur im Superpositionssinne und mit

$$\max(\mathbf{u}) = \max(\mathbf{v}) \Rightarrow \gamma_{\mathbf{u}} = \gamma_{\mathbf{v}}$$

die Struktur im Trunkationssinne.

Für eine Gewichtung der Form

$$\gamma_{\mathbf{u}} = 0 \text{ für } \mathbf{u} \subseteq \{1, \ldots, k\} \text{ und } \gamma_{\mathbf{u}} = 1 \text{ für } \mathbf{u} \nsubseteq \{1, \ldots, k\} \tag{3.13}$$

bedeutet die Minimierung von $\mathfrak{M}$, dass möglichst viel Varianz in den ersten k Trunkationsdimensionen konzentriert werden soll, unabhängig davon, wie die Verteilung in allen anderen Koordinatenrichtungen aussieht. Dies entspricht gerade der Maximierung von $\bar{d}_T$ aus (2.21).

Möchte man hingegen zu einem vorgegebenen $\alpha \in (0, 1]$ die Trunkations- oder Superpositionsdimension $d_T(\alpha)$ minimieren, so löst man nacheinander die Minimierungsprobleme, welche im k-ten Schritt den Gewichten (3.13) entsprechen. k wird solange erhöht, bis

$$d_T(\alpha)(f^\phi) \leq k \tag{3.14}$$

gilt. Damit ist dann k die minimale Trunkationsdimensionen zum Parameter α.

Abschneiden der Summe

Um $\mathfrak{M}(\phi)$ für ein einziges $\phi \in \Phi$ auszuwerten, müssen alle 2^d ANOVA-Terme von f^ϕ aufgestellt und ihre Normen berechnet werden, was aufgrund des Fluches der Dimension selbst für kleine Dimensionen bereits praktisch unmöglich ist. Die Summe im Funktional (3.12) bei einem gewissen $|\mathbf{u}| =: n$ abzuschneiden scheint der kanonische Ausweg – für $n \in \{1, 2, 3\}$ wären die Kosten zur Berechnung des abgeschnittenen Funktionals $\mathfrak{M}_n$ zumindest noch aktzeptabel.

Betrachten wir den Fehler

$$|\mathfrak{M}_n(\phi) - \mathfrak{M}(\phi)| = \sum_{|\mathbf{u}|>n} \gamma_{\mathbf{u}} \, \|f^\phi_{\mathbf{u}}\|_*,$$

den man dabei macht, so stellt man jedoch fest, dass ein Abschneiden bei kleinen n bedeuten würde, dass gerade die hohen ANOVA-Terme nicht mehr erfasst und minimiert werden – also gerade die Beträge zu $\mathfrak{M}$, die mit einem großen Gewicht $\gamma_{\mathbf{u}}$ in die Summe eingehen und daher stark zum Fehler beitragen.

Auf der anderen Seite ist es ebenfalls nicht möglich nur die hohen Terme auszuwerten, da zur Berechnung eines ANOVA-Terms $f_{\mathbf{u}}$ auch alle Terme $f_{\mathbf{v}}$ mit $\mathbf{v} \subset \mathbf{u}$ aufgestellt und ausgewertet werden müssen.

Wir benötigen also ein anderes Funktional, dessen Minimierung äquivalent zu der von $\mathfrak{M}$ ist, welches sich aber einfacher berechnen lässt. Dazu werden wir im nächsten Abschnitt ein äquivalentes Optimierungsproblem $\hat{\mathfrak{M}}$ einführen, dessen Extremstellen unter gewissen Vorraussetzungen identisch zu denen von $\mathfrak{M}$ sind, dessen Koeffizienten $\boldsymbol{\gamma}_{\mathbf{u}}$ jedoch monoton fallend in $|\mathbf{u}|$ sind, wodurch ein Abschneiden der Summe bei kleinem n vertretbar ist.

3.2.2 Das äquivalente Maximierungsproblem

Wir suchen ein Funktional, das die gleichen kritischen Punkte wie $\mathfrak{M}$ besitzt, aber einfacher auszuwerten ist. Wenn die Summe über die $\|f_{\mathbf{u}}\|_*$ beschränkt ist, lässt sich $\mathfrak{M}$ zumindest gegen ein günstigeres Funktional $\hat{\mathfrak{M}}$ abschätzen.

Falls es eine Funktion $C_f : \Phi \to \mathbb{R}^+$ mit

$$\sum_{\mathbf{u} \subseteq \mathcal{D}} \|f_{\mathbf{u}}^{\phi}\|_* \leq C_f(\phi) \text{ für alle } \phi \in \Phi \tag{3.15}$$

gibt, so folgt

$$\|f_{\{1..d\}}^{\phi}\|_* \leq C_f(\phi) - \sum_{|\mathbf{u}|<d} \|f_{\mathbf{u}}^{\phi}\|_*,$$

womit wir $\mathfrak{M}_f(\phi)$ gegen

$$\begin{aligned}
\mathfrak{M}_f(\phi) &= \sum_{\mathbf{u} \subseteq \mathcal{D}} \gamma_{\mathbf{u}} \|f_{\mathbf{u}}^{\phi}\|_* \\
&= \sum_{|\mathbf{u}|<d} \gamma_{\mathbf{u}} \|f_{\mathbf{u}}^{\phi}\|_* + \gamma_{\mathcal{D}} \|f_{\{1..d\}}^{\phi}\|_* \\
&\leq \sum_{|\mathbf{u}|<d} \gamma_{\mathbf{u}} \|f_{\mathbf{u}}^{\phi}\|_* + \gamma_{\mathcal{D}} \left(C_f(\phi) - \sum_{|\mathbf{u}|<d} \|f_{\mathbf{u}}^{\phi}\|_* \right) \\
&= \sum_{|\mathbf{u}|<d} (\gamma_{\mathbf{u}} - \gamma_{\mathcal{D}}) \, \|f_{\mathbf{u}}^{\phi}\|_* + \gamma_{\mathcal{D}} \, C_f(\phi)
\end{aligned}$$

abschätzen können.

Die Koeffizienten $\hat{\gamma}_{\mathbf{u}} := (\gamma_{\mathbf{u}} - \gamma_{\mathcal{D}})$ sind aufgrund der Monotonie von $\gamma_{\mathbf{u}}$ alle negativ und monoton fallend im Betrag. Skalieren wir mit $-\frac{1}{\gamma_{\mathcal{D}}}$, so erhalten wir das *äquivalente Maximierungsfunktional*

$$\begin{aligned}
\hat{\mathfrak{M}}_f(\phi) &:= -\frac{1}{\gamma_{\mathcal{D}}} \left(\sum_{|\mathbf{u}|<d} (\gamma_{\mathbf{u}} - \gamma_{\mathcal{D}}) \, \|f_{\mathbf{u}}^{\phi}\|_* + \gamma_{\mathcal{D}} \, C_f(\phi) \right) \\
&= \sum_{|\mathbf{u}|<d} \left(1 - \frac{\gamma_{\mathbf{u}}}{\gamma_{\mathcal{D}}} \right) \, \|f_{\mathbf{u}}^{\phi}\|_* + C_f(\phi).
\end{aligned}$$

Die Darstellung als Maximierungsproblem entspricht der Intuition, dass „Minimierung der hohen Terme“ auch einer „Maximierung der kleinen Terme“ entsprechen sollte, falls die Gesamtsumme beschränkt ist.

Für einen Spezialfall wird aus der Ungleichung in (3.15) sogar Gleichheit, nämlich wenn wir $\|\cdot\|_*$ als die durch das Skalarprodukt $(\cdot,\cdot)_\mu$ induzierte Norm

$$\|f_{\mathbf{u}}^{\mu}\|_* = \|f_{\mathbf{u}}\|_{2,\mu} = \sigma_{\mathbf{u},\mu} = \sqrt{\int_{\mathbb{R}^d} f_{\mathbf{u}}^{\mu}(\boldsymbol{x}_{\mathbf{u}})^2 \, d\mu_{\mathbf{u}}(\boldsymbol{x})}$$

wählen, wobei μ zugleich das Maß ist, welchem die ANOVA-Zerlegung $f_{\mathbf{u}}^{\mu}$ zugrunde liegt. Wegen

der Orthogonalität (Satz 2.9) folgt

$$C_f(\phi) := \sum_{\mathbf{u}\subseteq\mathcal{D}} \|f_{\mathbf{u}}^{\phi}\|_2^{\mu} = \|f^{\phi}\|_2^{\mu}$$

womit dann

$$\arg\min_{\phi\in\Phi} \mathfrak{M}_f(\phi) = \arg\max_{\phi\in\Phi} \hat{\mathfrak{M}}_f(\phi) \tag{3.16}$$

gilt.

Für den Fall, dass $\Omega^{(d)} = \mathbb{R}^d$, μ das d-dimensionale Gauß-Maß und $\Phi = \mathbf{SO}(d)$ ist, ist

$$C_f(\phi) = \sigma^2(f^{\phi}) = \sigma^2(f)$$

sogar konstant in ϕ, womit sich das normierte Maximierungsfunktional

$$\begin{aligned} \hat{\mathfrak{M}}: &\quad \mathbf{SO}(d) \to \mathbb{R} \\ \hat{\mathfrak{M}}(\boldsymbol{Q}) &:= \frac{1}{\sigma^2(f)} \sum_{|\mathbf{u}|<d} \left(1 - \frac{\gamma_{\mathbf{u}}}{\gamma_{\mathcal{D}}}\right) \sigma_{\mathbf{u}}^2(f^{\boldsymbol{Q}}) \end{aligned} \tag{3.17}$$

ergibt.

3.2.3 Reduktion der Trunkationsdimension

Eine weitere Vereinfachung ergibt sich, wenn man die Trunkationsdimension d_T reduzieren möchte. Bei einer Belegung der $\gamma_{\mathbf{u}}$ wie in (3.13) ergeben sich für die Gewichte $\hat{\gamma}_{\mathbf{u}}$ im äquivalenten Maximierungsfunktional $\hat{\mathfrak{M}}$ die Werte

$$\hat{\gamma}_{\mathbf{u}} = 1 \text{ für } \mathbf{u} \subseteq \{1,\dots,p\} \text{ und } \hat{\gamma}_{\mathbf{u}} = 0 \text{ für } \mathbf{u} \nsubseteq \{1,\dots,p\}. \tag{3.18}$$

Nun folgt wegen

$$P_{\{1\dots p\}}(f)(x_1,\dots,x_p) = \sum_{\mathbf{u}\subseteq\{1\dots p\}} f_{\mathbf{u}}^{\boldsymbol{Q}}(\boldsymbol{x}_{\mathbf{u}}),$$

dass

$$\begin{aligned} \hat{\mathfrak{M}}(\boldsymbol{Q}) = \sum_{\mathbf{u}\subseteq\mathcal{D}} \hat{\gamma}_{\mathbf{u}}\ \sigma_{\mathbf{u}}^2(f^{\boldsymbol{Q}}) &= \sum_{\mathbf{u}\subseteq\{1\dots p\}} \sigma_{\mathbf{u}}^2(f^{\boldsymbol{Q}}) \\ = \| \sum_{\mathbf{u}\subseteq\{1\dots p\}} f_{\mathbf{u}}^{\boldsymbol{Q}} \|_2 &= \|P_{\{1\dots p\}}(f^{\boldsymbol{Q}})\|_2 \end{aligned}$$

gilt.

Da $P_{\{1\dots p\}}(f^{\boldsymbol{Q}})$ nur von den Variablen $x_1,\dots,x_p$ abhängt, genügt es, diejenigen Einträge der Matrix $\boldsymbol{Q}$ zu betrachten, die mit diesen Komponenten von $\boldsymbol{x}$ multipliziert werden. Dabei handelt es sich offensichtlich gerade um die ersten p Spalten von $\boldsymbol{Q}$.

Betrachtet man – etwas allgemeiner – ein abgeschnittenes $\hat{\mathfrak{M}}$ der Form

$$\hat{\mathfrak{M}}_k(\boldsymbol{Q}) = \sum_{\mathbf{u}\subseteq\{1\ldots p\}} \hat{\gamma}_{\mathbf{u}} \, \|f_{\mathbf{u}}^{\boldsymbol{Q}}\|_2,$$

so stellen wir ebenfalls fest, dass ebenfalls nur die ersten p Spalten relevant für $\hat{\mathfrak{M}}$ sind, da nur diejenigen ANOVA-Terme $f_{\mathbf{u}}^{\boldsymbol{Q}}$ von der Summe durchlaufen werden, deren Eingansvariablen in $\{1\ldots p\}$ enthalten sind.

Während die Optimierung von $\hat{\mathfrak{M}}$ im Superpositionssinne also alle d Spalten einer orthogonalen Matrix $\boldsymbol{Q}$ in Betracht zieht und daher über ganz $\mathbf{SO}(d)$ erfolgen muss, genügt es sich im Falle der Trunkationsdimension auf die ersten p Spalten einzuschränken. Diese lassen sich ebenfalls mit der Struktur einer differenzierbaren Mannigfaltigkeit versehen – der so gennannten *Stiefel-Mannigfaltigkeit* $\mathbf{St}(p, d)$.

Im nächsten Abschnitt dieses Kapitels werden wir uns dann mit Optimierungsverfahren auf diesen beiden Mannigfaltigkeiten auseinandersetzen. Zuvor wollen wir jedoch noch abschließend bemerken, dass wir zwar festgestellt haben, dass die Kombination „Gauß-Maß auf dem $\mathbb{R}^d$ mit der Menge der zulässigen Diffeomorphismen $\Phi = \mathbf{SO}(d)$“ einige Vorzüge und Vereinfachungen bietet, es aber durchaus noch eine Reihe weiterer sinnvoller Möglichkeiten gibt, das Funktional $\mathfrak{M}$, bzw. $\hat{\mathfrak{M}}$ mit passenden Normen $\|\cdot\|_*$, ANOVA-Zerlegungen (durch das Maß μ definiert) und Dimensionsgewichten $\gamma_{\mathbf{u}}$ auf einer Menge $\Phi \subset \mathrm{Diff}(\Omega^{(d)})$ zu definieren. Je nach Anwendung sind hier andere Parameter sinnvoll, was jedoch im Allgemeinen auf eine komplexere Numerik führt.

3.3 Optimierung auf der Mannigfaltigkeit $\mathbf{SO}(d)$

In diesem Abschnitt werden wir einen iterativen Algorithmus zur Minimierung von Funktionalen $F : \mathbf{SO}(d) \to \mathbb{R}$ entwickeln. Obwohl $\mathbf{SO}(d)$ eine Teilmenge des Vektorraumes $\mathbb{R}^{d\times d}$ darstellt, lassen sich die gängigen Methoden für die Optimierung über Vektorräumen nicht ohne Weiteres auf unseren Fall übertragen, da die Mannigfaltigkeit $\mathbf{SO}(d)$ kein linearer Unterraum von $\mathbb{R}^{d\times d}$ ist.

Dies ist in Abb. 3.6 dargestellt, wo eine ein- und eine zweidimensionale Mannigfaltigkeit – jeweils eingebettet in den $\mathbb{R}^3$ – zu sehen sind. Bewegt man sich, ausgehend von einem Punkt auf diesen Mannigfaltigkeiten, entlang eines beliebigen Richtungsvektors, so befindet man sich zwar noch im umgebenden Vektorraum, jedoch im Allgemeinen nicht mehr auf der Mannigfaltigkeit selbst.

Dennoch betrachten wir zunächst iterative Liniensuch-Verfahren zur Optimierung über Vektorräumen um deren Konzepte dann später auf nichtlineare Mannigfaltigkeiten zu übertragen. Diese basieren darauf, dass man im k-ten Iterationsschritt eine Suchrichtung $\boldsymbol{d}^{(k)}$ wählt und sich entlang dieser um eine Schrittweite $\alpha^{(k)}$ bewegt, bis eine angemessene Verringerung der Kosten-Funktion F erzielt wurde. Beispiele für solche Verfahren sind die Methode des steilsten Abstiegs, das Newton-Verfahren und die verschiedenen Methoden der konjugierten Gradienten (CG-Verfahren nach Polak-Ribiere, Fletcher-Powell, Hestenes und Stiefel, etc).

Diese Verfahren übertragen wir dann mittels diverser Konzepte aus der Differentialgeometrie auf nichtlineare Mannigfaltigkeiten, was jedoch noch immer Gegenstand aktueller Forschung ist, deren gegenwärtiger Stand umfassend in [AMS08] dargelegt wird.

Ziel dieses Abschnitts soll es nun sein, ein nichtlineares CG-Verfahren für die Minimierung auf $\mathbf{SO}(d)$ zu entwickeln, welches mit möglichst wenigen Funktionsauswertungen auskommt, da diese bei unserem Kostenfunktional $\hat{\mathfrak{M}}$ ausgesprochen teuer sind.

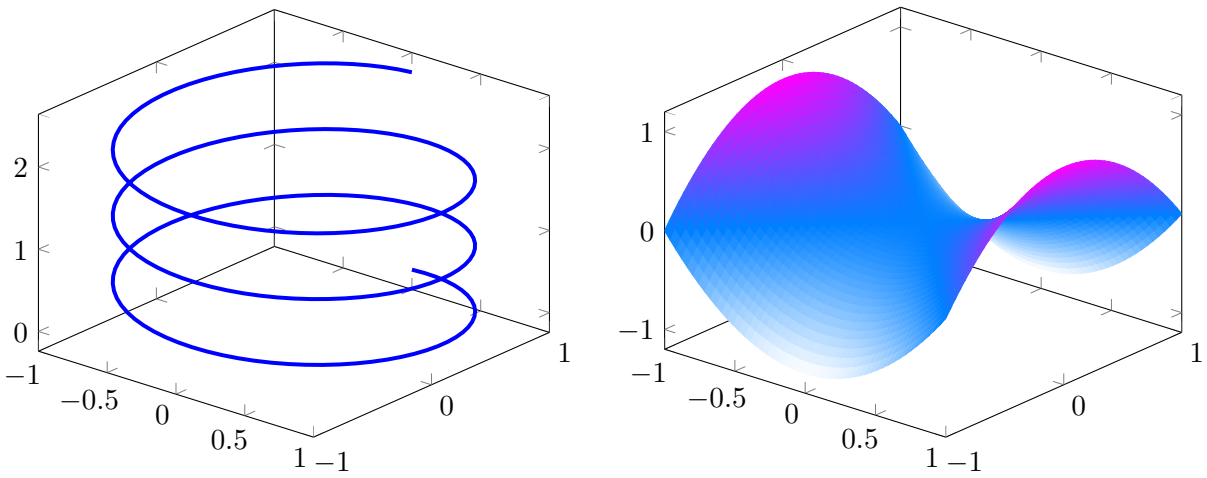


Abb. 3.6: Nichtlineare ein- und zweidimensionale Mannigfaltigkeiten – beide in den $\mathbb{R}^3$ eingebettet

Schränkt man sich bei der Minimierung auf die ersten p Spalten von $\mathbf{SO}(d)$ ein, so reduziert sich das Problem auf eine Minimierung über der Stiefelmannigfaltigkeit $\mathbf{St}(p, d)$ (vgl. Definition 3.9).

Dazu übertragen wir die notwendigen Begriffe aus der Optimierung über Vektorräumen in den Kontext eingebetteter Untermannigfaltigkeiten und diskutieren einige Besonderheiten der Mannigfaltigkeiten $\mathbf{SO}(d)$ und $\mathbf{St}(p, d)$.

3.3.1 Minimierung in Vektorräumen

In einem zum $\mathbb{R}^d$ isomorphen Vektorraum $\mathcal{E} \simeq \mathbb{R}^d$ lautet die allgemeine Iterationsvorschrift für Liniensuchverfahren zur Minimierung einer Funktion $F : \mathcal{E} \to \mathbb{R}$:

$$\boldsymbol{x}^{(k+1)} := \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{d}^{(k)}. \tag{3.19}$$

Durch die konkrete Wahl der Suchrichtung $\boldsymbol{d}^{(k)}$ und der Schrittweite $\alpha^{(k)}$ in jedem Interationsschritt ergeben sich dann die gängigen Verfahren. Einige Beispiele für die Wahl der Abstiegsrichtungen $\boldsymbol{d}^{(k)}$ sind im *Verfahren des steilsten Abstieges*

$$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}), \tag{3.20}$$

im *Newton-Verfahren*

$$\boldsymbol{d}^{(k)} = -\mathrm{Hess}_f(\boldsymbol{x}^{(k)})^{-1} \nabla f(\boldsymbol{x}_k^{(k)}) \tag{3.21}$$

und in der *Methode der konjugierten Gradienten* nach *Fletcher–Reeves*

$$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) + \frac{\|\nabla f(\boldsymbol{x}^{(k)})\|^2}{\|\nabla f(\boldsymbol{x}^{(k-1)})\|^2} \boldsymbol{d}^{(k-1)}. \tag{3.22}$$

Die Schrittweite α, um welche man sich in die Richtung $\boldsymbol{d}$ bewegt, wird durch eine Liniensuchmethode bestimmt, über die wir im Folgenden einen kurzen Überblick geben wollen.

Liniensuche

Ist ausgehend von einem Punkt $\boldsymbol{x}$ eine Abstiegsrichtung $\boldsymbol{d}$ gegeben, so gibt es die verschiedensten Möglichkeiten eine geeignete Schrittweite α zu bestimmen. An dieser Stelle wollen wir die drei gebräuchlichsten Möglichkeiten angeben – welche man letztlich verwendet, ist eine Frage des zugrundeliegenden Problems – insbesondere das Verhältnis der Kosten für Funktionsauswertungen zu Gradientenauswertungen spielt eine Rolle, bei der Entscheidung, wie exakt man eine Liniensuche durchführen möchte.

Exakte Liniensuche: Bestimme ein exaktes Minimum – etwa durch ein eindimensionales Newtonverfahren entlang des Richtungsvektors $\boldsymbol{d}$ oder durch eine Intervallschachtelungsmethode. Dies ist nur dann zu empfehlen, wenn die Berechnung eines Gradienten substanziell teurer als eine Funktionsauswertung ist.

Approximation durch Polynome: Bestimme ein quadratisches oder kubisches Interpolationspolynom der eindimensionalen Funktion $t \to f(\boldsymbol{x} + t\boldsymbol{d})$ und berechne dessen Minimum analytisch.

Armijo-Bedingungen: Zu einer gegebenen Konstante $c \in (0, 1)$ erfülle die Schrittweite α die Bedingung $F(\boldsymbol{x} + \alpha\boldsymbol{d}) \leq F(\boldsymbol{x}) + c\alpha\boldsymbol{d}^t\nabla f(\boldsymbol{x})$.

Weiterführende Informationen zur Minimierung auf Vektorräumen und Liniensuchverfahren finden sich etwa in [Arm66, Noc92, DS83, Pol97, Fle00].

3.3.2 Minimierung auf eingebetteten Untermannigfaltigkeiten

Es gibt nun verschiedene Herangehensweisen, um solche Verfahren von Vektorräumen auf die Minimierung über einer nichtlinearen Mannigfaltigkeit $\mathcal{M}$ zu übertragen. Für den Fall, dass $\mathcal{M}$ in einen Vektorraum $\mathcal{M} \subset \mathcal{E}$ eingebettet ist, gibt es dazu grundsätzlich drei Möglichkeiten:

- Nebenbedingungen durch Gleichungen formulieren und die Methode der *Lagrange-Multiplikatoren* anwenden. Dies hat den Nachteil, dass man im Vektorraum $\mathcal{E}$ operiert, welcher im Allgemeinen eine größere Dimension als $\mathcal{M}$ besitzt.
- In $\mathcal{E}$ minimieren, jedoch den Abstand zu $\mathcal{M}$ durch einen so genannten *Penalty-Term* bestrafen. Das Minimum der auf diese Art regularisierten Kostenfunktion liegt bei geeigneter Wahl des Straftermes in $\mathcal{M}$. Nachteilig ist auch hierbei, dass man im höherdimensionalen $\mathcal{E}$ operiert. Außerdem verwendet man kein Wissen über die lokale Struktur, bzw. Krümmung von $\mathcal{M}$, welches in vielen Fällen vorhanden ist.
- Man bewege sich nur innerhalb der Mannigfaltigkeit, indem man diese lokal über einer Teilmenge eines Vektorraumes parametrisiert. Dadurch nutzt man implizit Information über die lokale Struktur der Mannigfaltigkeit aus. Von Nachteil ist hierbei jedoch, dass die Wahl geeigneter Parametrisierungen schwierig und deren Berechnung numerisch aufwändig ist.

Wir betrachten letztere Methode, denn wie wir sehen werden, ist in dem uns betreffenden Fall $\mathcal{M} = \mathbf{SO}(d)$ eine lokale Darstellung als $\frac{d(d-1)}{2}$-dimensionales Funktional auf dem Tangentialraum von $\mathbf{SO}(d)$ möglich. Betrachtet man nur die ersten p Spalten von $\boldsymbol{Q} \in \mathbf{SO}(d)$, so ist $\mathcal{M} = \mathbf{St}(p, d)$ und der Tangentialraum nur $(dp - \frac{p(p+1)}{2})$-dimensional – also linear abhängig von der Dimension d.

Um nun Methoden des Typs (3.19) von Vektorräumen $\mathcal{E}$ auf Mannigfaltigkeiten $\mathcal{M}$ zu verallgemeinern, ist es notwendig, die Begriffe Richtungsvektor, die Bewegung entlang dieser und die Translation von Richtungen (Parallelverschiebung) auf eingebettete Untermannigfaltigkeiten zu übertragen.

Kurz gesagt geschieht dies, indem die Abstiegsrichtung $\boldsymbol{d}^{(k)}$ als Tangentialvektor an $\mathcal{M}$ im Punkt $\boldsymbol{x}^{(k)}$ aufgefasst wird. Die Liniensuche zur Bestimmung der Schrittweite vollzieht man dann entlang einer Kurve $\gamma(t)$ innerhalb von $\mathcal{M}$, deren Tangentialvektor in $t = 0$ gerade $\boldsymbol{d}^{(k)}$ ist. Die Parallelverschiebung von Richtungsvektoren überträgt sich auf Mannigfaltigkeiten durch das Konzept des *riemannschen Zusammenhangs* (auch Levi-Civita-Zusammenhang), welcher

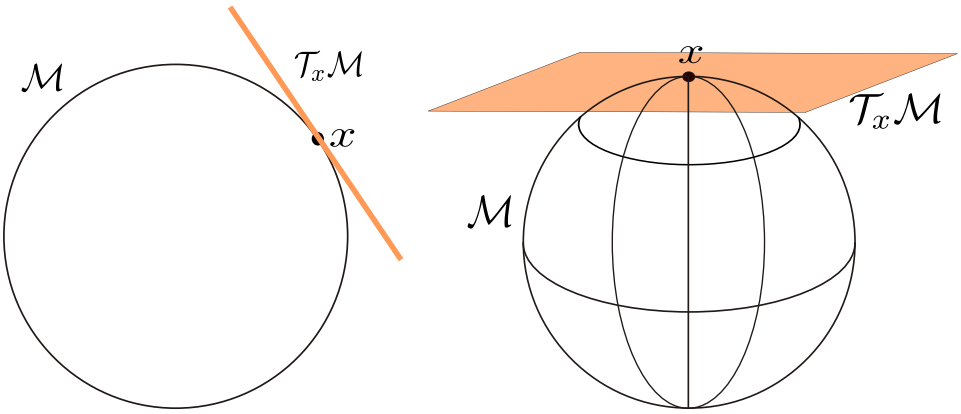


Abb. 3.7: Beispiele für Tangentialräume an die 1- und an die 2-Sphäre

verschiedene Tangentialräume zueinander in Beziehung setzt. Eine andere Möglichkeit Richtungsvektoren über Tangentialräume hinweg zu transportieren, stellt das allgemeinere Konzept des Vektortransports aus [AMS08] dar.

Die für das weitere Vorgehen nötigen Begriffe werden wir im Folgenden lediglich zusammenfassen, da eine gründliche Einführung in die zugrundeliegenden Konzepte der Differentialgeometrie an dieser Stelle zu weit führen würde.

Eine speziell auf Matrix-Mannigfaltigkeiten zugeschnittene Einführung in die Differentialgeometrie findet sich in den ersten Kapiteln von [AMS08]. Allgemeineres zu riemannschen Mannigfaltigkeiten und deren Geometrie lässt sich in [Lee97] und [dC92] nachlesen. Ein weiterer geometrischer Zugang zur Optimierung auf der Stiefel- und Grassman-Mannigfaltigkeit findet sich in [EAS+98].

Der Tangentialraum

Der grundlegendste Unterschied zur Minimierung in Vektorräumen ist eine andere Vorstellung von „Richtungen“. Denn während in $\mathcal{E}$ ein Richtungsvektor $\boldsymbol{d}$ auch selbst ein Element des Vektorraumes ist und man sich daher ausgehend von einem Punkt $\boldsymbol{x}$ in seine Richtung bewegt, indem man für eine Schrittweite α den neuen Punkt $\boldsymbol{x} + \alpha\boldsymbol{d}$ berechnet, definiert man über einer Mannigfaltigkeit $\mathcal{M}$ eine Richtung als ein Element des *Tangentialraumes* $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$. Dieser Vektorraum ist am Punkt $\boldsymbol{x}$ an $\mathcal{M}$ „angeheftet“ und enthält alle *Tangentialvektoren* an $\mathcal{M}$, die durch $\boldsymbol{x}$ gehen – das ist gerade die Menge

$$\mathcal{T}_{\boldsymbol{x}}\mathcal{M} = \{\gamma'(0) : \gamma \text{ ist } C^1\text{-Kurve in } \mathcal{M} \text{ mit } \gamma(0) = \boldsymbol{x}\}. \tag{3.23}$$

Der Tangentialraum ist stets ein Vektorraum und besitzt, wie in Abbildung 3.7 am Beispiel der in den $\mathbb{R}^2$ eingebetteten 1-Sphäre und der in den den $\mathbb{R}^3$ eingebetteten 2-Sphäre zu sehen ist, die gleiche Dimension wie $\mathcal{M}$. Die Menge aller Tangentialräume $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ zu Punkten $\boldsymbol{x} \in \mathcal{M}$ wird als *Tangentialbündel* $T\mathcal{M}$ bezeichnet.

Genau wie man sich das Differential einer Abbildung als deren lokale Approximation durch eine lineare Funktion vorstellen kann, ist der Tangentialraum die „lokale Approximation von $\mathcal{M}$ an der Stelle $\boldsymbol{x}$ durch einen Vektorraum“.

Um sich nun, ausgehend von $\boldsymbol{x} \in \mathcal{M}$, in Richtung $\boldsymbol{d} \in \mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ zu bewegen ohne die Mannigfaltigkeit zu verlassen, bewegt man sich entlang einer Kurve $\gamma : [-c, c] \to \mathcal{M}$ mit $\gamma(0) = \boldsymbol{x}$ und $\gamma'(0) = \boldsymbol{d}$. Solche Kurven (die nicht notwendigerweise eine Geodätische sein müssen) stellen die Verallgemeinerung einer „geraden Linie“ auf gekrümmte Räume dar. Für deren Berechnung wird eine Abbildung $\mathcal{T}_{\boldsymbol{x}}\mathcal{M} \to \mathcal{M}$ benötigt, welche wir im folgenden Abschnitt definieren werden.

Retraktionen

Das Konzept der *Retraktion* bietet die Möglichkeit, die Mannigfaltigkeit $\mathcal{M}$ lokal in einer Umgebung von $\boldsymbol{x} \in \mathcal{M}$ durch eine Abbildung

$$\mathcal{R}_{\boldsymbol{x}} : \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \to \mathcal{M}$$

darzustellen. Sie stellt damit eine Verallgemeinerung der Riemannschen Exponentialabbildung dar.

Definition 3.7. (Retraktionen)

Eine *Retraktion* $\mathcal{R}$ ist eine differenzierbare Abbildung vom Tangentialbündel $T\mathcal{M}$ auf die Mannigfaltigkeit $\mathcal{M}$, für die mit $R_{\boldsymbol{x}} := R_{|\mathcal{T}_{\boldsymbol{x}}\mathcal{M}}$ gilt

1. $\mathcal{R}_{\boldsymbol{x}}(\mathbf{0}) = \boldsymbol{x}$
2. $\mathrm{D}\mathcal{R}_{\boldsymbol{x}}(\mathbf{0}) = \mathrm{Id}_{\mathcal{T}_{\boldsymbol{x}}\mathcal{M}}$,

wobei wir mit $\mathrm{Id}_{\mathcal{T}_{\boldsymbol{x}}\mathcal{M}}$ die Identität des Tangentialraums $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ bezeichnen.

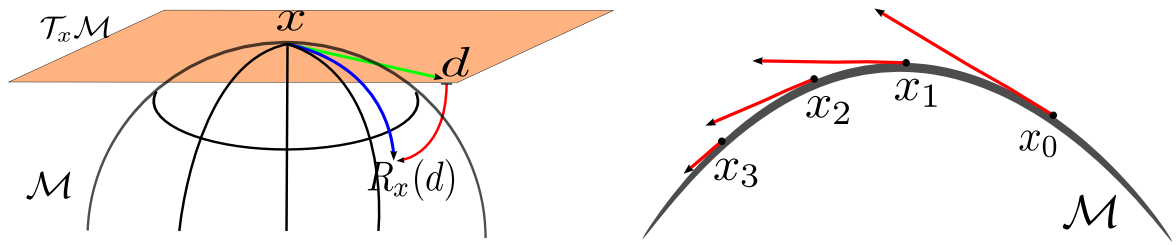
Ist $\mathcal{M}$ in einen Vektorraum $\mathcal{E}$ eingebettet, so ist $\mathcal{T}_{\boldsymbol{x}}\mathcal{M} \subset \mathcal{T}_{\boldsymbol{x}}\mathcal{E} \cong \mathcal{E}$.
(Genauer: $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ ist ein linearer Unterraum von $\mathcal{T}_{\boldsymbol{x}}\mathcal{E}$, der wiederum zu $\mathcal{E}$ isomorph ist).

Wir erhalten somit in einer Umgebung von $\boldsymbol{x}$ eine Darstellung der Kostenfunktion F über dem Vektorraum $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ anstatt auf der Mannigfaltigkeit $\mathcal{M}$ selbst. Damit lassen sich dann die bekannten Konzepte aus Vektorräumen – zumindest lokal – wieder verwenden.

Die klassische Exponentialabbildung erfüllt offensichtlich die Definition einer Retraktion, allerdings ist diese auf den Mannigfaltigkeiten, die uns in dieser Arbeit beschäftigen, nur mit großem numerischem Aufwand zu berechnen. Wir wollen daher auch eine andere Klasse von Retraktionen angeben, die numerisch etwas günstiger sind.

Retraktion durch Zerlegung

Wie bereits bemerkt, lässt sich der Tangentialraum $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ einer in einen Vektorraum $\mathcal{E}$ eingebetteten Untermannigfaltigkeit $\mathcal{M}$ als linearer Unterraum von $\mathcal{T}_{\boldsymbol{x}}\mathcal{E} \simeq \mathcal{E}$ betrachten.



(a) Retraktion (b) Vektortransport

Abb. 3.8: Retraktion und Vektortransport

Dies motiviert eine Retraktion nach folgendem Prinzip:

1. Ausgehend von $\boldsymbol{x} \in \mathcal{M}$ bewege man sich *innerhalb von* $\mathcal{E}$ in Richtung $\boldsymbol{d} \in \mathcal{T}_{\boldsymbol{x}}\mathcal{M}$, indem man $(\boldsymbol{x} + \alpha\boldsymbol{d}) \in \mathcal{E}$ berechnet.
2. Man „projiziere“ den Punkt $\boldsymbol{x} + \alpha\boldsymbol{d}$ zurück auf die Mannigfaltigkeit $\mathcal{M}$.

Eine Retraktion nach Definition 3.7 liefert dann der folgende Satz, welcher auf den ersten Blick wenig anschaulich wirken mag, später jedoch noch eine wichtige Rolle spielen wird.

Satz 3.8. (Retraktion durch Zerlegung)

$\mathcal{M}$ sei eine eingebettete Untermannigfaltigkeit eines Vektorraums $\mathcal{E}$ und $\mathcal{N}$ eine beliebige Mannigfaltigkeit, so dass $\dim\mathcal{M} + \dim\mathcal{N} = \dim\mathcal{E}$ gelte. Es gebe einen Diffeomorphismus

$$\begin{aligned} \Psi : \mathcal{M} \times \mathcal{N} &\to \mathcal{E}_* \\ (F, G) &\to \Psi(F, G), \end{aligned}$$

wobei $\mathcal{E}_* \subset \mathcal{E}$ eine offene Teilmenge und $I \in \mathcal{N}$ ein „neutrales Element“ sei, so dass

$$\Psi(F, I) = F \quad \text{für alle } F \in \mathcal{M}$$

gelte. Bezeichnen wir nun mit $\pi_1 : \mathcal{M} \times \mathcal{N} \to \mathcal{M}, (F, G) \to F$ die Projektion auf die erste Komponente, so definiert

$$R_{\boldsymbol{x}}(\boldsymbol{d}) := \pi_1(\Psi^{-1}(\boldsymbol{x} + \boldsymbol{d}))$$

eine Retraktion auf $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$.

Beweis. Bei der Behauptung handelt es sich um Proposition 4.1.2 aus [AMS08], wo sich auch ein Beweis findet. □

Wir werden diesen Satz später benutzen, um Retraktionen durch Matrix-Zerlegungen, wie etwa der QR- oder Polar-Dekomposition, zu definieren. Das „neutrale Element“ entspricht dann der Einheitsmatrix und der Diffeomorphismus Ψ der Matrix-Multiplikation.

Gradienten

Der Gradient $\nabla F(\boldsymbol{x})$ einer auf $\mathcal{M}$ definierten Funktion F ist das Element aus $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$, das in die Richtung des steilsten Anstieges von F zeigt. In der Literatur wird häufig empfohlen, diesen vermöge der Kettenregel, also durch

$$\nabla_{F\circ\mathcal{R}_{\boldsymbol{x}}}(\mathbf{0}) = \nabla F(\mathcal{R}_{\boldsymbol{x}}(\mathbf{0}))\, D_{\mathcal{R}_{\boldsymbol{x}}}(\mathbf{0}) \tag{3.24}$$

zu berechnen, da das Differential der Retraktion $D_{\mathcal{R}_{\boldsymbol{x}}}(\mathbf{0})$ oft analytisch gegeben, die Auswertung der Retraktion selbst jedoch teuer ist.

Dabei ist aber zu beachten, dass ∇F in $\mathcal{T}_{\boldsymbol{x}}\mathcal{E} \simeq \mathcal{E}$ liegt, während $\nabla_{F\circ\mathcal{R}_{\boldsymbol{x}}}$ zu $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ gehört, der eine geringere Dimension als $\mathcal{E}$ besitzt.

Ist die Auswertung von F also teurer als die von $\mathcal{R}_{\boldsymbol{x}}$ (was in unserer Anwendung der Fall sein wird), so empfiehlt es sich, den Gradienten direkt (ohne Anwendung der Kettenregel) zu berechnen, da aufgrund der kleineren Dimension des Tangentialraums von $\mathcal{M}$ weniger Richtungsableitungen bestimmt werden müssen.

Vektortransport

Beim CG-Verfahren (3.22) wird die neue Suchrichtung $\boldsymbol{d}^{(k)} \in \mathcal{T}_{\boldsymbol{x}^{(k)}}\mathcal{M}$ aus dem aktuellen Gradienten $\nabla f(\boldsymbol{x}^{(k)}) \in \mathcal{T}_{\boldsymbol{x}^{(k)}}\mathcal{M}$ und aus dem Gradienten des letzten Schrittes $\nabla f(\boldsymbol{x}^{(k-1)}) \in \mathcal{T}_{\boldsymbol{x}^{(k-1)}}\mathcal{M}$ durch Linearkombination gewonnen. Doch $\nabla f(\boldsymbol{x}^{(k-1)})$ ist kein Element von $\mathcal{T}_{\boldsymbol{x}^{(k)}}\mathcal{M}$.

Während man in Vektorräumen einen Richtungsvektor durch den Raum transportiert, indem man einfach seinen Anfangspunkt verschiebt, ist auf einer Mannigfaltigkeit eine Abbildung

$$\tau : \mathcal{T}_{\boldsymbol{x}^{(k-1)}}\mathcal{M} \to \mathcal{T}_{\boldsymbol{x}^{(k)}}\mathcal{M},$$

der so genannte *Parallel-Transport* bezüglich eines affinen Zusammenhangs nötig. Auf riemannschen Mannigfaltigkeiten, bietet sich der Levi-Civita-Zusammenhang (auch „riemannscher Zusammenhang“ genannt) an. Dieser ist jedoch in unserem Fall nicht analytisch gegeben und muss unter großem Aufwand numerisch bestimmt werden, weshalb wir einen anderen, in [QGA10] und [AMS08] vorgeschlagenen Weg beschreiten wollen.

Ist eine Mannigfaltigkeit $\mathcal{M}$ in einen Vektorraum $\mathcal{E}$ eingebettet, so sind sämtliche Tangentialräume an $\mathcal{M}$ affine Unterräume von $\mathcal{T}_{\boldsymbol{x}}\mathcal{E} \simeq \mathcal{E}$. Daher existiert stets eine Abbildung

$$P_{\boldsymbol{x}^{(k)}} : \mathcal{E} \to \mathcal{T}_{\boldsymbol{x}^{(k)}}\mathcal{M}, \tag{3.25}$$

die jeden Vektor aus $\mathcal{E}$ orthogonal auf $\mathcal{T}_{\boldsymbol{x}^{(k)}}\mathcal{M}$ projiziert. Wir werden feststellen, dass diese Projektion für unsere Anwendung besonders leicht zu berechnen ist.

Algorithm 1: Mannigfaltigkeiten CG-Verfahren – Metaalgorithmus

Vorraussetzungen: eingebettete Riemannsche Mannigfaltigkeit $\mathcal{M}$, reellwertige Funktion F auf $\mathcal{M}$, Retraktion $\mathcal{R}_{\boldsymbol{x}} : \mathcal{T}_{\boldsymbol{x}} \to \mathcal{M}$, Projektion $P_{\boldsymbol{x}} : \mathcal{E} \to \mathcal{T}_{\boldsymbol{x}}\mathcal{M}$, Abbruchbedingung $\varepsilon > 0$

Initialisiere: $k = 0$, wähle Startpunkt $\boldsymbol{x}^{(0)} \in \mathcal{M}$, berechne $\boldsymbol{d}^{(0)} := -\nabla F(\boldsymbol{x}^{(0)})$

for $k = 0, 1, 2, \ldots$ **do**

1) Wenn $\|\nabla F(\boldsymbol{x}^{(k)})\|_2 < \varepsilon$: Ausgabe
2) Berechne Schrittweite $\alpha^{(k)}$ durch Liniensuche
3) Setze $\boldsymbol{x}^{(k+1)} := \mathcal{R}_{\boldsymbol{x}^{(k)}}(\alpha^{(k)}\boldsymbol{d}^{(k)})$
4) Setze $\beta^{(k+1)} = \frac{\|\nabla F(\boldsymbol{x}^{(k)})\|^2}{\|\nabla F(\boldsymbol{x}^{(k-1)})\|^2}$
5) Setze $\tau\boldsymbol{d}^{(k)} := P_{\boldsymbol{x}^{(k+1)}}(\boldsymbol{d}^{(k)})$
6) Setze $\boldsymbol{d}^{(k+1)} := -\nabla F(\boldsymbol{x}^{(k+1)}) + \beta^{(k+1)}\tau\boldsymbol{d}^{(k+1)}$

Ausgabe

Konjugierte Gradienten Meta-Algorithmus

Sind

- eine eingebettete Riemannsche Mannigfaltigkeit $\mathcal{M} \subset \mathcal{E}$
- eine reellwertige Funktion F auf $\mathcal{M}$
- eine Retraktion $\mathcal{R}_{\boldsymbol{x}} : \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \to \mathcal{M}$ für alle $\boldsymbol{x} \in \mathcal{M}$
- eine Projektion $P_{\boldsymbol{x}} : \mathcal{E} \to \mathcal{T}_{\boldsymbol{x}}\mathcal{M}$
- eine Abbruchbedingung $\varepsilon > 0$

gegeben, so stellt Algorithmus 1 ein nichtlineares CG-Verfahren zur Minimierung von F über der eingebetteten Mannigfaltigkeit $\mathcal{M}$ dar.

In den folgenden beiden Abschnitten, werden wir nun für die Fälle $\mathcal{M} = \mathbf{SO}(d)$ und $\mathcal{M} = \mathbf{St}(p, d)$ geeignete Projektionen $P_{\boldsymbol{x}}$ und Retraktionen $\mathcal{R}_{\boldsymbol{x}}$ entwickeln, die eine effiziente Maximierung des Funktionals $\hat{\mathfrak{M}}$ gestatten werden.

3.3.3 Ein CG-Verfahren für $\mathbf{St}(p, d)$

In diesem Unterabschnitt stellen wir ein nichtlineares Verfahren der konjugierten Gradienten nach Fletcher–Reeves für die Minimierung von Funktionalen auf der Stiefel-Mannigfaltigkeit $\mathbf{St}(p, d)$ vor. Es ist im wesentlichen aus [AMS08] entnommen, bzw. aus den dort vorgestellten Konzepten zusammengestellt.

Die Stiefel-Mannigfaltigkeit

Definition 3.9. (Stiefel-Mannigfaltigkeit)

Die (orthogonale) Stiefel-Mannigfaltigkeit $\mathbf{St}(p,d)$ zur Dimension d ist für ein $p \leq d$ die Menge aller orthonormalen $d \times p$ Matrizen, also

$$\mathbf{St}(p,d) := \{\boldsymbol{X} \in \mathbb{R}^{d\times p} : \boldsymbol{X}^t\boldsymbol{X} = \mathbf{Id}_p\}, \tag{3.26}$$

wobei $\mathbf{Id}_p$ die p-dimensionale Einheitsmatrix bezeichne.

$\mathbf{St}(p,d)$ ist eine eingebettete Untermannigfaltigkeit der linearen Mannigfaltigkeit $\mathbb{R}^{d\times p}$ und erbt somit deren natürliche Metrik

$$(\boldsymbol{A},\boldsymbol{B})_{\boldsymbol{X}} = \mathrm{Tr}(\boldsymbol{A}^t\boldsymbol{B}) \quad \text{für } \boldsymbol{A},\boldsymbol{B} \in \mathcal{T}_{\boldsymbol{X}}\mathbf{St}(p,d).$$

Differenziert man $\boldsymbol{X}^t\boldsymbol{X} = \mathbf{Id}_p$, so ergibt sich für den Tangentialraum

$$\mathcal{T}_{\boldsymbol{X}}\mathbf{St}(p,d) = \{\boldsymbol{Z} \in \mathbb{R}^{d\times p} :\ \boldsymbol{X}^t\boldsymbol{Z} = -\boldsymbol{Z}^t\boldsymbol{X}\},$$

d.h. $\boldsymbol{X}^t\boldsymbol{Z}$ ist schiefsymmetrisch. Eine äquivalente Charakterisierung ergibt sich durch

$$\mathcal{T}_{\boldsymbol{X}}\mathbf{St}(p,d) = \{\boldsymbol{X}\boldsymbol{S} + (\mathbf{Id} - \boldsymbol{X}\boldsymbol{X}^t)\boldsymbol{A} :\ \boldsymbol{S}^t = -\boldsymbol{S} \text{ und } \boldsymbol{A} \in \mathbb{R}^{d\times p}\},$$

wobei $\boldsymbol{S} \in \mathbb{R}^{p\times p}$ eine schiefsymmetrische Matrix und $\boldsymbol{A} \in \mathbb{R}^{d\times p}$ eine beliebige Matrix seien.

Die orthogonale Projektion auf den Tangentialraum $\mathcal{T}_{\boldsymbol{X}}\mathbf{St}(p,d)$ ist durch

$$P_{\boldsymbol{X}}(\boldsymbol{Y}) = (\mathbf{Id} - \boldsymbol{X}\boldsymbol{X}^t)\boldsymbol{Y} + \boldsymbol{X}\mathrm{skew}(\boldsymbol{X}^t\boldsymbol{Y}) \tag{3.27}$$

definiert, wobei wir mit $\mathrm{skew}(\boldsymbol{A}) := \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^t)$ den schiefsymmetrischen Anteil von $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ bezeichnen.

Desweiteren ist $\mathbf{St}(p,d)$ abgeschlossen und beschränkt in $\mathbb{R}^{d\times p}$, woraus nach Heine-Borell Kompaktheit folgt. Ihre Dimension ist $\dim\mathbf{St}(p,d) = dp - \frac{p(p+1)}{2}$.

Für den Fall $p = 1$ ergibt sich die $(d-1)$-dimensionale Einheitssphäre S^{d-1} und für $p = d$ die $\frac{d(d-1)}{2}$-dimensionale Orthogonale Gruppe $\mathbf{O}(d)$.

Retraktion durch QR-Zerlegung

Mit $\mathbf{qf}(\boldsymbol{A})$ bezeichnen wir die Matrix $\boldsymbol{Q} \in St(p,d)$, welche aus der QR-Zerlegung von $\boldsymbol{A} \in \mathbb{R}^d$ in eine orthonormale Matrix $\boldsymbol{Q} \in \mathbb{R}^{d\times p}$ und eine rechte obere Dreiecksmatrix $\boldsymbol{R} \in \mathbb{R}^{p\times p}$ gewonnen wird.

Damit definieren wir für $\boldsymbol{X} \in \mathbf{St}(p,d)$ die Abbildung $\mathcal{R}_{\boldsymbol{X}} : \mathcal{T}_{\boldsymbol{X}}\mathbf{St}(p,d) \to \mathbf{St}(p,d)$ durch

$$\mathcal{R}_{\boldsymbol{X}}(\boldsymbol{A}) := \mathbf{qf}(\boldsymbol{X} + \boldsymbol{A}), \tag{3.28}$$

Algorithm 2: CG-Verfahren für die Stiefelmannigfaltigkeit

Vorraussetzungen: reellwertige Funktion F auf $\mathbf{St}(p,d)$, Abbruchbedingung $\varepsilon > 0$

Initialisiere:

- Startpunkt $\boldsymbol{X}^{(0)} \in \mathbf{St}(p,d)$
- berechne $\boldsymbol{d}^{(0)} := -\nabla F(\boldsymbol{X}^{(0)}) \in \mathbb{R}^{d\times p}$

for $k = 0, 1, 2, \ldots$ **do**

1) Wenn $\|\nabla F(\boldsymbol{X}^{(k)})\|_2 < \varepsilon$: Ausgabe
2) Berechne Schrittweite $\alpha^{(k)}$ durch Liniensuche
3) Berechne die QR-Zerlegung $(\boldsymbol{X}^{(k)} + \alpha^{(k)}\boldsymbol{d}^{(k)}) = \boldsymbol{QR}$
4) Setze $\boldsymbol{X}^{(k+1)} := \boldsymbol{X}^{(k)}\boldsymbol{Q}$
5) Setze $\beta^{(k+1)} = \frac{\|\nabla F(\boldsymbol{X}^{(k)})\|^2}{\|\nabla F(\boldsymbol{X}^{(k-1)})\|^2}$
6) Setze $\tau\boldsymbol{d}^{(k)} := P_{\boldsymbol{X}^{(k+1)}}(\boldsymbol{d}^{(k)})$ (Projektion nach (3.28))
7) Setze $\boldsymbol{d}^{(k+1)} := -\nabla F(\boldsymbol{X}^{(k+1)}) + \beta^{(k+1)}\tau\boldsymbol{d}^{(k)}$

Ausgabe

die wegen Satz 3.8 eine wohldefinierte Retraktion darstellt, indem man ψ als die Matrix-Matrix-Multiplikation und I als die p-dimensionale Einheitsmatrix $\boldsymbol{I}_p$ wählt.

Setzen wir diese Retraktion nun in den Metaalgorithmus 1 ein, so erhalten wir Algorithmus 2 – ein nichtlineares CG-Verfahren nach Polak-Ribiere zur Minimierung von Funktionalen auf der Mannigfaltigkeit $\mathbf{St}(p,d)$.

Kosten:

Die Zahl der Multiplikationen für einen CG-Schritt nach Algorithmus 2 setzen sich folgendermaßen zusammen:

In jedem Schritt müssen insgesamt $\dim\mathcal{T}_{\boldsymbol{X}}\mathbf{St}(p,d) = dp - \frac{p(p+1)}{2}$ Richtungsableitungen bestimmt werden. Für jede Auswertung der Retraktion (3.28) muss eine QR-Zerlegung berechnet werden, was mit Householder-Reflektionen $\mathcal{O}(p^2d)$ Multiplikationen kostet.

Insgesamt ergibt sich damit für $p \ll d$ ein Aufwand von $\mathcal{O}(d^2p^3)$ und für $p \sim d$ ein Aufwand von $\mathcal{O}(d^5 - d^4)$ für jede CG-Iteration.

3.3.4 Ein CG-Verfahren für $\mathbf{SO}(d)$

Wegen $\mathbf{O}(d) = \mathbf{St}(d,d)$ könnte man das oben beschriebene Verfahren auch für die orthogonale Gruppe verwenden. Allerdings wollen wir hier einen anderen Weg beschreiten, indem wir uns auf $\mathbf{SO}(d) \subset \mathbf{O}(d)$ einschränken, was keinen Verlust darstellt, da hierdurch lediglich Spiegelungen/Reflektionen außer Acht gelassen werden, welche die effektive Dimension nicht beeinflussen.

Nun können wir die Tatsache ausnutzen, dass es sich bei $\mathbf{SO}(d)$ um eine Lie-Gruppe mit zugehöriger Lie-Algebra $\mathfrak{so}(d) = \{\boldsymbol{S} \in \mathbb{R}^{d\times d} : \boldsymbol{S}^t = -\boldsymbol{S}\}$ handelt. Dies gestattet es, die Mannigfaltigkeit lokal über ihrer Lie-Algebra durch die Exponentialabbildung zu parametrisieren, wodurch sich ein $\frac{d(d-1)}{2}$-dimensionales Problem ergibt.

Um die Kosten zur Berechnung des Matrix-Exponentials weiter zu verringern, verwenden wir eine Padé-Approximation, also eine Darstellung als rationales Polynom über $\mathfrak{so}(d)$.

Retraktion durch das Matrix-Exponential

Eine besonders „natürliche“ Retraktion auf riemannschen Mannigfaltigkeiten stellt die Exponentialabbildung $\exp : \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \to \mathcal{M}$ dar [Lee97]. Im Falle von Lie-Gruppen fällt diese mit der gebräuchlichen Definition des Exponentials

$$\exp s := \sum_{k=0}^{\infty} \frac{s^k}{k!}$$

auf der zur Lie-Gruppe gehörenden Lie-Algebra zusammen.

Die zu $\mathbf{SO}(d)$ gehörende Lie-Algebra $\mathfrak{so}(d)$ ist die Menge aller schiefsymmetrischen Matrizen

$$\mathfrak{so}(d) = \{\boldsymbol{S} \in \mathbb{R}^{d\times d} :\ \boldsymbol{S}^t = -\boldsymbol{S}\}.$$

Für jedes $\boldsymbol{S} \in \mathfrak{so}(d)$ ist also $\exp \boldsymbol{S} \in \mathbf{SO}(d)$. Umgekehrt existiert zu jedem $\boldsymbol{Q} \in \mathbf{SO}(d)$ aus einer Umgebung von $\boldsymbol{Q}_0 \in \mathbf{SO}(d)$ ein $\boldsymbol{S} \in \mathfrak{so}(d)$, so dass

$$\boldsymbol{Q} = \boldsymbol{Q}_0 \exp \boldsymbol{S}$$

gilt.

Somit können wir also für $\boldsymbol{Q}_0 \in \mathbf{SO}(d)$ die Retraktion

$$\mathcal{R}_{\boldsymbol{Q}_0}(\boldsymbol{S}) = \boldsymbol{Q}_0 \exp(\boldsymbol{S}), \tag{3.29}$$

definieren.

Die naive Berechnung des Matrix-Exponentials ist jedoch mit sehr hohen Kosten verbunden, weshalb wir uns im folgenden Unterabschnitt mit geeigneten Alternativen befassen werden.

Das Matrix-Exponential schiefsymmetrischer Matrizen

Mit $\boldsymbol{s} \in \mathbb{R}^{\frac{d(d-1)}{2}}$ definieren wir

$$\mathbf{S}(\boldsymbol{s}) = \begin{pmatrix} 0 & s_1 & s_2 & \dots & s_{d-1} \\ -s_1 & 0 & s_d & \dots & \vdots \\ -s_2 & -s_d & \ddots & \dots & \vdots \\ \vdots & \dots & & 0 & s_{\frac{d(d-1)}{2}} \\ -s_{d-1} & \dots & & -s_{\frac{d(d-1)}{2}} & 0 \end{pmatrix} \tag{3.30}$$

und damit das Matrix-Exponential

$$\exp\left(\boldsymbol{S}(\boldsymbol{s})\right) := \sum_{k=0}^{\infty} \frac{\boldsymbol{S}(\boldsymbol{s})^k}{k!}. \tag{3.31}$$

Für $d = 2$ lässt sich dann

$$\exp\begin{pmatrix} 0 & s_1 \\ -s_1 & 0 \end{pmatrix} = \begin{pmatrix} \cos(s_1) & \sin(s_1) \\ -\sin(s_1) & \cos(s_1) \end{pmatrix}$$

und für $d = 3$ die Rodriguez-Formel, mit $\theta = \sqrt{s_1^2 + s_2^2 + s_3^2}$

$$\exp(\mathbf{S}(\boldsymbol{s})) = \mathbf{Id} + \sin(\theta)\ \mathbf{S}(\boldsymbol{s}) + (1 - \cos(\theta)\ \mathbf{S}(\boldsymbol{s})^2) \tag{3.32}$$

zeigen.

Eine verallgemeinerte Rodriguez-Formel für $d \geq 3$ wird in [GX00] entwickelt. Die grundsätzliche Idee besteht in einem Eigenwert-Argument – jedoch ist diese Methode nur von theoretischem Interesse, da die numerische Berechnung durch die von Gallier angegebene Formel ausgesprochen kostspielig ist.

Für den Fall, dass die schiefsymmetrische Matrix $\boldsymbol{S}$ bis auf $\boldsymbol{S}_{i,j} = s$ und $\boldsymbol{S}_{j,i} = -s$ identisch Null ist, ergibt sich jedoch aus [GX00], dass

$$\exp(\boldsymbol{S}(\boldsymbol{s})) = G(i,j)(s_k), \tag{3.33}$$

gilt, wobei wir mit $G(i,j)(s_k)$ die Givensmatrix zum Winkel s_k bezeichen, bei der jeweils nur

die Diagonale und die i-ten und j-ten Zeilen und Spalten belegt sind:

$$G(i,j)(s_k) := \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos(s_k) & \cdots & \sin(s_k) & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -\sin(s_k) & \cdots & \cos(s_k) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

Um das Matrix-Exponential beliebiger schiefsymmetrischer Matrizen zu berechnen, verwenden wir eine aus [CL10] entnommene Technik, die auf der Pade-Approximation und geeigneter Skalierung basiert.

Padé-Approximation des Matrix-Exponentials

Allgemein bezeichnet die Pade-Approximation einer Funktion deren Annäherung durch eine rationale Funktion, also den Quotienten aus einem Polynom P_n vom Grad n und Q_m vom Grad m.

$$\exp(\mathbf{S}) \approx \frac{P_n(\mathbf{S})}{Q_m(\mathbf{S})}$$

Wählt man $m = n$ so ergibt sich nach [CL10] für exp die Darstellung

$$\exp(x) \approx r_m(x) = \frac{p_m(x)}{p_m(-x)}.$$

Mit

$$p_m(x) := \sum_{k=0}^{m} \frac{(2m-k)!m!}{(2m)!k!(m-k)!} x^k$$

gilt für den Fehler in einer Umgebung der $\mathbf{0}$

$$|\exp(x) - r_m(x)| = \mathcal{O}(x^{m+1}).$$

Wegen

$$\exp(\boldsymbol{S}) = \left(\exp\left(\frac{1}{k}\boldsymbol{S}\right)\right)^k$$

kann man $\boldsymbol{S}$ nun geeignet skalieren, so dass man bereits mit einem Polynom siebten Grades eine ausreichend gute Näherung durch die Pade-Approximierende

$$\exp(x) = r_7(x) := \frac{p_7(x)}{p_7(-x)}$$

Algorithm 3: Padé-Approximation des Matrix-Exponentials schiefsymmetrischer Matrizen

Vorraussetzungen: schiefsymmetrische Matrix $\boldsymbol{S} \in \mathfrak{so}(d)$

1) Finde $k \in \mathbb{N}^+ : \frac{\|\mathbf{S}\|_2}{2^k} \leq 1$
2) $B := \frac{1}{2^k}\mathbf{S}$
3) $B_1 := B^2$, $B_2 := B_1 B_1$, $B_3 := B_1 B_2$
4) $P_1 := 17297280\ \mathbf{Id} + 1995840\ B_1 + 25200\ B_2 + 56\ B_3$
5) $P_2 := B(8648640\ \mathbf{Id} + 277200\ B_1 + 1512\ B_2 + B_3)$
6) Berechne die LR-Zerlegung von $(P1 - P2) = \mathbf{LR}$
7) **for** $k = 1, \ldots, d$ **do**
  a) Löse das LGS $\mathbf{R}\boldsymbol{x} = \mathbf{L}^{-1}(P_1 + P_2)_{\cdot,k}$
  b) Setze $r_7(B)_{\cdot,i} := \boldsymbol{x}$
8) $\exp(\mathbf{S}) := r_7(B)^{2^k}$

Ausgabe

erhält, wobei

$$p_7(x) = x^7 + 56x^6 + 1512x^5 + 25200x^4 + 277200x^3 + 1995840x^2 + 8648640x + 17297290$$

bezeichne [CL10].

Kosten:

Für die Berechnung der Matrix-Produkte B_1, B_2, B_3 sind jeweils d^3 Multiplikationen nötig. Die LR-Zerlegung in Schritt 6 läuft in $\mathcal{O}(d^3)$. Vorwärts- und Rückwärts-Substitution in Schritt 7a kostet $\mathcal{O}(d^2)$. Zur Berechnung von $r_7(B)^{2^k}$ sind k Matrix-Multiplikationen nötig, womit sich für Algorithmus 3 ein Gesamtaufwand von ingsesamt $\mathcal{O}((3+k)d^3)$ Multiplikationen ergibt.

Givensmatrizen zur Berechnung der Gradienten

Wie bereits erwähnt, gilt für $\boldsymbol{s} \in \mathbb{R}^{\frac{d(d-1)}{2}}$ mit

$$\boldsymbol{s} = (0, \ldots, 0, s_k, 0, \ldots, 0)$$

die Identität (3.33).

Bei der Berechnung des Gradienten einer Abbildung

$$F(\boldsymbol{s}) := \tilde{F}(\boldsymbol{Q}\exp(\boldsymbol{S}(\boldsymbol{s}))) \quad \text{mit } \tilde{F} : \mathbf{SO}(d) \to \mathbb{R}$$

lässt sich das ausnutzen, da stets nur Richtungsableitungen im Punkt $\mathbf{0}$ zu berechnen sind.

Algorithmus

Wir setzen die Erkenntnisse dieses Abschnitts in den Metaalgorithmus 1 ein und erhalten somit Algorithmus 4.

Algorithm 4: CG-Verfahren für die spezielle orthogonale Gruppe

Vorraussetzungen: reellwertige Funktion F auf $\mathbf{SO}(p,d)$, Abbruchbedingung $\varepsilon > 0$

Initialisiere: wähle Startpunkt $\boldsymbol{Q}^{(0)} \in \mathbf{SO}(p,d)$, berechne $\boldsymbol{d}^{(0)} := -\nabla F(\boldsymbol{Q}^{(0)}) \in \mathbb{R}^{\frac{d(d-1)}{2}}$

for $k = 0, 1, 2, \ldots$ **do**

1) Wenn $\|\nabla F(\boldsymbol{Q}^{(k)})\|_2 < \varepsilon$: Ausgabe

2) Berechne Schrittweite $\alpha^{(k)}$ durch Liniensuche

3) Berechne $\exp(\boldsymbol{S})$ durch Algorithmus 3.

4) Setze $\boldsymbol{Q}^{(k+1)} := \boldsymbol{Q}^{(k)} \exp(\boldsymbol{S})$

5) Setze $\beta^{(k+1)} = \frac{\|\nabla F(\boldsymbol{Q}^{(k)})\|^2}{\|\nabla F(\boldsymbol{Q}^{(k-1)})\|^2}$

6) Setze $\tau\boldsymbol{d}^{(k)} := P_{\boldsymbol{Q}^{(k+1)}}(\boldsymbol{d}^{(k)})$ (Projektion nach (3.28))

7) Setze $\boldsymbol{d}^{(k+1)} := -\nabla F(\boldsymbol{Q}^{(k+1)}) + \beta^{(k+1)}\tau\boldsymbol{d}^{(k)}$

Ausgabe

Kosten:

Für einen CG-Schritt nach Algorithmus 4 sind $\frac{d(d-1)}{2}$ Richtungsbleitungen zu berechnen. Die Berechnung des Matrix-Exponentials kostet durch die Verwendung der Givensdarstellung (3.33) nur $2d$ Multiplikationen. Für die Funktionsauswertungen muss jedoch das volle Matrix-Exponential mit der Pade-Approximation (Algorithmus 3) berechnet werden.

Damit ergibt sich in jedem Iterationsschritt ein Gesamtaufwand von $\mathcal{O}(d^3)$.

3.3.5 Anwendung auf Eigenwertprobleme

Um die beschriebenen Verfahren zu testen, wollen wir es auf die Berechnung der Eigenwerte einer symmetrischen Matrix $\boldsymbol{H} \in \mathbb{R}^{d\times d}$ anwenden.

Exemplarisch betrachten wir eine Testmatrix, die durch $\boldsymbol{H}_{ij} = \frac{i}{j}$ definiert ist.

$$\boldsymbol{H} := \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{d} \\ \frac{1}{2} & 1 & \frac{2}{3} & \cdots & \frac{2}{d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{d-1} & \cdots & & 1 & \frac{d-1}{d} \\ \frac{1}{d} & \cdots & \frac{d-2}{d} & \frac{d-1}{d} & 1 \end{pmatrix} \tag{3.34}$$

Diese wollen wir durch Minimierung des Funktionales

$$F(\boldsymbol{Q}) := -\mathrm{Tr}(\boldsymbol{Q}^t \boldsymbol{H} \boldsymbol{Q}\, \boldsymbol{Q}^t \boldsymbol{H} \boldsymbol{Q}) = -\sum_{i=1}^{d} (\boldsymbol{Q}^t \boldsymbol{H} \boldsymbol{Q})_{i,i}^2 \tag{3.35}$$

diagonalisieren.

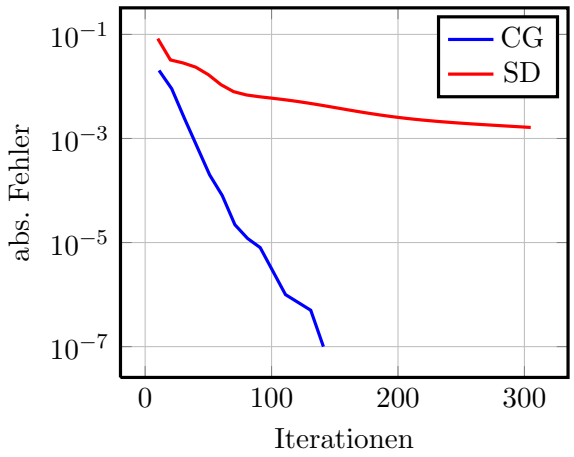
$\boldsymbol{H}$ ist nach dem Spektralsatz über $\mathbb{R}$ diagonalisierbar. Aufgabe soll es zum einen sein eine vollständige Orthonormalbasis zu finden (das sind alle normierten Eigenvektoren), und zum anderen nur die zu den jeweils größten p Eigenwerten gehörigen Eigenvektoren zu bestimmen.

Als erstes wollen wir die Überlegenheit des CG-Verfahrens gegenüber eines Verfahrens des steilsten Abstieges demonstrieren indem wir in $d = 8$ Dimensionen die Konvergenz des Gradienten gegen 0 für das CG-Verfahren aus Algorithmus 4 und ein naives Verfahren des steilsten Abstieges gegenüberstellen. Als Retraktion wurde in beiden Fällen die Exponentialabbildung verwendet. In Abbildung 3.9(a) ist die superlineare Konvergenz des CG-Verfahrens deutlich zu erkennen. Mit dem absoluten Fehler im k-ten Schritt bezeichnen wir dabei den Abstand zum tatsächlichen Minimum.

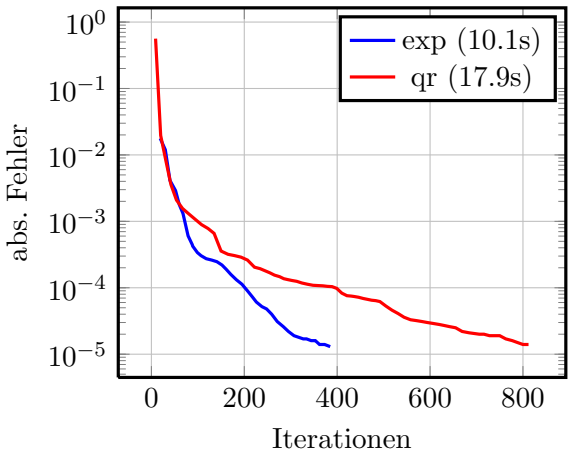
In Abbildung 3.9(b) stellen wir Algorithmus 2 mit $p = d$ dem auf der Exponentialabbildung basierenden Algorithmus 4 gegenüber. Man sieht, dass die Exponentialretraktion mit deutlich weniger Iterationen (und damit Funktionalauswertungen) eine relative Genauigkeit von 10^{-4} erreicht, als die QR-basierte Retraktion, welche ca. doppelt so viele Zyklen benötigt.

In Abbildung 3.9(c) betrachten wir den Fall, dass nur der zum betraglich größten Eigenwert gehörende Eigenvektor gefunden werden soll. In diesem Fall benötigt Algorithmus 2 zwar immer noch mehr Iterationen, jedoch deutlich weniger Rechenzeit als der teurere Algorithmus 4. Mit steigender Zahl p der benötigten Eigenvektoren, reduziert sich dieser Vorteil jedoch wieder, wie in Abbildung 3.9(d) zu sehen ist.

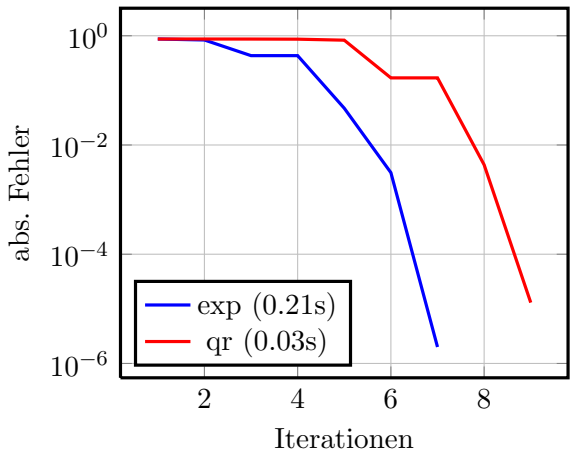
Die angegebenen Rechenzeiten beziehen sich auf ein System mit Intel Xeon CPU X7460 mit 2.66GHz und 16384 KB Cache.



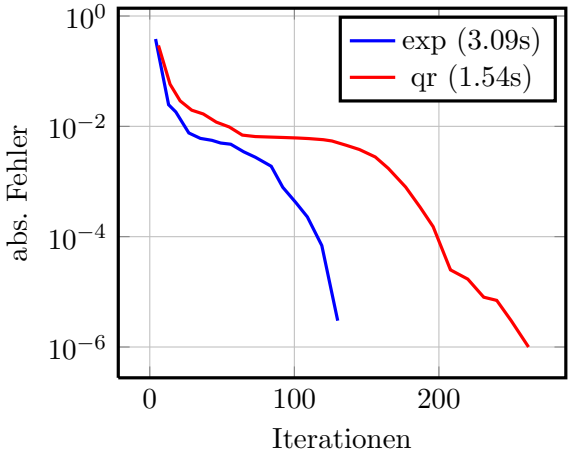
(a) CG vs. SD, $d = 8$



(b) Exp vs. QR, $d = 16$



(c) $d = 16, p = 1$



(d) $d = 16, p = 4$

Abb. 3.9: Vergleich verschiedener Minimierungsverfahren auf $\mathbf{SO}(d)$.

3.4 Auswertung der Integrale

Nachdem wir gesehen haben, wie wir das Funktional $\hat{\mathfrak{M}}$ prinzipiell über $\mathbf{SO}(d)$, bzw. $\mathbf{St}(p,d)$ minimieren können, wollen wir uns nun der Frage zuwenden, wie man dieses möglichst einfach und effizient auswerten kann. Der kostenintensive und damit kritische Teil ist dabei das Aufstellen der ANOVA-Terme $f_{\mathbf{u}}^{\phi}$ und die Berechnung ihrer $\mathcal{L}^2$-Normen $\sigma_{\mathbf{u}}^2 = \|f_{\mathbf{u}}^{\phi}\|_2^2$.

3.4.1 Berechnung der Sensitivitätskoeffizienten

In diesem Abschnitt werden wir beschreiben, wie man die Sensitivitätskoeffizienten einer Funktion $f : \Omega^{(d)} \to \mathbb{R}$ möglichst einfach und effizient berechnen kann.

Um die $\sigma_{\mathbf{u}}^2$ auszurechnen müsste man eigentlich folgendermaßen vorgehen:

$$\begin{aligned}\sigma_{\mathbf{u}}^2(f) &= \int_{\Omega^{\mathbf{u}}} f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})^2 \, d\mu_{\mathbf{u}}(\boldsymbol{x}) \\ &= \int_{\Omega^{\mathbf{u}}} \left(P_{\mathbf{u}}^{\mu}(f)(\boldsymbol{x}_{\mathbf{u}}) - \sum_{\mathbf{v} \subsetneq \mathbf{u}} f_{\mathbf{v}}(\boldsymbol{x}_{\mathbf{v}}) \right)^2 d\mu_{\mathbf{u}}(\boldsymbol{x}) \\ &= \int_{\Omega^{\mathbf{u}}} \left(\sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} \int_{\Omega^{\mathbf{v}^c}} f(\boldsymbol{x}) d\mu_{\mathbf{v}^c}(\boldsymbol{x}) \right)^2 d\mu_{\mathbf{u}}(\boldsymbol{x})\end{aligned} \tag{3.36}$$

Stattdessen berechnen wir jedoch die Werte

$$D_{\mathbf{u}}(f) := \sum_{\mathbf{v} \subseteq \mathbf{u}} \sigma_{\mathbf{v}}^2(f) = \int_{\Omega^{\mathbf{u}}} \left(\int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x}) d\mu_{\mathbf{u}^c}(\boldsymbol{x}) \right)^2 d\mu_{\mathbf{u}}(\boldsymbol{x}) \tag{3.37}$$

und errechnen daraus dann rekursiv die einzelnen $\sigma_{\mathbf{u}}^2(f)$ mittels

$$\sigma_{\mathbf{u}}^2(f) = D_{\mathbf{u}}(f) - \sum_{\mathbf{v} \subsetneq \mathbf{u}} \sigma_{\mathbf{v}}^2(f). \tag{3.38}$$

Dieses Vorgehen hat Vorteile gegenüber (3.36), denn (3.37) lässt sich vermöge des folgenden Lemmas aus [Sob01] durch ein einzelnes Integral berechnen.

Lemma 3.10. Es gilt:

$$D_{\mathbf{u}}(f) = \int_{\Omega^{2d-|\mathbf{u}|}} f(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{x}_{\mathbf{u}^c}) \, f(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{y}_{\mathbf{u}^c}) \, d\mu_{\mathbf{u}}(\boldsymbol{x}) d\mu_{\mathbf{u}^c}(\boldsymbol{x}) d\mu_{\mathbf{u}^c}(\boldsymbol{y}). \tag{3.39}$$

Beweis.

$$D_{\mathbf{u}}(f) := \int_{\Omega^{\mathbf{u}}} \left(\int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x}) d\mu_{\mathbf{u}^c}(\boldsymbol{x}) \right)^2 d\mu_{\mathbf{u}}(\boldsymbol{x})$$

$$= \int_{\Omega^{\mathbf{u}}} \left(\int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{x}_{\mathbf{u}^c}) d\mu_{\mathbf{u}^c}(\boldsymbol{x}) \right) \left(\int_{\Omega^{\mathbf{u}^c}} f(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{y}_{\mathbf{u}^c}) d\mu_{\mathbf{u}^c}(\boldsymbol{y}) \right) d\mu_{\mathbf{u}}(\boldsymbol{x})$$
$$= \int_{\Omega^{\mathbf{u}} \times \Omega^{\mathbf{u}^c} \times \Omega^{\mathbf{u}^c}} f(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{x}_{\mathbf{u}^c})\, f(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{y}_{\mathbf{u}^c})\; d\mu_{\mathbf{u}}(\boldsymbol{x}) d\mu_{\mathbf{u}^c}(\boldsymbol{x}) d\mu_{\mathbf{u}^c}(\boldsymbol{y})$$

□

3.4.2 Quasi-Monte Carlo- und Dünngitter Quadratur

Bei der Auswertung des Funktionals $\hat{\mathfrak{M}}$ sind die Sensitivitätskoeffizienten von $f^\phi := f \circ \phi$ zu berechnen, was sich auf die Berechnung des $(2d - |\mathbf{u}|)$-dimensionalen Integrals (3.39) zurückführen lässt. Um nun $D_{\mathbf{u}}(f^\phi)$ zu berechnen, muss man f^ϕ geeignet diskretisieren, wobei die Verwendung der Dünngitter- und der Quasi-Monte Carlo Quadratur naheliegend ist. Beide Verfahren werden wir ausführlich in 5.2 beschreiben und diskutieren, doch vorerst wollen wir beide Methoden in der allgemeinen Form

$$\int f(x)\, d\mu(\boldsymbol{x}) \approx \sum_{i=1}^{N} w_i\; f(\boldsymbol{x}^{(i)}) \tag{3.40}$$

darstellen, wobei die $w_i \in \mathbb{R}$ einer Folge vorgegebener Quadraturgewichte und die $\boldsymbol{x}^{(i)} \in \mathbb{R}^d$ den zugehörigen Stützstellen entsprechen.

Die Approximation von $D_{\mathbf{u}}(f^\phi)$ durch (3.39) vollziehen wir entsprechend mit einer $(2d - |\mathbf{u}|)$-dimensionalen Quadraturregel der Form (3.40), welche wir für ein gegebenes $f \in V^{(d)}$ in Abhängigkeit von ϕ als

$$Q_{\mathbf{u}}^N(\phi) = \sum_{i=1}^{N} w_i\; f^\phi(\boldsymbol{x}_{\mathbf{u}}^{(i)}, \boldsymbol{x}_{\mathbf{u}^c}^{(i)})\, f^\phi(\boldsymbol{x}_{\mathbf{u}}^{(i)}, \boldsymbol{y}_{\mathbf{u}^c}^{(i)}) \tag{3.41}$$

definieren wollen.

Offensichtlich ist für stetiges f auch $Q_{\mathbf{u}}^N(\phi)$ stetig auf jeder zusammenhängenden Menge $\Phi \subset \mathrm{Diff}(\Omega^{(d)})$, also insbesondere auf $\mathbf{SO}(d)$. (Dies gilt nur für deterministische Quadraturverfahren der Form (3.40). Für Monte Carlo Methoden ist $Q_{\mathbf{u}}^N$ nicht stetig!).

3.4.3 Verfahren I (Quadraturmethode)

Wir definieren nun ein Funktional $\bar{M} : \mathbf{SO}(d) \to \mathbb{R}$, welches im Sinne von 3.2 die Reduktion der effektiven Dimension im Superpositionssinne (Verfahren Ia) und im Trunkationssinne (Verfahren Ib) formalisiert.

Dabei wird

$$\bar{M}(\phi) := \sum_{\mathbf{u}} \gamma_{\mathbf{u}} Q_{\mathbf{u}}^N(\phi)$$

über $\Phi = \mathbf{SO}(d)$ mit Algorithmus 4, bzw. 2 minimiert.

Algorithm 5: Auswertung des diskretisierten Maximierungsfunktionals $\bar{\mathfrak{M}}_f$ für die Trunkationsdimension

Vorraussetzungen: reellwertige Funktion f, geeignete Dimensionsgewichten $\hat{\gamma}_{\mathbf{u}}$, Integrationsverfahren mit N Stützstellen $\boldsymbol{x}^{(i)}$ und passenden Gewichten w_i

Initialisiere:: $S = 0$, $S_0 = 0$

for $k = 1, 2, \ldots, p$ **do**

1) $\mathbf{u} = \{1, \ldots, k\}$

2) Berechne $S_k := D_{\mathbf{u}}(f^{\boldsymbol{Q}})$ durch Formel (3.41)

3) $S = S + \hat{\gamma}_{\mathbf{u}}(S_k - S_{k-1})$

Ergebnis: $(-1) \cdot S$

Verfahren Ia

Verfahren Ia reduziert die Trunkationsdimension einer Funktion $f : \mathbb{R}^d \to \mathbb{R}$. Wir beschreiben die Auswertung des Funktionals $\bar{M}(\boldsymbol{Q})$ in Algorithmus 5. Die Minimierung vollzieht man dann mit Algorithmus 2 über der Stiefelmannigfaltigkeit.

Kostenanalyse: Für jede mit Algorithmus 5 ausgeführte Funktionalauswertung fallen $2N \cdot p$ Auswertungen von f an. Diese setzen sich zusammen aus den $2N$ Auswertungen von f in Formel (3.41) zu Berechnung der $D_{\mathbf{u}}$. Von diesen müssen insgesamt p Stück berechnet werden.

Verfahren Ib

Verfahren Ib reduziert die Superpositionsdimension einer Funktion $f : \mathbb{R}^d \to \mathbb{R}$. Wir beschreiben die Auswertung des Funktionals $\bar{M}(\boldsymbol{Q})$ in Algorithmus 6, wobei wir annehmen, dass ab einem gewissen Superpositionslevel k für die Dimensiomsgewichte $\hat{\gamma}_{\mathbf{u}} = 0$, $|\mathbf{u}| > k$ gilt. Die Minimierung vollzieht man dann mit Algorithmus 4 über der speziellen orthogonalen Gruppe.

Kostenanalyse: Ingesamt müssen $\sum_{j=0}^{k} \binom{d}{j} = \binom{d+k}{d}$ der $D_{\mathbf{u}}(f^{\boldsymbol{Q}})$ berechnet werden. Diese schlagen jeweils mit $2N$ Funktionsauswertungen zu Buche, was einen Gesamtaufwand von $2N\binom{d+k}{d}$ ergibt.

Algorithm 6: Auswertung des diskretisierten Maximierungsfunktionals $\bar{\mathfrak{M}}_f$ für die Superpositionsdimension

Vorraussetzungen: reellwertige Funktion f, geeignete Dimensionsgewichten $\hat{\gamma}_{\mathbf{u}}$, Integrationsverfahren mit N Stützstellen $\boldsymbol{x}^{(i)}$ und passenden Gewichten w_i

Initialisiere:: $S = 0$, $S_\emptyset = 0$

for $|\mathbf{u}| \leq k$ **do**

- **1)** $\mathbf{u} = \{1, \ldots, k\}$
- **2)** Berechne $D_{\mathbf{u}}(f^{\boldsymbol{Q}})$ durch Formel (3.41)
- **3)** Setze $S_{\mathbf{u}} := D_{\mathbf{u}}(f) - \sum_{\mathbf{v} \subsetneq \mathbf{u}} S_{\mathbf{u}}$
- **4)** $S = S + \hat{\gamma}_{\mathbf{u}} S_{\mathbf{u}}$

Ergebnis: $(-1) \cdot S$

3.5 Diskretisierung durch globale Polynome

Die großen Kosten bei der Ausführung von Verfahren I (a+b) bestehen darin, dass für jede Funktionalauswertung f^ϕ erneut diskretisiert werden muss. Dies liegt daran, dass die Punkte eines dünnen Gitters nach Drehung nicht mehr auf dem gleichen dünnen Gitter liegen. Das gleiche gilt für die QMC-Quadratur – auch hier ist die gedrehte Punktfolge nicht mehr in der Ausgangsfolge enthalten. Die Konsequenz ist, dass man für jede Drehung – und damit für jede Auswertung des Funktionals $\bar{\mathfrak{M}}$ neu diskretisieren muss.

Wünscheswert wäre also eine Basis, deren Span unter allen orthogonalen Transformationen abgeschlossen ist. Dies wollen wir im Folgenden konkretisieren.

Rotationsinvariante Funktionenraum Basis

Wir suchen einen Unterraum $\hat{V} \subset V^{(d)}$ mit einer Basis $\mathcal{B}_{\hat{V}}$, so dass

$$f \circ \boldsymbol{Q} \in \hat{V} \text{ für alle } f \in \hat{V}, \boldsymbol{Q} \in \mathbf{SO}(d)$$

gilt und sich für alle Basiselemente $\psi, \tilde{\psi} \in \mathcal{B}_{\hat{V}}$ das Integral

$$(\psi, \tilde{\psi})_\mu = \int \psi \, \tilde{\psi} \, d\mu$$

analytisch berechnen lässt.

Für alle $f \in \hat{V}$ mit $f = \sum_{i=1}^N C_i \psi_i$ soll dann

$$f(\boldsymbol{Q}\boldsymbol{x}) = \sum_{i=1}^N C_i \psi_i(\boldsymbol{Q}\boldsymbol{x}) = \sum_{i=1}^N \tilde{C}_i \psi_i(\boldsymbol{x})$$

gelten, was zu

$$\psi_i \circ \boldsymbol{Q} \in \langle \mathcal{B} \rangle \text{ für alle Basiselemente } \psi_i \in \mathcal{B} \tag{3.42}$$

äquivalent ist.

3.5.1 Die homogene Polynombasis

Tensoransatz

Wir betrachten die eindimensionale Monombasis für Polynome vom Grad $\leq n$

$$\mathcal{B}_n^{(1)} := \{x_1^k, k = 0, \ldots, n\}$$

und definieren

$$\Delta\mathcal{B}_n^{(1)} := \mathcal{B}_n^{(1)} \backslash \mathcal{B}_{n-1}^{(1)} = \{x_1^n\}.$$

Eine d-dimensionale Monombasis erhält man dann mit dem Tensoransatz

$$\bigotimes_{i=1}^{d} \mathcal{B}_{\alpha_i}^{(i)} = \bigoplus_{|\boldsymbol{\alpha}|_\infty \leq n} \bigotimes_{i=1}^{d} \Delta\mathcal{B}_{\alpha_i}^{(i)} = \{\boldsymbol{x}^{\boldsymbol{\alpha}}, |\boldsymbol{\alpha}|_\infty \leq n\}.$$

Diese besitzt $(n+1)^d$ Freiheitsgrade, ist also für unsere Zwecke ungeeignet, da sie wieder dem Fluch der Dimension unterworfen ist.

Statt dessen wählen wir angelehnt an das Dünngitter-Prinzip

$$\mathcal{B}_n^{(d)} := \bigoplus_{|\boldsymbol{\alpha}|_1 \leq n} \bigotimes_{i=1}^{d} \Delta\mathcal{B}_{\alpha_i}^{(i)} = \{\boldsymbol{x}^{\boldsymbol{\alpha}}, |\boldsymbol{\alpha}|_1 \leq n\}$$

die globale Basis homogener Polynome, welche lediglich $K := \binom{d+n}{d}$-Elemente besitzt [Beb08].

Wir werden im Folgenden ausführen, dass $\mathcal{B}_n^{(d)}$ eine für uns geeignete Wahl darstellt. Dazu zeigen wir zunächst die Drehinvarianz und leiten anschließend die analytischen Lösungen der auftretenden Integrationsprobleme her.

Drehinvarianz homogener Polynome Die Polynom/Monom-Basis

$$\mathcal{B}_n := \{\boldsymbol{x}^{\boldsymbol{\alpha}} : |\boldsymbol{\alpha}|_1 \leq n\}$$

erfüllt Bedingung (3.42), was wir im folgenden Lemma beweisen werden.

Lemma 3.11. (Drehinvarianz homogener Polynome)

Die d-dimensionale Basis der homogenen Polynome $\mathcal{B}_n$ vom Grade $\leq n$ spannt einen Funktionenraum auf, der unter allen Drehungen des Koordinatensystems $\mathbb{R}^d$ mit einem Element $\boldsymbol{Q} \in \mathbf{SO}(d)$ invariant ist, also

$$\psi \circ \boldsymbol{Q} \in \langle \mathcal{B}_n \rangle \quad \text{für alle } \psi \in \mathcal{B}_n$$

Beweis. Die Elemente $\psi_{\boldsymbol{\alpha}} = \boldsymbol{x}^{\boldsymbol{\alpha}}$ von $\mathcal{B}$ wollen wir mit einem Multiindex $\boldsymbol{\alpha}$ indizieren, so dass $|\boldsymbol{\alpha}|_1 \leq n$ gilt. Damit rechnen wir unter Verwendung des Multinomialtheorems

$$\begin{aligned}
\psi_\alpha \circ \boldsymbol{Q}(\boldsymbol{x}) &= \prod_{i=1}^{d} (Q_{i\cdot}^t \boldsymbol{x})^{\alpha_i} \\
&= \prod_{i=1}^{d} \left(\sum_{j=1}^{d} Q_{ij} x_j \right)^{\alpha_i} \\
&= \prod_{i=1}^{d} \sum_{|\boldsymbol{\beta}|=\alpha_i} \binom{\alpha_i}{\boldsymbol{\beta}} \prod_{j=1}^{d} (Q_{ij} x_j)^{\beta_j}
\end{aligned}$$

$$= \prod_{i=1}^{d} \underbrace{\sum_{|\boldsymbol{\beta}|=\alpha_i} \binom{\alpha_i}{\boldsymbol{\beta}} \left(\prod_{j=1}^{d} Q_{ij}^{\beta_j} \right) \psi_{\boldsymbol{\beta}}(\boldsymbol{x})}_{\in \langle \mathcal{B}_{\alpha_i} \rangle}$$

nach. Wegen $\deg(P \cdot Q) = \deg(P) + \deg(Q)$ folgt dann die Behauptung. □

Analytische Berechnung der Integrale

Die Monombasis $\mathcal{B}_n$ besitzt den Vorteil, dass die bei der Auswertung von $\mathfrak{M}$ auftretenden Integrale leicht zu berechnen sind.

Lemma 3.12. (Erwartungswert eines Monoms)

Der Erwartungswert eines Basiselementes, also eines Monoms der Form $\boldsymbol{x}^{\boldsymbol{\alpha}}$ mit $|\boldsymbol{\alpha}|_1 = n$, bezüglich des Gauß-Maßes ist

$$\int_{\mathbb{R}^d} \boldsymbol{x}^{\boldsymbol{\alpha}} \; \varphi^d(\boldsymbol{x}) \, d\boldsymbol{x} = \begin{cases} \prod_{i=1}^{d} (\alpha_i - 1)!! & \text{alle } \alpha_i \text{ gerade} \\ 0 & \text{sonst} \end{cases}, \tag{3.43}$$

wobei $n!! = n(n-2) \cdot (n-4) \cdot \ldots \cdot 1$ die Doppelfakultät bezeichne.

Beweis. Nach dem Satz von Fubini gilt

$$\int_{\mathbb{R}^d} \boldsymbol{x}^{\boldsymbol{\alpha}} \; \varphi^d(\boldsymbol{x}) \, d\boldsymbol{x} = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^{\alpha_i} \; \exp\left(-\frac{x_i^2}{2} \right) \, dx_i = \prod_{i=1}^{d} \mathbb{E}(x^{\alpha_i}) \tag{3.44}$$

Es müssen also die Momente der Normalverteilung berechnet werden. Diese sind gerade

$$\mathbb{E}(x^p) = \begin{cases} (p-1)!! & p \text{ gerade} \\ 0 & p \text{ ungerade} \end{cases}$$

□

Der Übersicht halber definieren wir (3.43) als $\mathcal{I}_M(\boldsymbol{\alpha}) := \int_{\mathbb{R}^d} \boldsymbol{x}^{\boldsymbol{\alpha}} \; \varphi^d(\boldsymbol{x}) \, d\boldsymbol{x}$ mit der kanonischen Einschränkung $\mathcal{I}_M(\boldsymbol{\alpha}_{\mathbf{u}})$ auf die in $\mathbf{u}$ enthaltenen Richtungen und leiten damit die analytische Berechnung der Sensitivitätskoeffizienten homogener Polynome her.

Satz 3.13. (Berechnung der Sensitivitätskoeffizienten homogener Polynome)

$$D_{\mathbf{u}}(P_n) = \sum_{|\boldsymbol{\alpha}|_1 \leq n} \sum_{|\boldsymbol{\beta}|_1 \leq n} C_{\boldsymbol{\alpha}} C_{\boldsymbol{\beta}} \; \mathcal{I}_M(\boldsymbol{\alpha}_{\mathbf{u}} + \boldsymbol{\beta}_{\mathbf{u}}) \cdot \mathcal{I}_M(\boldsymbol{\alpha}_{\mathbf{u}^c}) \cdot \mathcal{I}_M(\boldsymbol{\beta}_{\mathbf{u}^c}) \tag{3.45}$$

Beweis. Wir benutzen die Formel aus Lemma 3.10, um die Behauptung zu beweisen.

$$\begin{aligned}
D_{\mathbf{u}}(P_n) &= \int_{\mathbb{R}^{2d-|\mathbf{u}|}} P_n(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{x}_{\mathbf{u}^c})\, P_n(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{y}_{\mathbf{u}^c})\ \varphi^d(\boldsymbol{x})\varphi^{d-|\mathbf{u}|}(\boldsymbol{y}_{\mathbf{u}^c})\ d\boldsymbol{x}\, d\boldsymbol{y}_{\mathbf{u}^c} \\
&= \int_{\mathbb{R}^{2d-|\mathbf{u}|}} \sum_{|\boldsymbol{\alpha}|_1 \le n} \sum_{|\boldsymbol{\beta}|_1 \le n} C_{\boldsymbol{\alpha}} C_{\boldsymbol{\beta}}\ \boldsymbol{x}_{\mathbf{u}}^{\boldsymbol{\alpha}_{\mathbf{u}}}\, \boldsymbol{x}_{\mathbf{u}^c}^{\boldsymbol{\alpha}_{\mathbf{u}^c}}\, \boldsymbol{x}_{\mathbf{u}}^{\boldsymbol{\beta}_{\mathbf{u}}}\, \boldsymbol{y}_{\mathbf{u}^c}^{\boldsymbol{\beta}_{\mathbf{u}^c}}\ \varphi^d(\boldsymbol{x})\varphi^{d-|\mathbf{u}|}(\boldsymbol{y}_{\mathbf{u}^c})\ d\boldsymbol{x}\, d\boldsymbol{y}_{\mathbf{u}^c} \\
&= \sum_{|\boldsymbol{\alpha}|_1 \le n} \sum_{|\boldsymbol{\beta}|_1 \le n} C_{\boldsymbol{\alpha}} C_{\boldsymbol{\beta}} \int_{\mathbb{R}^{2d-|\mathbf{u}|}} \boldsymbol{x}_{\mathbf{u}}^{\boldsymbol{\alpha}_{\mathbf{u}}+\boldsymbol{\beta}_{\mathbf{u}}}\, \boldsymbol{x}_{\mathbf{u}^c}^{\boldsymbol{\alpha}_{\mathbf{u}^c}}\, \boldsymbol{y}_{\mathbf{u}^c}^{\boldsymbol{\beta}_{\mathbf{u}^c}}\ \varphi^d(\boldsymbol{x})\varphi^{d-|\mathbf{u}|}(\boldsymbol{y}_{\mathbf{u}^c})\ d\boldsymbol{x}\, d\boldsymbol{y}_{\mathbf{u}^c} \\
&= \sum_{|\boldsymbol{\alpha}|_1 \le n} \sum_{|\boldsymbol{\beta}|_1 \le n} C_{\boldsymbol{\alpha}} C_{\boldsymbol{\beta}}\, \mathcal{I}_M(\boldsymbol{\alpha}_{\mathbf{u}} + \boldsymbol{\beta}_{\mathbf{u}}) \cdot \mathcal{I}_M(\boldsymbol{\alpha}_{\mathbf{u}^c}) \cdot \mathcal{I}_M(\boldsymbol{\beta}_{\mathbf{u}^c})
\end{aligned}$$

□

Somit lassen sich nun alle in Verfahren I (a+b) auftretenden Integrale analytisch berechnen, was das Verfahren deutlich beschleunigen wird. Im nächsten Abschnitt werden wir jedoch feststellen, dass sich unter Umständen sogar eine analytische Lösung für das gesamte Maximierungsproblem angeben lässt.

3.5.2 Lösung im linearen und quadratischen Fall

Betrachtet man nur Polynome vom Grad eins, so ergeben sich weitere Vereinfachungen. In diesem Fall lässt sich das gesamte Minimierungsproblem analytisch berechnen. Für den Fall $n = 2$ kann man die Minimierung auf eine Hauptachsentransformation, letztlich also ein Eigenwertproblem zurückführen.

Lemma 3.14. (Lineare Polynome)

Für eine lineare Abbildung $l : \mathbb{R}^d \to \mathbb{R}$, $l(\boldsymbol{x}) = \boldsymbol{w}^t\boldsymbol{x}$ kann man stets eine orthogonale Matrix $\boldsymbol{Q} \in \mathbf{SO}(d)$ angeben, so dass $l \circ \boldsymbol{Q}(\boldsymbol{x}) = \|\boldsymbol{w}\|_2 x_1$ gilt – d.h. jede lineare Funktion lässt sich achsenparallel ausrichten.

Beweis. Wir konstruieren die gesuchte Drehmatrix $\boldsymbol{Q}$, indem wir ihre erste Spalte auf

$$\boldsymbol{Q}_{\cdot 1} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2} \tag{3.46}$$

setzen und die verbleibenden $d-1$ Spalten von $\boldsymbol{Q}$ paarweise orthonormal wählen. Dies kann man günstig mit einem Gram-Schmidt Verfahren erreichen, numerisch stabiler ist jedoch die Orthogonalisierung mittels QR-Zerlegung oder Givensrotationen.

Aus $\langle \boldsymbol{Q}_{\cdot i}, \boldsymbol{Q}_{\cdot j} \rangle = 0$ für $i \neq j$ folgt dann

$$l \circ \boldsymbol{Q}(\boldsymbol{x}) = \boldsymbol{w}^t\boldsymbol{Q}\boldsymbol{x} = \sum_{j=1}^{d} \langle \boldsymbol{w}, \boldsymbol{Q}_{\cdot j} \rangle x_j = \langle \boldsymbol{w}, \boldsymbol{Q}_{\cdot 1} \rangle x_1 = \|\boldsymbol{w}\|_2 x_1,$$

womit die Behauptung bewiesen ist. □

Mit symmetrischer Matrix $\boldsymbol{H} \in \mathbb{R}^{d\times d}$ defnieren wir quadratische Polynome als

$$P_2(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{H}\boldsymbol{x} = \sum_{i=1}^{d}\sum_{j=i}^{d} H_{ij} x_i\, x_j.$$

Lemma 3.15. (Quadratische Polynome)
Für jedes quadratische Polynom P_2 existiert ein $\boldsymbol{Q} \in \mathbf{O}(d)$ so dass $P_2 \circ \boldsymbol{Q}$ Superpositionsdimension eins besitzt.

Beweis. Aus der Symmetrie folgt, dass $\boldsymbol{H}$ als Bilinearform selbstadjungiert ist, womit sich aus dem Spektralsatz ergibt, dass es eine orthogonale Matrix $\boldsymbol{Q} \in \mathbf{O}(d)$ und eine Diagonalmatrix $\mathcal{D}$ gibt, so dass

$$\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^t$$

gilt. Damit folgt

$$P_2(\boldsymbol{Q}\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{D}\boldsymbol{x} = \sum_{i=1}^{d} D_{ii} x_i^2$$

□

3.6 Basiswechsel durch Differentiation

Nachdem wir nun einen geeigneten Untervektorraum $\mathcal{P}_n$ von $\mathcal{L}^2$ mit Basis $\mathcal{B}$ gefunden haben, stellt sich die Frage, wie wir eine beliebige Funktion $f \in \mathcal{L}^2(\mathbb{R}^d, \eta)$ in dieser Basis darstellen, ohne dass die für die Minimierung relevante Information über die dimensionale Struktur von f verloren geht.

Einen möglichen Ansatz stellt die abgeschnittene Taylorreihe an einem Punkt $\boldsymbol{a} \in \mathbb{R}^d$

$$T_N(\mathbf{x}) = \sum_{|\alpha|\leq N} \frac{(\mathbf{x}-\mathbf{a})^\alpha}{\alpha!} \mathbf{D}^\alpha f(\mathbf{a}) \tag{3.47}$$

dar, sofern f im Entwicklungspunkt $\boldsymbol{a}$ hinreichend differenzierbar ist.

Obwohl wir aus der eindimensionalen Polynominterpolation wissen, dass die Taylorreihe deutlich schlechtere Konvergenzeigenschaften als etwa die Lagrange-Interpolation besitzt, wollen wir diesen Ansatz erläutern da er für lineare Polynome auf ein Verfahren führt, welches in den letzten Jahren sehr viel Anerkennung gefunden hat – die *Lineare Transformation* (LT) [IT04].

3.6.1 Die Lineare Transformation nach Imai/Tan

Die *Lineare Transformation* (LT) [IT06] verwendet das Taylorpolynom ersten Grades an einem Punkt $\boldsymbol{a} \in \Omega^{(d)}$, um die Polynomkoeffizienten in $\mathcal{B}_1$ zu bestimmen.

Um mit der LT-Methode nun eine passende Drehung für eine Funktion $f : \mathbb{R}^d \to \mathbb{R}$ zu finden, wird

$$\boldsymbol{Q}_{\cdot 1} = \frac{\nabla f(\boldsymbol{a})}{\|\nabla f(\boldsymbol{a})\|_2} \tag{3.48}$$

gesetzt und die anderen Spalten können beliebig, aber orthonormal gewählt werden. Für lineare $f = \boldsymbol{w}^t\boldsymbol{x} + w_0$ ist die LT wegen $\nabla f = \boldsymbol{w}$ offensichtlich optimal. Wir können jedoch beweisen, dass sie auch für eine weitaus größere Klasse von Funktionen die bestmögliche Drehung liefert.

Definition 3.16. (Ridge-Funktionen)

Als *Ridge-Funktion* bezeichnen wir eine Funktion $f : \mathbb{R}^d \to \mathbb{R}$, welche eine Darstellung

$$f(\boldsymbol{x}) = g\left(w_0 + \sum_{i=1}^{d} w_i x_i\right) = g(w_0 + \boldsymbol{w}^t\boldsymbol{x})$$

besitzt, wobei $g : \mathbb{R} \to \mathbb{R}$ beliebig, aber am Punkte $w_0 + \boldsymbol{w}^t\boldsymbol{a}$ differenzierbar sei.

Zu dieser Klasse gehören praxisrelevante Funktionen wie etwa $\max(0, h(\boldsymbol{x}))$ oder $\exp(\boldsymbol{w}^t\boldsymbol{x})$.

Satz 3.17. (Achsenparallele Ausrichtung von Ridge-Funktionen)

Für jede d-dimensionale Funktion der Form $f(\boldsymbol{x}) = g(\boldsymbol{w}^t\boldsymbol{x})$ gilt mit $\boldsymbol{Q}$ wie in (3.46)

$$f \circ \boldsymbol{Q}(\boldsymbol{x}) = g(\|\boldsymbol{w}\|_2 x_1)$$

Beweis. Kennt man den Vektor $\boldsymbol{w}$ so lässt sich Lemma 3.14 anwenden und die Behauptung folgt sofort.

Da man den Vektor $\boldsymbol{w}$ jedoch a-priori nicht kennt, errechnet man ihn über

$$\frac{\nabla f(\boldsymbol{a})}{\|\nabla f(\boldsymbol{a})\|_2} = \frac{g'(\boldsymbol{w}^t\boldsymbol{a})\boldsymbol{w}}{|g'(\boldsymbol{w}^t\boldsymbol{a})|\|\boldsymbol{w}\|_2} = \pm\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2},$$

falls $g'(\boldsymbol{w}^t\boldsymbol{a}) \neq 0$ gilt. Dies entspricht gerade der durch die LT definierten Spalte von $\boldsymbol{Q}$. □

Verbleibende Freiheitsgrade

Obwohl die Spalten $2, \ldots, d$ die Orthonormalitätsbedingung (3.7) erfüllen müssen, stehen noch Freiheitsgrade zur Verfügung. Die LT versucht diese zu nutzen, indem f sukzessive an $d-1$ weiteren Punkten linearisiert wird und die dabei auftrenden Gradienten nach Orthogonalisierung die weiteren Spalten von $\boldsymbol{Q}$ belegen.

In der Praxis ist der Nutzen durch weitere Linearisierungen allerdings eher gering. Imai und Tan empfehlen höchstens die ersten 20 Spalten zu verwenden, da andernfalls der numerische Aufwand zur Orthogonalisierung der Gradienten in jedem zusätzlichen Schritt zu groß wird.

Bemerkung: Die LT reduziert die effektive Dimension im Trunkationssinne.

3.6.2 Eine Hauptachsentransformation

Die naheliegende Erweiterung der LT ist natürlich die Verwendung quadratischer Polynome – denn auch für diese existiert eine analytische Lösung nach Lemma 3.15, was gerade der Diagonal-Methode (DM) aus [Mor98, Wan10] entspricht. Dazu bestimmen wir die Hesse-Matrix von f, berechnen für diese die Hauptachsentransformation und erhalten damit nach Lemma 3.15 eine Summe eindimensionaler Funktionen.

Quadratische Funktionen

Gegeben sei eine quadratische Funktion

$$f(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{H} \boldsymbol{x} = \sum_{i=1}^{d} \sum_{j=i}^{d} H_{ij} x_i \, x_j,$$

wobei $\boldsymbol{H} \in \mathbb{R}^{d \times d}$ symmetrisch ist.

Wir berechnen die Hesse-Matrix im Punkt $\boldsymbol{a}$

$$\mathrm{Hess} f(\boldsymbol{a}) = \left(\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{a}) \right)_{i,j}$$

und verwenden Lemma 3.15, welches ein orthogonale Matrix $\boldsymbol{Q}$ liefert, so dass für quadratische Funktionen

$$f(\boldsymbol{Q}\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{D} \boldsymbol{x} = \sum_{i=1}^{d} D_{ii} x_i^2$$

folgt. Nicht-quadratische Funktionen werden jedoch nur in einer Umgebung von $\boldsymbol{a}$ „dekorreliert“.

Bemerkung: Diese Methode reduziert die Superpositionsdimension quadratischer Polynome exakt auf eins.

Auch für die DM können wir zeigen, dass sie für Ridge-Funktionen die optimale Drehung liefert, also etwas allgemeiner eingesetzt werden kann als die LT.

Satz 3.18. (Optimalität der DM für Ridge-Funktionen)

Für jede Ridge-Funktion (siehe Definition 3.16) liefert die DM im Punkt $\boldsymbol{a}$ eine achsenparallele Ausrichtung, also die optimale Transformation auf eine eindimensionale Funktion, falls $\mathrm{Hess} f(\boldsymbol{a})$ von Null verschiedenen ist.

Beweis. Wir betrachten $f(\boldsymbol{x}) = g(\boldsymbol{w}^t \boldsymbol{x} + w_0)$. Dann gilt $\frac{\partial^2}{\partial x_i x_j} f(\boldsymbol{x}) = w_i w_j g''(\boldsymbol{w}^t \boldsymbol{x} + w_0)$ und somit für die Hessematrix von f im Punkt $\boldsymbol{a}$

$$\mathrm{Hess} f(\boldsymbol{a}) = \underbrace{g''(\boldsymbol{w}^t \boldsymbol{a} + w_0)}_{=:C} \, \boldsymbol{w}\boldsymbol{w}^t.$$

Es handelt sich also um eine Rang 1 Matrix. Für diese gilt, dass sie höchstens einen von Null verschiedenen Eigenwert λ besitzen kann. Sei $\boldsymbol{v}$ ein zu λ gehörender Eigenvektor, so gilt

$$(\mathrm{Hess} f(\boldsymbol{a}) - \lambda \mathbf{Id})\boldsymbol{v} = \mathbf{0} \Leftrightarrow (C\boldsymbol{w}^t\boldsymbol{w} - \lambda)\boldsymbol{w}^t\boldsymbol{v} = \mathbf{0},$$

womit folgt, dass $\lambda = C\boldsymbol{w}^t\boldsymbol{w}$ gelten muss.

Der zu λ gehörige Eigenvektor ist nun die bis auf Skalierung eindeutige Lösung des linearen Gleichungssystems

$$(\boldsymbol{w}\boldsymbol{w}^t - \boldsymbol{w}^t\boldsymbol{w}\mathbf{Id})\,\boldsymbol{x} = \mathbf{0}. \tag{3.49}$$

Offensichtlich ist $\boldsymbol{w}$ eine Lösung von (3.49) und spannt somit den zu λ gehörenden Eigenraum auf. Da die DM eine Orthonormalbasis von $\mathrm{Hess} f(\boldsymbol{a})$ findet, ist bei einer absteigenden Sortierung der Eigenwerte die erste Spalte von $\boldsymbol{Q}$ also gerade $\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}$, was der Lösung entspricht, welche die LT liefert (siehe Satz 3.17). □

3.7 Basiswechsel durch Projektion

Nachteilig an der Projektion durch Abschneiden der Taylorreihe sind die schlechten Approximationseigenschaften – insbesondere wenn f nicht hinreichend oft differenzierbar ist.

Daher berechnen wir die orthogonale Projektion von $\mathcal{L}^2$ auf $\mathcal{P}_n$, welche nach Lemma 2.1 aus Kapitel 2 bezüglich der $\mathcal{L}^2$-Norm optimal ist.

Sei also $P_n(\boldsymbol{x})$ ein d-dimensionales homogenes Polynom vom Grad $\leq n$, d.h.

$$P_n(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}|_1 \leq n} c_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}.$$

Um P_n aufzustellen (also f auf $\mathcal{P}_n$ zu projizieren), suchen wir nun das Polynom, welches den geringsten $\mathcal{L}_2$-Abstand zu f besitzt, also

$$\operatorname*{arg\,min}_{P_n \in \mathcal{P}_n} \|f - P_n\|_2 = \operatorname*{arg\,min}_{P_n \in \mathcal{P}_n} \int_{\mathbb{R}^d} (f(\boldsymbol{x}) - P_n(\boldsymbol{x}))^2 \, d\eta(\boldsymbol{x}). \tag{3.50}$$

Diskretisiert man dieses Integral mit einem Quadraturverfahren der Form

$$\int f \, d\mu \approx \sum_{i=1}^{N} w^{(i)} f(\boldsymbol{x}^{(i)}) \tag{3.51}$$

welche für $i = 1, \ldots, N$ durch ihre Integrationsgewichte $w^{(i)}$ mit zugehörigen Stützstellen $\boldsymbol{x}^{(i)}$ definiert sind, so kommt man auf das diskrete Problem

$$\operatorname*{arg\,min}_{P_n \in \mathcal{P}_n} \sum_{i=1}^{N} w^{(i)} \left(f_k(\boldsymbol{x}^{(i)}) - P_n(\boldsymbol{x}^{(i)}) \right)^2. \tag{3.52}$$

Geeignete Verfahren stellen zum Beispiel die dünnen Gitter aus Abschnitt 5.2.2 oder Quasi-Monte Carlo Methoden aus Abschnitt 5.2.1 dar.

3.7.1 Darstellung als lineares Ausgleichsproblem

Die Minimierung (3.52) über alle $P_n = \sum_{|\boldsymbol{\alpha}|\leq n} c_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$ lässt sich als ein *gewichtetes Least-Squares*-Problem formulieren, welches man effizient mit der QR- oder SVD-Zerlegung lösen kann.

Dazu definieren wir $\boldsymbol{D} := \operatorname{diag}(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}, \ldots, \boldsymbol{w}^{(N)})$, $\bar{\mathbf{f}} := (f(\boldsymbol{x}^{(1)}), \ldots, f(\boldsymbol{x}^{(N)}))^t$ und $\bar{\boldsymbol{y}} := (P_n(\boldsymbol{x}^{(1)}), \ldots, P_n(\boldsymbol{x}^{(N)})^t$. Damit entspricht (3.52) dann

$$\arg\min_{P_n \in \mathcal{P}_n} \|D(\bar{\mathbf{f}} - \bar{\boldsymbol{y}})\|_2.$$

Definieren wir nun

$$\boldsymbol{A} = \begin{pmatrix} 1 & \boldsymbol{x}_{(1)}^{(1,0,..)} & \ldots & \boldsymbol{x}_{(1)}^{\boldsymbol{\alpha}} & \ldots \\ 1 & \boldsymbol{x}_{(2)}^{(1,0,..)} & \ldots & \boldsymbol{x}_{(2)}^{\boldsymbol{\alpha}} & \ldots \\ \vdots & \vdots & & \vdots & \\ 1 & \boldsymbol{x}_{(K)}^{(1,0,..)} & \ldots & \boldsymbol{x}_{(K)}^{\boldsymbol{\alpha}} & \ldots \end{pmatrix} \quad \text{und} \quad \bar{\boldsymbol{c}} = \begin{pmatrix} c_0 \\ \vdots \\ c_{\boldsymbol{\alpha}} \\ \vdots \end{pmatrix}, \tag{3.53}$$

wobei $K = \binom{d+n}{d}$ der Zahl der Basiselemente entspricht, so folgt wegen $y_i = P_n(\boldsymbol{x}^{(i)}) = \boldsymbol{A}_{i.}\bar{\boldsymbol{c}}$

$$\arg\min_{P_n \in \mathcal{P}_n} \|f - P_n\|_2 \approx \arg\min_{\bar{\boldsymbol{c}} \in \mathbb{R}^N} \|\boldsymbol{D}(\bar{\mathbf{f}} - \boldsymbol{A}\bar{\boldsymbol{c}})\|_2. \tag{3.54}$$

Das lineare Ausgleichsproblem (3.54) ist also die diskrete Version des Variantionsproblems (3.52), welches sich nun mit einer QR- oder SVD-Zerlegung von $\boldsymbol{A}$ stabil lösen lässt [FH07].

3.7.2 Verfahren II (Polynom-Methode)

Wir projizieren f auf den Raum der homogenen Polynome vom Grad n und berechnen dort das Minimum von $-\hat{\mathfrak{M}}_{P_n}$ mit Verfahren I (a oder b), wobei wir die dabei auftretenden Integrale nun analytisch bestimmen können.

Kostenanalyse

Da ein homogenes Polynom vom Grad kleiner-gleich n in d Dimensionen $K = \binom{d+n}{d}$ Freiheitsgrade besitzt und die QR-Zerlegung einer $K \times K$-Matrix mit anschließender Vorwärts-Substitution einen Aufwand von $\mathcal{O}(K^3)$ Multiplikationen kostet, ist die Lösung des gesamten Ausgleichsproblem mit $\mathcal{O}(\binom{d+n}{d}^3)$ Multiplikationen verbunden. Diese fallen jedoch nur einmalig bei der Projektion auf $\mathcal{P}_n$ an.

Das Aufstellen der Matrix, welches implizit eine Dünngitter/QMC-Quadratur beinhaltet kostet nur $\mathcal{O}(\binom{d+n}{d}^2)$ und wird daher von den Kosten für die QR-Zerlegung dominiert.

Algorithm 7: $\mathcal{L}^2$-orthogonale Projektion auf $\mathcal{P}_n$

Vorraussetzungen: reellwertige Funktion f, Integrationsverfahren mit N Stützstellen $\boldsymbol{x}^{(i)}$ und passenden Gewichten w_i, $N > \binom{d+n}{d}$

1) Stelle die Matrix $\boldsymbol{A}$, den Vektor $\bar{\mathbf{f}}$ und die Diagonalmatrix $\boldsymbol{D}$ aus aus (3.53) auf

2) Löse das lineare Ausgleichsproblem (3.54) mit der QR-Zerlegung von $\boldsymbol{A}$ und Lösungsvektor $\boldsymbol{c}$.

3) Besetze die Koeffizienten von P_n mit $\boldsymbol{c}$.

Ausgabe: Homogenes Polynom vom Grad n mit minimalem $\mathcal{L}^2$-Abstand zu f

Für die analytische Berechnung der $D_{\mathbf{u}}$ nach Satz 3.13 und Lemma 3.11 fallen bei *jeder* Auswertung von $\bar{\mathfrak{M}}$ im CG-Verfahren Kosten der Ordnung $\mathcal{O}(n\binom{d+n}{d}^2)$ an.

3.8 Nichtlineare Transformationen

Wie bereits zu Beginn dieses Kapitels bemerkt, beinhaltet die Diskretisierung allgemeiner d-dimensionaler Diffeomorphismen den Fluch der Dimension. Um diesen bei der Darstellung reellwertiger Funktionen $f : \Omega^{(d)} \to \mathbb{R}$ zu umgehen, kommen also nur solche Diffeomorphismen zur Transformation des zugrundeliegenden Gebietes $\Omega^{(d)}$ in Frage, deren Freiheitsgrade höchstens polynomiell von d abhängen.

Zwei nichtlineare Beispiele für solche Bijektionen haben wir bereits genannt - die komponentenweise Gradierung des Gitters (Beispiel 3.2) und die stückweise orthogonalen Transformationen aus Abschnitt 3.1.3.

Während sich Gittergradierungen in moderaten Dimensionen[1] als ortsadaptive Gitterverfeinerung interpretieren lassen und damit schon relativ gut erforscht sind (etwa [Feu10]), bieten die krummlinigen Koordinatentransformationen aus Abschnitt 3.1.3 interessantes Potential, welches wir im Folgenden im zweidimensionalen Raum untersuchen wollen. Für $d \geq 3$ lässt sich das Konzept zwar ebenfalls verwenden, allerdings wird dann die Definition eines Matrix-Logarithmus auf $\mathbf{SO}(d)$ nötig, was an dieser Stelle zu weit führen würde.

3.8.1 Diskretisierung von stückweise-orthogonalen Abbildungen

An Abschnitt 3.1.3 anknüpfend suchen wir eine stetig-differenzierbare Abbildung $\boldsymbol{q} : \mathbb{R}^+ \to \mathbf{SO}(d)$ durch die wir dann eine Abbildung $\hat{\boldsymbol{Q}} : \mathbb{R}^d \to \mathbb{R}^d$ mit

$$\hat{\boldsymbol{Q}}(\boldsymbol{x}) := \underbrace{\boldsymbol{q}(\|\boldsymbol{x}\|_2^2)}_{\in \mathbb{R}^{d\times d}} \boldsymbol{x}$$

[1] In hohen Dimensionen ist die Erzeugung ortsadaptiver Gitter aufgrund der dazu notwendigen Datenstrukturen nicht mehr praktikabel. Für Dimensionen ≥ 20 könnte eine a-priori definierte Gittergradierung also wieder sinnvoll sein.

Algorithm 8: Auswertung der Funktion $f \circ \hat{\boldsymbol{Q}}(\boldsymbol{c})$

Vorraussetzungen: reellwertige Funktion f
Eingabe:: Polynomkoeffizienten $\boldsymbol{c} = (c_0, \ldots, c_n) \in \mathbb{R}^{n+1}$, Vektor $\boldsymbol{x} \in \mathbb{R}^2$

1) Berechne die Norm $r = \|\boldsymbol{x}\|_2$
2) Berechne $g := P_n(r) = c_0 + c_1 r + \ldots + c_n r^n$
3) Setze $\boldsymbol{Q} := \tilde{\exp}(g) = \begin{pmatrix} \cos(g) & \sin(g) \\ -\sin(g) & \cos(g) \end{pmatrix}$
4) Setze $\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}$

Ergebnis: $f(\boldsymbol{y})$

definieren können.

Im Folgenden werden wir Verfahren zur Bestimmung einer derartigen Abbildung für den Fall $d = 2$ vorschlagen, welches zur Reduktion der effektiven Dimension eingesetzt werden kann. Dabei benutzen wir, dass die Abbildung

$$\tilde{\exp} : s \to \exp \begin{pmatrix} 0 & s \\ -s & 0 \end{pmatrix}$$

das Intervall $[0, \infty)$ surjektiv auf $\mathbf{SO}(2)$ abbildet. Mit einem eindimensionalen Polynom vom Grad n

$$P_n(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n,$$

definieren wir nun die stückweise-orthogonale Abbildung

$$\hat{\boldsymbol{Q}}(\boldsymbol{x}) := \tilde{\exp}(P_n(\|\boldsymbol{x}\|_2))\boldsymbol{x}, \tag{3.55}$$

welche offensichtlich stetig-differenzierbar ist und durch die Koeffizienten des Polynoms $\boldsymbol{c} = (c_0, \ldots, c_n) \in \mathbb{R}^{n+1}$ parametrisiert wird.

Damit können wir nun das Maximierungsfunktional $\bar{\mathfrak{M}}_f(\hat{\boldsymbol{Q}}(\boldsymbol{c}))$ auf dem Vektorraum $\mathbb{R}^{n+1}$ definieren, welches wir nach Algorithmus 6 auswerten. Die dabei nötigen Auswertungen von $f \circ \hat{\boldsymbol{Q}}(\boldsymbol{c}) : \mathbb{R}^d \to \mathbb{R}$ ist in Algorithmus 8 beschrieben, welchen wir im Weiteren als **Verfahren III** bezeichen werden.

4 Numerische Ergebnisse

In diesem Kapitel werden wir untersuchen, inwiefern sich die effektive Dimension durch die vorgestellten Verfahren tatsächlich reduzieren lässt. Dazu betrachten wir verschiedene Modellfunktionen und vergleichen deren effektive Dimensionen in Standardkoordinaten mit ihren dimensionsoptimierten Transformationen.

Durchführung der Versuche

Um die Verfahren für eine gegebene Funktion f miteinander zu vergleichen starten wir an einem zufälligen Punkt auf $\mathbf{SO}(d)$ und maximieren von dort ausgehend das Funktional $\bar{\mathfrak{M}}_f$ mit der direkten Quadraturmethode (SG_{10^5}, Qmc_{10^5}), der Polynom-Methode zum Grad $n = 1, 2, 3$ (Poly1, Poly2, Poly3), sowie für differenzierbare Funktionen mit der Linearen Transformation (LT) und der Diagonal-Methode von Morokoff (DM). Die auf diese Weise dimensionsoptimierte Funktion vergleichen wir mit der ursprünglichen Funktion in Standardkoordinaten (Std) und im Falle der Optionspreisberechnung auch mit der Brownschen Brücke (BB) und der PCA-Pfadkonstruktion.

Zur Definition des diskretisierten Minimierungsfunktionals $\bar{\mathfrak{M}}$ verwenden wir in diesem Kapitel, sofern nicht anders angegeben, als Dimensionsgewichte $\hat{\gamma}_{\mathbf{u}} = 1$ für $|\mathbf{u}| = 1$, $\hat{\gamma}_{\mathbf{u}} = e^{-1}$ für $|\mathbf{u}| = 2$ und $\hat{\gamma}_{\mathbf{u}} = 0$ für $|\mathbf{u}| \geq 3$.

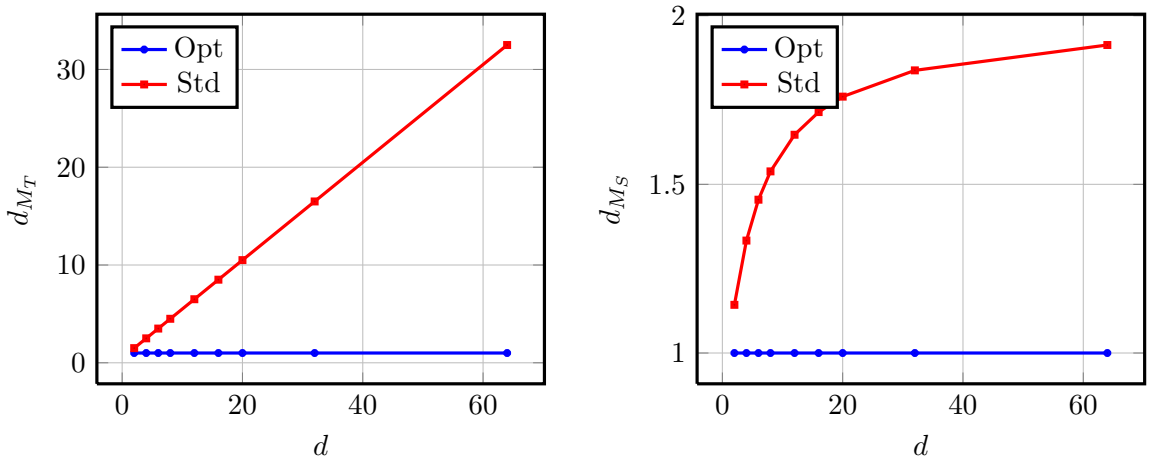
Das zugrundeliegende Maß für die untersuchten Begriffe von Superpositions- und Trunkationsdimension wird, sofern nicht anders angegeben, stets das Gauß-Maß auf dem Ganzraum $\mathbb{R}^d$ sein.

Wir bemerken, dass das Funktional $\bar{\mathfrak{M}}_f$ im Allgemeinen nicht konvex ist und es daher möglich ist, dass unsere Liniensuchverfahren in ein lokales Minimum konvergieren, welches nicht zwingend das globale Minimum darstellen muss.

4.1 Polynome

Die Funktionenklasse der homogenen Polynome vom Grad $\leq n$ werden wir besonders genau untersuchen, da wir zum einen mit der Polynommethode die Reduktion allgemeiner Funktionen auf den Fall homogener Polynome zurückführen, und sie es zum anderen ermöglichen die dimensionale Struktur sehr einfach zu modellieren.

Dazu betrachten wir ausgewählte Polynome in verschiedenen Dimensionen d über der Menge $\mathbb{R}^d$ und berechnen die in Verfahren I auftretenden Integrale analytisch durch die Formeln aus Abschnitt 3.5.1.



(a) linear (b) quadratisch

Abb. 4.1: Mittlere Dimensionen für ein lineares und ein quadratisches Polynom des Typs (4.1) in $d = 2, 4, 6, 8, 12, 16, 20, 32, 64$ Dimensionen.

Polynom 1

Wir betrachten das homogene Polynom vom Grad $\leq n$, bei dem für jeden Koeffizienten $C_{\boldsymbol{\alpha}} = 1$, also

$$P_n(\boldsymbol{x}) := \sum_{|\boldsymbol{\alpha}|_1 \leq n} \boldsymbol{x}^{\boldsymbol{\alpha}} \tag{4.1}$$

gilt.

Eine solches Polynom repräsentiert Funktionen, deren Ableitungen sowohl isotrop, als auch in gemischten Richtungen gleichermaßen zur Funktion beitragen.

In Abschnitt 3.5.2 haben wir für den Fall $n = 1$ gezeigt, dass sich P_1 stets auf eine Funktion mit Trunkationsdimension 1 (also eine eindimensionale Funktion) reduzieren lässt. Ähnliches gilt für quadratische Polynome, für welche nach Lemma 3.15 stets eine Transformation auf eine Funktion mit Superpositionsdimension 1 existiert. In Abbildung 4.1 ist für verschiedene Dimensionen die mittlere Dimension im Trunkationssinne d_{M_T} für ein lineares, und die mittlere Dimension im Superpositionssinne d_{M_S} für ein quadratisches Polynom des oben beschriebenen Typs dargestellt. In beiden Fällen ist zu sehen, dass die mittlere Dimension des nichttransformierten Polynoms in der nominellen Dimension d monoton steigt, während die mittleren Dimensionen $d_{M_T} = 1$ und $d_{M_S} = 1$ der mit den Verfahren Ia, bzw. Ib berechneten Drehungen des Polynoms von der Dimension d unabhängig konstant 1 ist. Unser Verfahren konvergiert in diesem Falle also gegen die analytisch bestimmte optimale Lösung.

In Abbildung 4.2 sind für die Fälle $n = 3$ und $n = 4$ (für die keine analytischen Lösungen mehr existieren) die relative Superpositionsdimension $\bar{d}_S(k)$ für $k \in \{1, 2\}$ dargestellt. Deutlich ist zu sehen, dass mit steigender Dimension d immer weniger Varianz in den ersten beiden Superpositionsdimensionen des nicht-transfomierten Polynoms enthalten ist, während das gedrehte

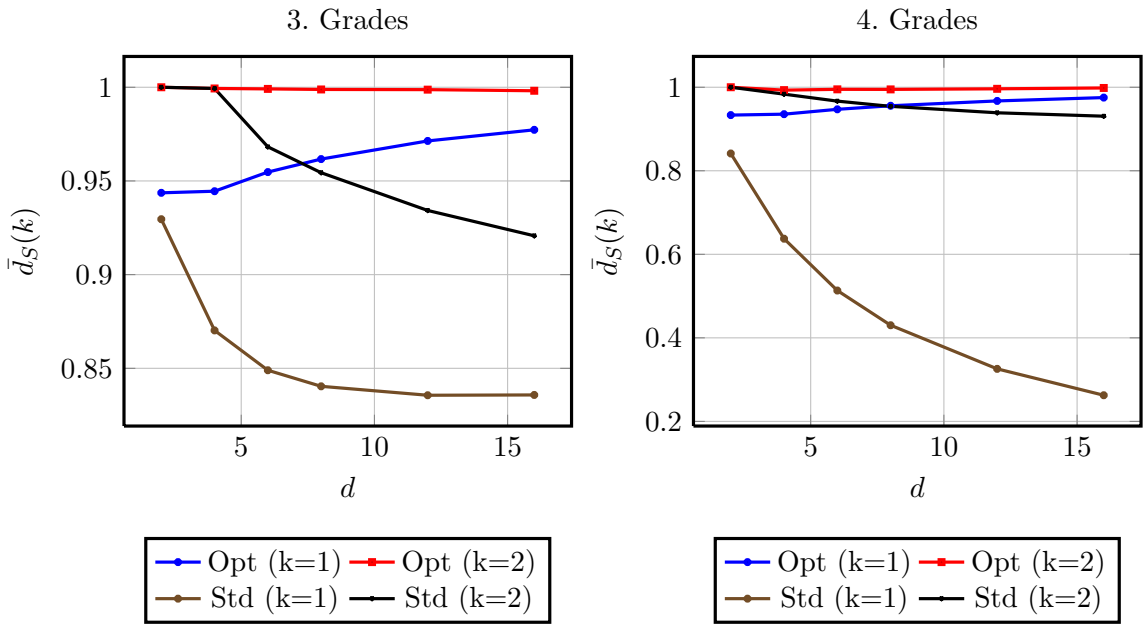


Abb. 4.2: Relative Superpositionsdimension $\bar{d}_S(k)$ von Polynom 1 für $n \in \{3, 4\}$.

Polynom fast unabhängig von der Dimension bis auf den Faktor 10^{-3} durch Terme erster und zweiter Ordnung beschrieben werden kann.

Polynom 2

Wir betrachten homogene Polynome vom Grad $\leq n$

$$P_n(\boldsymbol{x}) := \sum_{|\boldsymbol{\alpha}|_1 \leq n} C_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}, \tag{4.2}$$

deren Koeffizienten $C_{\boldsymbol{\alpha}}$ wir durch

$$C_{\boldsymbol{\alpha}} = \begin{cases} 1 & |\boldsymbol{\alpha}|_1 > 1 \text{ und } \alpha_i \leq 1 \text{ für alle } i = 1, \ldots, d \\ 0 & \textit{sonst} \end{cases} \tag{4.3}$$

definieren, was der Form

$$P_n(\boldsymbol{x}) = \sum_{i<j} x_i x_j + \sum_{i<j<k} x_i x_j x_k + \ldots$$

entspricht.

Dieser Fall ist besonders für die Approximation durch dünne Gittern interessant, da sie derart konstruiert sind, dass keinerlei Varianzmasse auf den Achsen liegt, sondern nur die gemisch-

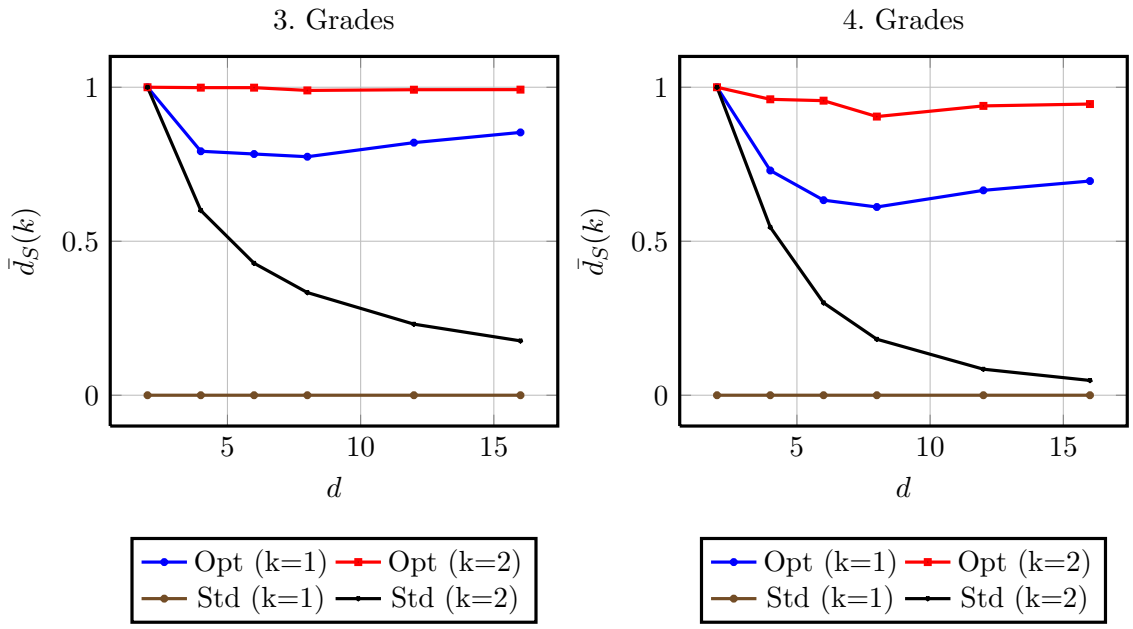


Abb. 4.3: Relative Superpositionsdimension $\bar{d}_S(k)$ von Polynom 2 für $n \in \{3, 4\}$.

ten Monome (im Kontext einer Taylorreihe also die gemischten Ableitungen) zu P_n beitragen. In Abschnitt 5.1 werden wir solche Funktionen hinsichtlich ihrer Darstellbarkeit in einer stückweise-linearen Dünngitterbasis noch näher untersuchen.

In Abbildung 4.3 sind die relativen Superpositionsdimensionen dargestellt. Man sieht, dass die ANOVA-Terme erster Ordnung des untransformierten Polynoms nicht zur Gesamtvarianz beitragen und der Anteil der Terme zweiter Ordnung mit steigender Dimension stark abfällt.

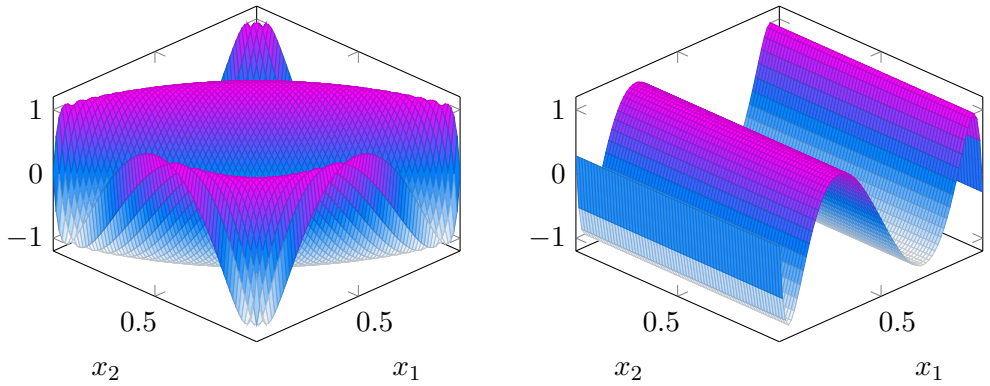
Wir wenden wieder Verfahren Ib an, wobei wir $\hat{\gamma}_{\mathbf{u}} = 1$ für $|\mathbf{u}| = 1$, $\hat{\gamma}_{\mathbf{u}} = e^{-1}$ für $|\mathbf{u}| = 2$ und $\hat{\gamma}_{\mathbf{u}} = 0$ für $|\mathbf{u}| > 2$ wählen.

Für das somit dimensionsoptimierte Polynom sind die relativen Superpositionsdimensionen $\bar{d}_S(k)$ zwar nicht mehr konstant in d, ihr Abfall ist jedoch deutlich geringer - so dass man durchaus behaupten kann, den Fluch der Dimension für diese Art von Problemen durch eine geeignete orthogonale Transformation brechen zu können.

Die Tatsache, dass der relative Anteil der Terme erster Ordnung erst leicht abfällt, bei $d = 12,\ 16$ dann jedoch wieder größer wird, erklären wir dadurch, dass unser Liniensuchverfahren vermutlich lokale Minima findet, also nicht immer ein globales Minimum von $\bar{\mathfrak{M}}$ erreicht.

4.2 Synthetische Testfunktionen

An dieser Stelle wollen wir komplexere Funktionen konstruieren, anhand derer wir die Wirksamkeit unserer Verfahren testen können. Insbesondere betrachten wir Funktionen, für deren



(a) Std (b) Opt

Abb. 4.4: $\sin(\boldsymbol{b}^t\boldsymbol{x})$ als Beispiel für eine Ridge-Funktion, die achsenparallel ausgerichtet werden kann.

ANOVA-Terme und Varianzen keine analytischen Lösungen bekannt sind. Wir testen daher Verfahren I (a+b) mit numerischen Quadraturverfahren, wie den dünnen Gittern (SG) (siehe Abschnitt 5.2.2) und Quasi-Monte Carlo Folgen (QMC) (siehe Abschnitt 5.2.1) und Verfahren II mit verschiedenen Polynomgraden (Poly1, Poly2, Poly3). Wir vergleichen die Resultate mit den in der Literatur gebräuchlichen Verfahren LT und DM, sofern diese anwendbar sind.

Ridge-Funktionen

Eine Funktion $f : \mathbb{R}^d \to \mathbb{R}$ heißt *Ridge-Funktion*, wenn sie eine Darstellung

$$f(\boldsymbol{x}) = g\left(w_0 + \sum_{i=1}^{d} w_i x_i\right) \tag{4.4}$$

besitzt, wobei $g : \mathbb{R} \to \mathbb{R}$ eine beliebige eindimensionale Funktion bezeichne.

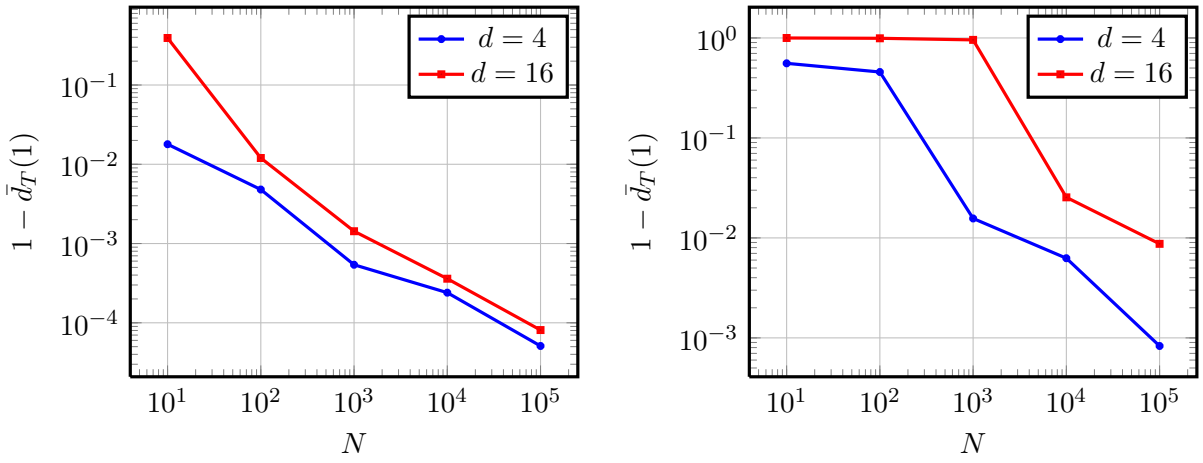
In Satz 3.17 haben wir bewiesen, dass eine Drehung des Koordinatensystems existiert, bezüglich welcher f eine Darstellung als eindimensionale Funktion besitzt. Anhand von zwei Beispielfunktionen wollen wir auch hier verifizieren, dass unsere Verfahren in der Lage sind, diese Koordinatentransformation aufzufinden.

Wir betrachten die Testfunktionen $f : \mathbb{R}^d \to \mathbb{R}$

$$f(\boldsymbol{x}) := \exp(\boldsymbol{w}^t\boldsymbol{x}), \quad \boldsymbol{w} = \frac{1}{\sqrt{d}}(1,\ldots,1)^t \tag{4.5}$$

und

$$f(\boldsymbol{x}) := \sin(2\ \boldsymbol{w}^t\boldsymbol{x}), \quad \boldsymbol{w} = \frac{1}{\sqrt{d}}(1,\ldots,1)^t, \tag{4.6}$$



(a) glatte Ridge Funktion (4.5)

(b) Ridge Funktion mit Sprung (4.7)

Abb. 4.5: Verfahren I konvergiert für die Funktionen (4.5) und (4.7) mit steigender Integrationsgenauigkeit gegen die analytische Lösung

als Beispiele für glatte Ridge-Funktionen und

$$f(\boldsymbol{x}) := \max\left(\exp(\boldsymbol{w}^t\boldsymbol{x}) - \frac{3}{2}, 0\right), \quad \boldsymbol{w} = \frac{1}{\sqrt{d}}(1, \ldots, 1)^t \tag{4.7}$$

als ein Beispiel für eine Funktion mit Sprung. Zudem gilt für (4.7), dass sowohl der Gradient, als auch die Hessematrix im Entwicklungspunkt $\boldsymbol{a} = \boldsymbol{0}$ identisch Null sind, was ein einfaches Beispiel für Funktionen darstellt, bei denen die LT und die DM nutzlos sind.

In Abbildung 4.5 haben wir den Abstand der relativen Trunkationsdimension $\bar{d}_T(1)$ vom theoretisch optimalen Wert 1 in doppelt-logarithmischen Achsen dargestellt. Man sieht, dass das Verfahren mit steigender Integrationsgenauigkeit gegen die analytische Lösung konvergiert. Für die glatte Funktion (4.5) haben wir ein reguläres dünnes Gitter mit Clenshaw-Curtis Diskretisierung [Hol08] verwendet, während bei der Sprung-Funktion (4.7) eine entsprechende Zahl von Quasi-Monte Carlo Punkten der Sobol Folge zum Einsatz kam.

Im letzteren Fall ist deutlich zu sehen, dass sich eine Konvergenz erst dann einstellt, wenn hinreichend viele Stützstellen hinzugenommen werden. Dies erklären wir damit, dass auf einem großen Teil des Integrationsgebiets der Integrand identisch Null ist, weshalb mit wenigen Punkten im eigentlich interessanten Bereich keine gute Approximation an f gelingen kann und die dimensionale Struktur daher nicht korrekt erfasst wird.

In Tabelle 4.1 vergleichen wir nun in $d = 8$ Dimensionen für die Funktionen (4.6) und (4.7) die Lineare Transformation und die Diagonal Methode aus Abschnitt 3.6.1 mit den von uns entwickelten Verfahren. Dabei betrachten wir zum einen die direkte Auswertung der Integrale mit numerischer Quadratur durch dünne Gitter im glatten und QMC im nicht-glatten Fall. Zum anderen projizieren wir die Funktionen vermöge Verfahren II in die Polynomräume P_1, P_2

und P_3 und führen dort die Optimierung mit der analytischen Auswertung der Integrale durch.

Wir sehen, dass für die glatte Ridge-Funktion (4.6) alle betrachteten Verfahren die optimale Drehung auf eine eindimensionale Funktion finden. Im Fall der Sprungfunktion (4.7) sind die LT und die DM jedoch nicht zu gebrauchen, da die Funktion selbst und auch ihre Differentiale am Punkte $\boldsymbol{a} = \mathbf{0}$ identisch Null sind. Die Projektionsmethoden und die direkte Quadraturmethode finden jedoch auch hier gleichermaßen die optimale Drehung auf eine achsenparallele Funktion.

Verallgemeinerte Ridge-Funktionen

Ridge-Funktionen werden durch eine eindimensionale Funktion $g : \mathbb{R} \to \mathbb{R}$ und einen Richtungsvektor $\boldsymbol{w}$ definiert. Eine Verallgemeinerung dieses Ansatzes erhält man, wenn man für $i = 1, \ldots, d$ Funktionen der Form $g_i : \mathbb{R} \to \mathbb{R}$ und paarweise orthogonale Richtungsvektoren $\boldsymbol{w}^{(i)}$ zu einer Abbildung

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} g_i(w_0^{(i)} + (\boldsymbol{w}^{(i)})^t \boldsymbol{x}) \tag{4.8}$$

zusammenfasst.

Im Folgenden wählen wir die Richtungsvektoren $\boldsymbol{w}^{(1)}, \ldots \boldsymbol{w}^{(d)}$ als die Spalten einer zufälligen orthonormalen Matrix $\tilde{\boldsymbol{Q}}$ und untersuchen, inwiefern unser Verfahren die optimale Drehung $\boldsymbol{Q} = \tilde{\boldsymbol{Q}}^t$ finden kann.

Dazu betrachten wir zum einen

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} \sin(2\, \tilde{\boldsymbol{Q}}_{\cdot,i}^t \boldsymbol{x})) \tag{4.9}$$

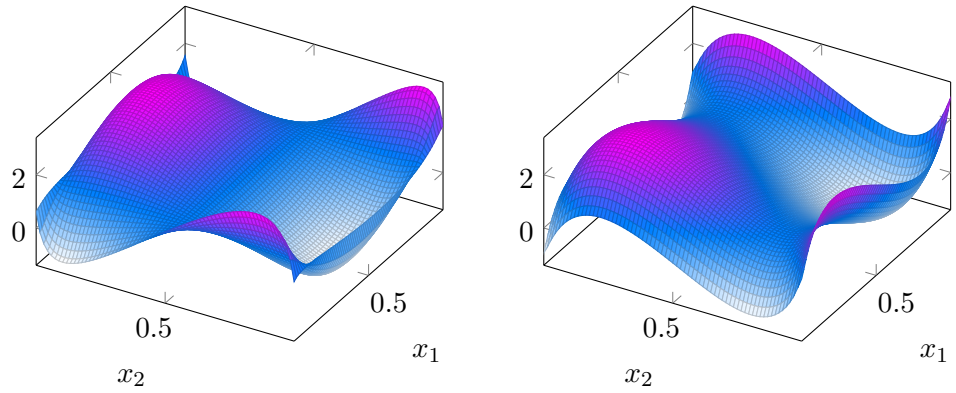
als eine glatte Testfunktion und

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} \max\left\{\exp(\tilde{\boldsymbol{Q}}_{\cdot,i}^t \boldsymbol{x}) - 1, 0\right\} \tag{4.10}$$

als eine Testfunktion mit Sprüngen.

	$\bar{d}_T(1)$ (Trunkation)		$\bar{d}_S(1)$ (Superposition)	
Methode	sin (4.6)	Sprung (4.7)	sin (4.9)	Sprung (4.10)
Std	0.008	0.592	0.194	0.698
LT	1.000	-	0.170	-
DM	1.000	-	0.222	-
Poly1	1.000	1.000	0.171	0.701
Poly2	1.000	1.000	0.373	0.901
Poly3	1.000	1.000	1.000	0.966
$\mathrm{SG}_{10^5}/\mathrm{Qmc}_{10^5}$	0.999	0.999	0.999	0.973

Tabelle 4.1: Vergleich verschiedener Reduktionsmethoden in $d = 8$.



(a) Std (b) Opt

Abb. 4.6: Verallgemeinerte Ridge-Funktion (4.9).

In $d = 8$ Dimensionen haben wir in Tabelle 4.1 die LT, die DM, die Polynommethoden zum Grad $n = 1, 2, 3$ und die direkten Quadraturmethoden mit dünnen Gittern (glatte Funktion (4.9)) und mit Quasi-Monte Carlo (Sprungfunktion (4.10)) verglichen.

Wir stellen fest, dass die einzigen Methoden, die hier noch zuverlässig funktionieren die Polynommethode vom Grad 3 und die direkten Quadraturmethoden sind. Das lineare und das quadratische Modell reichen für dieses Beispiel nicht mehr aus, um die dimensionale Struktur der Funktion genügend gut aufzulösen.

Im Falle der nicht-glatten Funktion (4.10) erzielt man selbst mit der direkten Quadratur nur noch einen Wert von 0.973, was wir darauf zurückführen, dass die eindimensionalen Funktionen Sprünge besitzen, welche durch die eingeschobene Drehung schräg im Raum liegen, wodurch sich die Konvergenz der Quadraturverfahren drastisch verschlechtert. Auch die Approximation mit einem glatten Polynom kann hier nicht mehr gut funktionieren, selbst wenn es sich um die orthogonale $\mathcal{L}^2$-Projektion handelt.

Spiralförmige Funktionen

Wir betrachten die Funktion

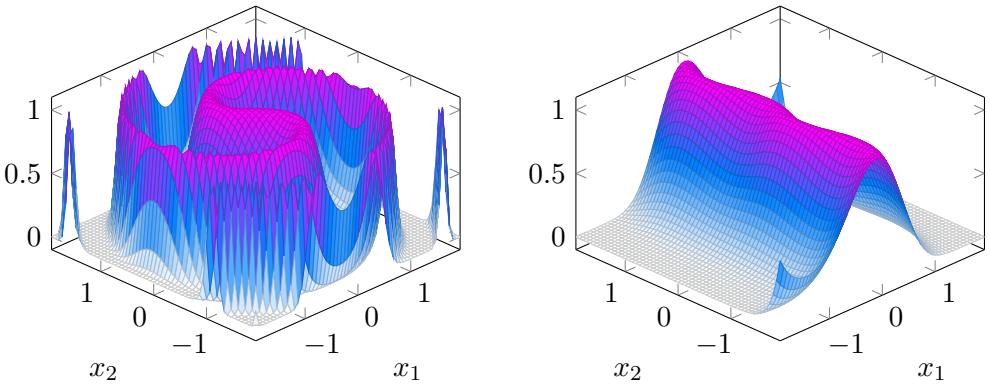
$$f(x_1, x_2) = \exp\left(-\left(\cos(x^2 + y^2)(x + y) + \sin(x^2 + y^2)(x - y)\right)^2\right), \tag{4.11}$$

welche in Abbildung 4.7(a) dargestellt ist.

Wir bemerken, dass in [Gar04] ein ähnliches Testproblem als Benchmark für Klassifikationsalgorithmen im Data-Mining betrachtet wird.

Wir reduzieren nun die effektive Dimension, indem wir Verfahren III anwenden, wobei die auftretenden Integrale mit 50.000 Quasi-Monte Carlo Punkten approximiert werden.

Während die untransformierte Funktion (4.11) eine relative Superpositionsdimension von



(a) Std (b) Opt

Abb. 4.7: Spiralfunktion (4.11) in kartesischen und in gekrümmten Koordinaten.

$\bar{d}_S(1) = 0.506$ besitzt, gilt für die durch eine stückweise-orthogonale Abbildung transformierte Funktion $f \circ \hat{\boldsymbol{Q}}$, dass $\bar{d}_S(1) = 0.983$. In Abbildung 5.4(b) ist die Konvergenzgeschwindigkeit des für die Minimierung verwendeten CG-Verfahrens dargestellt.

4.3 Bewertung von Optionspreisen

Ein verbreitetes Problem in der Finanzmathemtik ist die Berechnung eines „fairen“ Wertes von Derivaten, welche der Emmittent dann mit einem Ausgabeaufschlag (seinem Gewinn) an private oder institutionelle Anleger verkauft. Dazu modelliert man die Kursbewegungen der zugrundeliegenden Wertpapiere durch stochastische Prozesse und integriert eine auf diesen definierte Auszahlungsfunktion über der Menge aller Pfadrealisierungen.

Modelliert man die Kursbewegung nach dem Black-Scholes Modell durch eine geometrische Brownsche Bewegung [Gla04], so erhält man Integrale der Form

$$E[f] = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\boldsymbol{C})}} \int_{\mathbb{R}^d} f(\boldsymbol{w}) \exp\left(-\frac{1}{2}\boldsymbol{w}^t \boldsymbol{C}^{-1} \boldsymbol{w}\right) d\boldsymbol{w}, \tag{4.12}$$

wobei

- $\boldsymbol{w} \in \mathbb{R}^d$ den Werten eines (zeitdiskreten) stochastischen Prozesses[1],
- $\boldsymbol{C}$ dessen Kovarianz-Matrix und
- $f : \mathbb{R}^d \to \mathbb{R}$ einer auf dem Prozess definierten Funktion

entsprechen.

[1] In diesem Fall der Wiener-Prozess [Gla04].

Zerlegt man die Kovarianzmatrix in $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^t$ und substituiert $\boldsymbol{w} = \boldsymbol{A}\boldsymbol{x}$, so lässt sich das Integral in der Form

$$\int_{\mathbb{R}^d} f(\boldsymbol{A}\boldsymbol{x})\, \varphi^d(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.13}$$

darstellen, wobei $\varphi^d(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\boldsymbol{x}^t\boldsymbol{x}}{2}}$ die Dichtefunktion der d-dimensionalen Standardnormalverteilung bezeichnet.

Da die Zerlegung $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^t$ nicht eindeutig ist, gibt es verschiedene Möglichkeiten die Pfade von $\boldsymbol{w}$ zu generieren, welche anschaulich in der aus [Hol08] entnommenen Abbildung 4.8 dargestellt sind.

Random Walk (RW): Der Pfad des Prozesses wird sequentiell erzeugt – dadurch besitzt der Prozess bezüglich jedem Zeitschritt (und damit bezüglich jeder Eingansdimension) die gleiche Varianz. Die Matrix $\boldsymbol{A}$ entspricht in diesem Falle der Cholesky-Zerlegung von $\boldsymbol{C}$.

Brownsche Brücke (BB): Der Pfad des Prozesses wird hierarchisch generiert, wodurch führende Dimensionen einen größeren Beitrag zur Varianz liefern. Details zur Konstruktion und Eigenschaften der BB finden sich in [MC96].

Karhunen Loève Transformation (PCA): Die Matrix $\boldsymbol{C}$ wird durch eine Eigenwertzerlegung von $\boldsymbol{A}$ bestimmt. Dies entspricht einer stochastischen Fourieranalyse.

Allgemeiner gilt für $\boldsymbol{Q} \in \mathbf{O}(d)$ wegen $\boldsymbol{C} = (\boldsymbol{A}\boldsymbol{Q})(\boldsymbol{A}\boldsymbol{Q})^t$, dass man eine gültige Matrix $\boldsymbol{A}$ mit einer beliebigen orthogonalen Matrix $\boldsymbol{Q}$ multiplizieren kann, was wir uns im Folgenden zunutze machen werden, um die LT, die DM, sowie die von uns entwickelte Polynommethode (Poly1, Poly2, Poly3) und das Verfahren mit direkter Quadratur (Qmc_{10^5}) anzuwenden.

Wir vergleichen diese Ansätze mit den oben beschriebenen Pfadkonstruktionen RW, BB und PCA. Dabei werden wir feststellen, dass BB und PCA die effektive Dimension asiatischer Optionen deutlich reduzieren können, sich durch die Anpassung an die Struktur der konkreten Funktion f jedoch weitere substanzielle Verbesserungen erzielen lassen.

Asiatische Optionen

Wir betrachten das verbreitete Testproblem asiatischer Optionen, die durch eine Ausahlungsfunktion definiert werden, welche über den Kurs des zugrundeliegenden Wertes an d verschiedenen Zeitpunkten mittelt [Hol08, IT06].

Dabei verwenden wir sowohl für das geometrische Mittel, als auch für das arithmetische Mittel den Parametersatz

$$S(0) = 100, \sigma = 0.2, r = 0.1, T = 1 \text{ und } K = 100,$$

wobei $S(0)$ den Kursstand zum Zeitpunkt $t_0 = 0$, σ die Volatilität, r den risikofreien Zinssatz, T den Ausübungszeitpunkt und K den Ausübungspreis bezeichne.

Die gleichen Paramter werden auch in [WF03, Hol08, WS03, Wan06, SW97] verwendet.

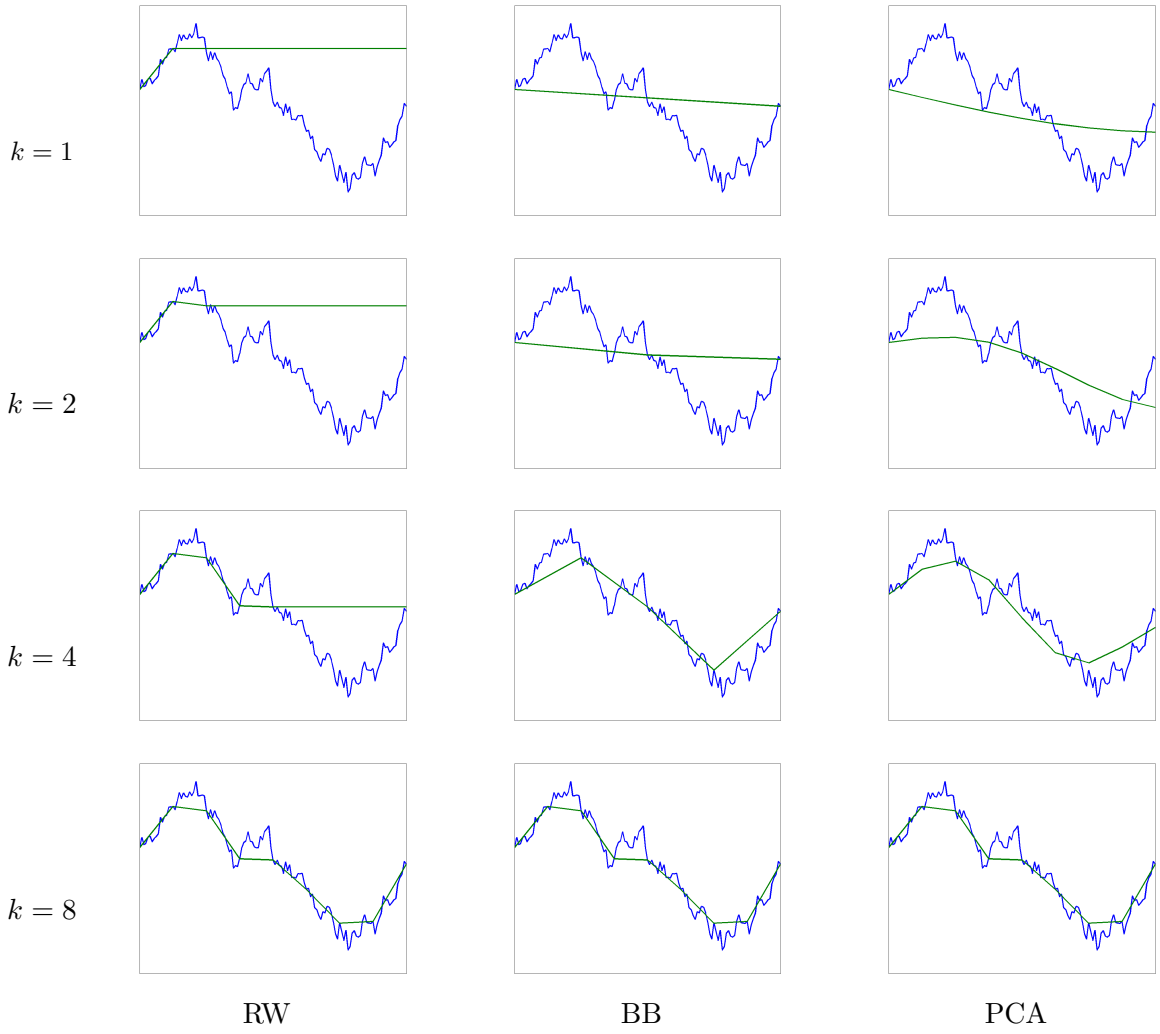


Abb. 4.8: Pfadkonstruktion mit RW, BB und PCA

Außerdem wollen wir auch noch die Fälle $K = 120$ und $K = 0$ betrachten. Im ersten Fall erhält man einen Integranden, der auf dem grösten Teil des Integrationsgebietes identisch Null ist, wodurch auf der Taylorreihe basierende Verfahren, wie die LT und die DM nicht mehr angewendet werden können. Im zweiten Fall erhalten wir einen auf ganz $\mathbb{R}^d$ stetig differenzierbaren Integranden, was in Abschnitt 5.2 für die Integration mit dünnen Gittern eine Rolle spielen wird.

In Tabelle 4.2 haben wir sämtliche (uns bekannten) Dimensionsreduktionverfahren für Funktionen diesen Types untersucht. „Std“ bezeichnet dabei stets die untransfomierte Funktion, „BB“ die Pfaddiskretisierung mit der Brownschen Brücke, „PCA“ die Karhunen-Loève Transformation (also die Hauptachsentransformation der Kovarianzmatrix), „LT“ die Lineare Transformation nach Imai/Tan und „DM“ die Hauptachsentransformation der Hessematrix, welche in der

Literatur teilweise als “Diagonal Methode“ bezeichnet wird [Mor98].

Die Verfahren „Poly1“„Poly2“ und „Poly3“ entsprechen unserer Polynommethode, also der orthogonalen Projektion des Integranden auf den Raum der homogenen Polynome mit anschließender analytischer Berechnung der optimalen Drehung im Raum der Polynome und „Qmc_{10^5}“ bezeichnet das Verfahren Ia, wobei die auftretenden Integrale mit 10^5 Quasi-Montecarlo Punkten (Sobol) approximiert werden.

	Geometrisches Mittel			Arithmetisches Mittel		
Methode	$K = 0$	$K = 100$	$K = 120$	$K = 0$	$K = 100$	$K = 120$
Std	0.170	0.144	0.066	0.166	0.138	0.062
BB	0.771	0.735	0.596	0.778	0.749	0.634
PCA	0.986	0.982	0.967	0.987	0.987	0.983
LT	1.000	1.000	-	0.999	0.999	-
DM	1.000	1.000	-	0.986	0.985	-
Poly1	1.000	1.000	0.998	0.999	0.999	0.992
Poly2	0.999	0.985	0.989	0.991	0.984	0.989
Poly3	0.999	0.871	0.951	0.934	0.785	0.954
Qmc_{10^5}	0.998	0.997	0.974	0.999	0.995	0.997

Tabelle 4.2: Relative Trunkationsdimension $\bar{d}_T(1)$ von 16-dimensionalen asiatische Optionen des geometischen und des arithmetischen Typs.

Geometrische Asiatische Option

Geometrische asiatische Optionen werden durch die Auszahlungsfunktion

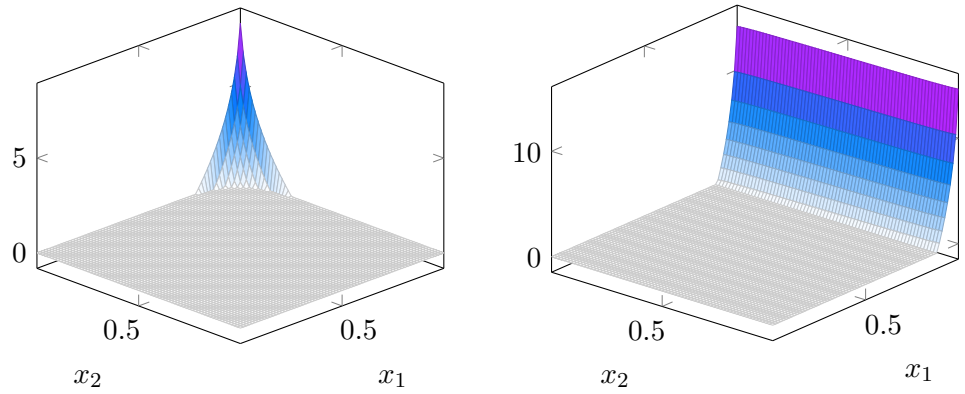
$$A(S(t_1), \ldots, S(t_d)) := \max\left(\prod_{i=1}^{d} (S(t_i))^{1/d} - K, 0\right) \tag{4.14}$$

definiert, wobei die $S(t_i)$ den durch eine geometrische Brownsche Bewegung modellierten Kursständen des zugrundeliegenden Wertpapieres zum Zeitpunkt t_i entsprechen.

Geometrische asiatische Optionen sind ein häufig verwendetes Modellproblem, da eine analytische Lösung für ihren Erwartungswert existiert.

Interessant ist, dass sich der durch (4.14) ergebende Integrand durch eine geeignete orthogonale Transformation auf eine eindimensionale Funktion reduzieren lässt. In Tabelle 4.2 sind in den ersten drei Spalten die effektiven Trunkationsdimensionen $\bar{d}_T(1)$, also der Anteil der ersten Eingangsdimension an der Gesamtvarianz für die verschiedenen Dimensionsreduktionsmethoden dargestellt.

Interessant ist hierbei, dass ein höherer Polynomgrad nicht immer zu einem besseren Ergebnis führt, was wir darauf zurückführen, dass der Projektor auf den Polynomraum auch den Bereich aufzulösen versucht, in dem der Integrand indentisch null ist.



(a) Std (RandomWalk) (b) Opt (Poly1)

Abb. 4.9: Asiatische Option, geometrisches Mittel, $K = 120$.

Die Methode der direkten Quadratur scheint in sämtlichen betrachteten Fällen zuverlässig zu funktionieren, allerdings ist sie auch mit Abstand die aufwändigste – für die in Tab. 4.2 dargestellten Probleme benötigt die Minimierung des Funktionals $\mathfrak{M}$ ca. 30 Minuten, während die poly3 Methode bereits nach ca. 5 Minuten das Minimum erreicht hat[2].

In Abbildung 4.9 betrachten wir für den Fall $d = 16$ und $K = 120$ sowohl den Standardintegranden, als auch seine dimensionsoptimierte Variante. Der Effekt der Drehung ist offensichtlich. Deutlich wird ebenfalls, weshalb die Differentiationsmethoden, wie die LT und die DM hier keine brauchbaren Ergebnisse liefern können, da die Funktion am Punkt der Taylorentwicklung $\boldsymbol{a} = \mathbf{0}$ identisch Null ist. Um LT und DM in solchen Fällen anwenden zu können, wäre es nötig, verschiedene Entwicklungspunkte durchzutesten, bis man einen brauchbaren Gradienten oder Hessematrix erhält.

Arithmetische Asiatische Option

Arithmetische asiatische Optionen werden durch die Auszahlungsfunktion

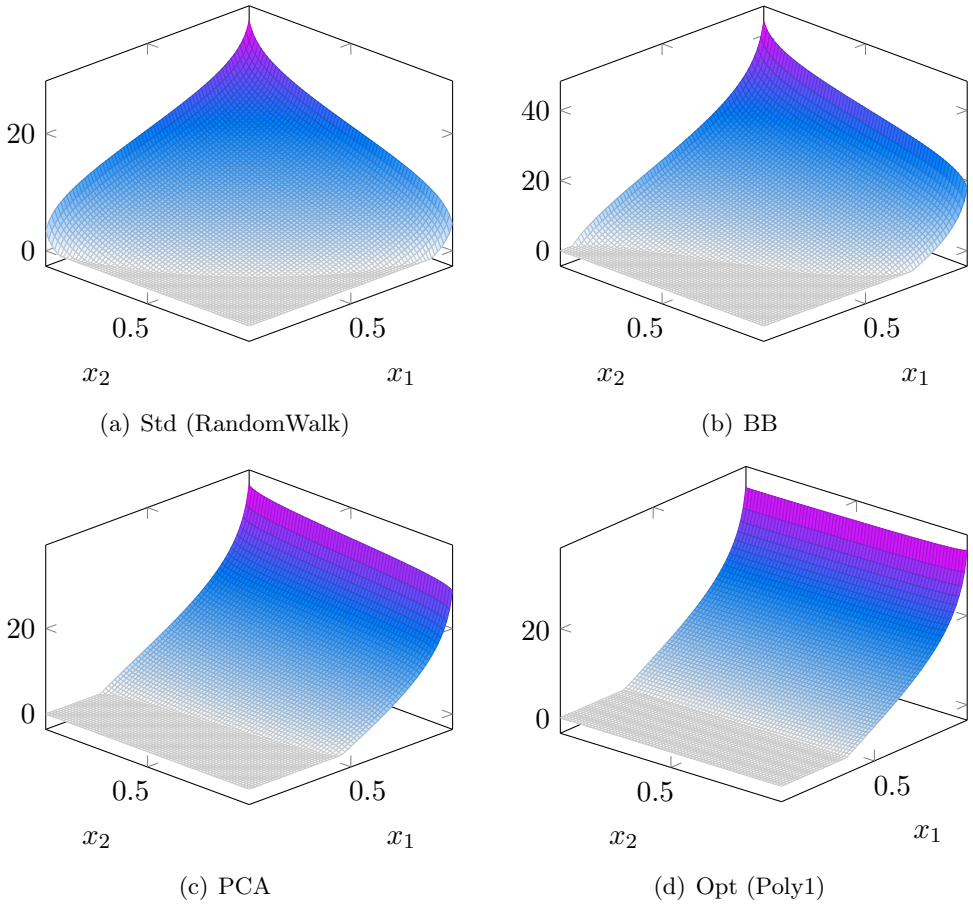
$$A(S(t_1), \ldots, S(t_d)) := \max\left(\frac{1}{d}\sum_{i=1}^{d} S(t_i) - K, 0\right) \tag{4.15}$$

definiert.

In den letzten drei Spalten von Tabelle 4.2 sind die relativen Trunkationsdimensionen $\bar{d}_T$ für das arithmetische Mittel zu verschiedenen Ausübungspreisen K dargestellt.

Eine Transformation auf eine tatsächlich eindimensionale Funktion ist nicht mehr möglich. Es lassen sich jedoch Drehungen finden, so dass sich in allen drei Fällen $K = 0, 100, 120$ für den Toleranzparameter $\alpha = 0.99$ ein *effektiv* eindimensionaler Integrand ergibt.

[2]Verwendet wurde ein System mit Intel Xeon CPU X7460 mit 2.66GHz und 16384 KB Cache.



(a) Std (RandomWalk) (b) BB

(c) PCA (d) Opt (Poly1)

Abb. 4.10: Verschiedene Pfadkonstruktionen, asiatische Option, arithmetisches Mittel, $K = 100$.

Basket-Optionen

Im Folgenden betrachten wir Europäische Basket-Call Optionen. Die Kurzverläufe der Underlyings (dem „Basket“) werden nach dem Black-Scholes Modell modelliert – wir verwenden den Parametersatz

$$S_i(0) = 100, \sigma_i = 0.2, r = 0.1, T = 1 \text{ und } K = 100,$$

wobei $i = 1, \ldots, d$ den im Basket enthaltenen Underlyings entspricht.

Als Korrelationskoeffizient betrachten wir zum einen den Fall eines stark korrelierenden Baskets ($\varrho_{ij} = 0.3$), sowie den Fall eines unkorrelierten Baskets ($\varrho_{ij} = 0$). Ersteres könnte etwa einem Call auf das Mittel einer ganzen Branche (etwa Automobilhersteller im DAX) entsprechen, während der zweite Fall einer durch Risiko-Diversifikation motivierten Anlagestrategie entspricht.

Methode	Stark korreliert ($\varrho = 0.3$) $\bar{d}_T(1)$	$\bar{d}_T(2)$	Unkorreliert ($\varrho = 0$) $\bar{d}_T(1)$	$\bar{d}_T(2)$
Std	0.3444	0.5287	0.0624	0.1249
PCA	0.9999	0.9999	0.0624	0.1249
LT	0.9999	0.9999	0.9992	0.9928
DM	0.0014	0.9598	0.9992	0.9993
Poly1	0.9999	0.9999	0.9992	0.9991
Poly2	0.9990	0.9999	0.9982	0.9993
Poly3	0.9991	0.9999	0.9931	0.9993
Qmc_{10^5}	0.9997	0.9990	0.9916	0.9991

Tabelle 4.3: $\bar{d}_T(1)$ und $\bar{d}_T(2)$ einer 16-dimensionalen Basket-Option mit stark- und schwachkorrelierenden Underlyings.

Wir definieren arithmetische Basket Optionen durch die Auszahlungsfunktion

$$A(S_1(T), \ldots, S_d(T) := \max\left(\frac{1}{d}\sum_{i=1}^{d} S_i(T) - K, 0\right), \tag{4.16}$$

d.h. man betrachtet das gleichmäßige arithmetische Mittel über die Kursstände der Underlyings zum Ausübungszeitpunkt T.

In Tabelle 4.3 sind ist der Varianzanteil der ersten beiden Dimensionen für korrelierte und unkorrelierte Baskets dargestellt. Im korrelierten Fall liefern bis auf die DM alle betrachteten Verfahren eine deutliche Reduktion der effektiven Trunkationsdimension, bzw. eine starke Varianzkonzentration in den ersten beiden Eingansvariablen.

Die kontraproduktive Drehung durch die DM erklären wir mit einer ungünstigen Hessematrix am Entwicklungspunkt $\boldsymbol{a} = \mathbf{0}$.

Interessant ist, dass sowohl DM, als auch LT und die Poly1-Methode im unkorrelierten Fall den gleichen Wert $\bar{d}_T(1) = 0.9992$ liefern, der von keiner anderen der von uns untersuchten Methoden übertroffen wird.

Wie in Abbildung 5.14 auch anschaulich zu sehen ist versagt im unkorrelierten Fall jedoch die PCA, da die Kovarianzmatrix für $\varrho = 0$ bereits Diagonalstruktur besitzt. Alle anderen Methoden finden aber Koordinatensysteme, in denen die Funktion (zum Parameter $\alpha = 0.99$) effektiv eindimensional ist.

5 Anwendung auf hochdimensionale Probleme

Nachdem wir festgestellt haben, dass es für eine Vielzahl von Funktionen möglich ist, deren effektive Dimension zu verringern, werden wir in diesem Kapitel nun untersuchen, inwiefern die von uns entwickelten Verfahren, sowie andere aus der Literatur bekannte Dimensionsreduktionsmethoden geeignet sind, hochdimensionale Algorithmen, wie dünne Gitter zur Interpolation, Integration und Approximation, sowie die Quasi-Monte Carlo Quadratur Verfahren zu beschleunigen.

Intuitiv sollte dies möglich sein, da sowohl für die dünnen Gitter, als auch für Quasi-Monte Carlo Verfahren ein Zusammenhang zwischen deren Konvergenzeigenschaften und der effektiven Dimension festgestellt wurde [Hol08, Feu10, SW97, WF03, IT04].

Zu diesem Zweck untersuchen wir ausgewählte Testfunktionen $f : \mathbb{R}^d \to \mathbb{R}$, deren reguläre und deren dimensionsoptimierten Versionen wir vor dem Hintergrund verschiedener Anwendungen miteinander vergleichen werden.

Dabei betrachten wir:

Interpolation: Wir verwenden ein Verfahren zur Interpolation mit stückweise-linearen, orts- und dimensionsadaptiven dünnen Gitter aus [Feu10] und messen die Approximationsgüte in der $\mathcal{L}^2$-Norm.

Quadratur: Wir untersuchen die Konvergenzeigenschaften der Quasi-Monte Carlo Quadratur mit Sobol- und Halton Folgen, sowie verschiedene dimensionsadaptive Dünngitter-Quadratur Verfahren aus [Hol08]. Wir betrachten sowohl Einheitswürfel (Clenshaw-Curtis, Gauß-Legendre), als auch die direkte Intgegration über dem Ganzraum $\mathbb{R}^d$ mit Gauß-Hermite.

Regression: Wir verwenden ein Verfahren aus [Boh10], welches im wesentlichen eine orts-adaptive Variante der in [Gar04] vorgestellten Ideen darstellt und untersuchen, inwiefern eine geeignete Drehung des Latentspace die Güte der Dünngitterapproximation an die Daten verbessern kann.

5.1 Interpolation mit Dünnen Gittern

Dünne Gitter (eng. *sparse grids*) erlauben eine effiziente Diskretisierung hochdimensionaler Funktionen[1] und basieren auf einem Tensorprodukt eindimensionaler Multiskalenbasen.

Da die Zahl ihrer Freiheitsgrade im Gegensatz zum vollen Gitter (n^d) nur $\mathcal{O}(n \log(n)^{d-1})$ beträgt, haben dünne Gitter in den letzten Jahren in vielen Bereichen, in denen hochdimensionale

[1] Sofern deren gemischte Ableitungen beschränkt sind.

Probleme auftreten an Popularität gewonnen, da die Maschenweite $\frac{1}{n}$ nur noch in einem logarithmischen Term exponentiell in die Komplexität eingeht und DG damit den Fluch der Dimension brechen können.

So werden dünne Gitter beispielsweise als Ansatzraum bei der Lösung von partiellen Differentialgleichungen [Feu10, Rei04, LO08], bei der Quadratur [Hol08, GH10b, GG98], bei der Erkennung von Mannigfaltigkeiten [FG09, Hul09] und für die multivariate Regression [Gar04, GGG10, Boh10] erfolgreich eingesetzt.

Im Folgenden geben wir eine kurze Einführung in das Konzept der dünnen Gittern – für eine ausführliche Auseinandersetzung sei auf [Gri06, BG04, Feu10] verwiesen. Im Anschluss werden wir untersuchen, inwiefern sich die Approximationsgüte von dünnen Gittern durch eine Reduktion der effektiven Dimension weiter verbessern lässt.

5.1.1 Hierarchische Basis und Dünne Gitter

Zu einem gegebenen Level l, einer Maschenweite h_l und Indizes $j = 0, \dots, 2^l$ seien die Hütchenfunktionen

$$\phi(x) := \begin{cases} 1 - |x| & x \in [-1, 1] \\ 0 & sonst \end{cases}$$

und damit $\phi_{l,j}(x) := \phi(\frac{x - jh_l}{h_l})$ auf $[(j-1)h_l, (j+1)h_l]$ definiert.

Mit $V_l := \text{span}\{\phi_{l,j} : j = 0, \dots 2^l\}$ bezeichnen wir den Vektorraum dieser Funktionen und konstruieren daraus die Differenzenräume $W_l := V_l - V_{l-1}$. Offensichtlich gilt damit $V_l = \oplus_{i=1}^{l} W_i$.

Zu einem Multiindex $\boldsymbol{l} \in \mathbb{N}^d$ und dem dazu zugehörigen Maschenweitenvektor

$$\boldsymbol{h} := \mathbf{2}^{-\boldsymbol{l}} = (2^{-l_1}, 2^{-l_2}, \dots, 2^{-l_d})$$

können wir dann d-dimensionale Hütchenfunktion

$$\phi_{\boldsymbol{l},\boldsymbol{j}}(\boldsymbol{x}) := \prod_{k=1}^{d} \phi_{l_k,j_k}(x_k)$$

konstruieren. Mit $V_{\boldsymbol{l}}$ wollen wir wieder den Vektorraum aller d-dimensionalen Hütchenfunktionen zu einem gegebenen Level $\boldsymbol{l} = (l_1, \dots, l_d)$ bezeichen – $W_{\boldsymbol{l}}$ bezeichne die zugehörigen Differenzenräume:

$$V_{\boldsymbol{l}} := \text{span}\{\phi_{\boldsymbol{l},\boldsymbol{j}} : \boldsymbol{j} = \mathbf{0}, \dots \mathbf{2}^{\boldsymbol{l}}\}, \quad W_{\boldsymbol{l}} := \text{span}\{\phi_{\boldsymbol{l},\boldsymbol{j}} : \boldsymbol{j} = \mathbf{0}, \dots \mathbf{2}^{\boldsymbol{l}}, l_k \text{ gerade}\}.$$

Entsprechend ist

$$V_{\boldsymbol{l}} = \bigoplus_{\mathbf{0} \le \boldsymbol{j} \le \mathbf{2}^{\boldsymbol{l}}} W_{\boldsymbol{j}} \quad \text{und} \quad V = \bigoplus_{\boldsymbol{j} \in \mathbb{N}^d} W_{\boldsymbol{j}},$$

wobei die Ausdrücke $\mathbf{2}^{\boldsymbol{l}}$ und $\mathbf{0} \le \boldsymbol{j}$ komponentenweise zu verstehen sind.

Wir definieren die zum einem Levelindex $\boldsymbol{l}$ gehörende Indexmenge als

$$\mathbf{I}(\boldsymbol{l}) := \{\boldsymbol{j} = \mathbf{0}, \ldots \mathbf{2}^{\boldsymbol{l}}, l_k \text{ gerade}\},$$

und definieren damit Gitterfunktionen durch

$$f(\boldsymbol{x}) := \sum_{\boldsymbol{l} \in \mathcal{C}} f_{\boldsymbol{l}}, \quad f_{\boldsymbol{l}} := \sum_{\boldsymbol{j} \in \mathbf{I}(\boldsymbol{l})} f_{\boldsymbol{l},\boldsymbol{j}} \phi_{\boldsymbol{l},\boldsymbol{j}}. \tag{5.1}$$

Durch die Wahl der Multiindexmenge $\mathcal{C} \subset \mathbb{N}^d$ bestimmt man die Beschaffenheit des Gitters. Für $\mathcal{C} = \{1 \leq |\boldsymbol{l}|_\infty \leq \ell\}$ ergibt sich etwa ein isotropes Produktgitter der Maschenweite $2^{-\ell}$, während $\mathcal{C} = \{1 \leq |\boldsymbol{l}|_1 \leq \ell\}$ ein reguläres dünnes Gitter zum Level ℓ liefert.

Die Indexmenge $\mathcal{C}$ lässt sich jedoch auch allgemeiner aufbauen – solange sie die Zulässigkeitsbedingung

$$\boldsymbol{l} \in \mathcal{C} \text{ und } \boldsymbol{k} \leq \boldsymbol{l} \Rightarrow \boldsymbol{k} \in \mathcal{C}$$

erfüllt.

Auf diese Weise lassen sich auch *dimensionsadaptive dünne Gitter* erzeugen [GG03, Hol08, Feu10], welche ohne a-priori Information zu einer Funktion zu besitzen, in der Lage sind deren relevante Dimensionen zu erkennen und entsprechend zu verfeinern.

In [Feu10] und [Boh10] werden ferner ortsadaptive Varianten der dünnen Gitter beschrieben, wodurch sich in niedrigen Dimensionen auch nicht-reguläre Funktionen effizient approximieren lassen.

5.1.2 Numerische Versuche

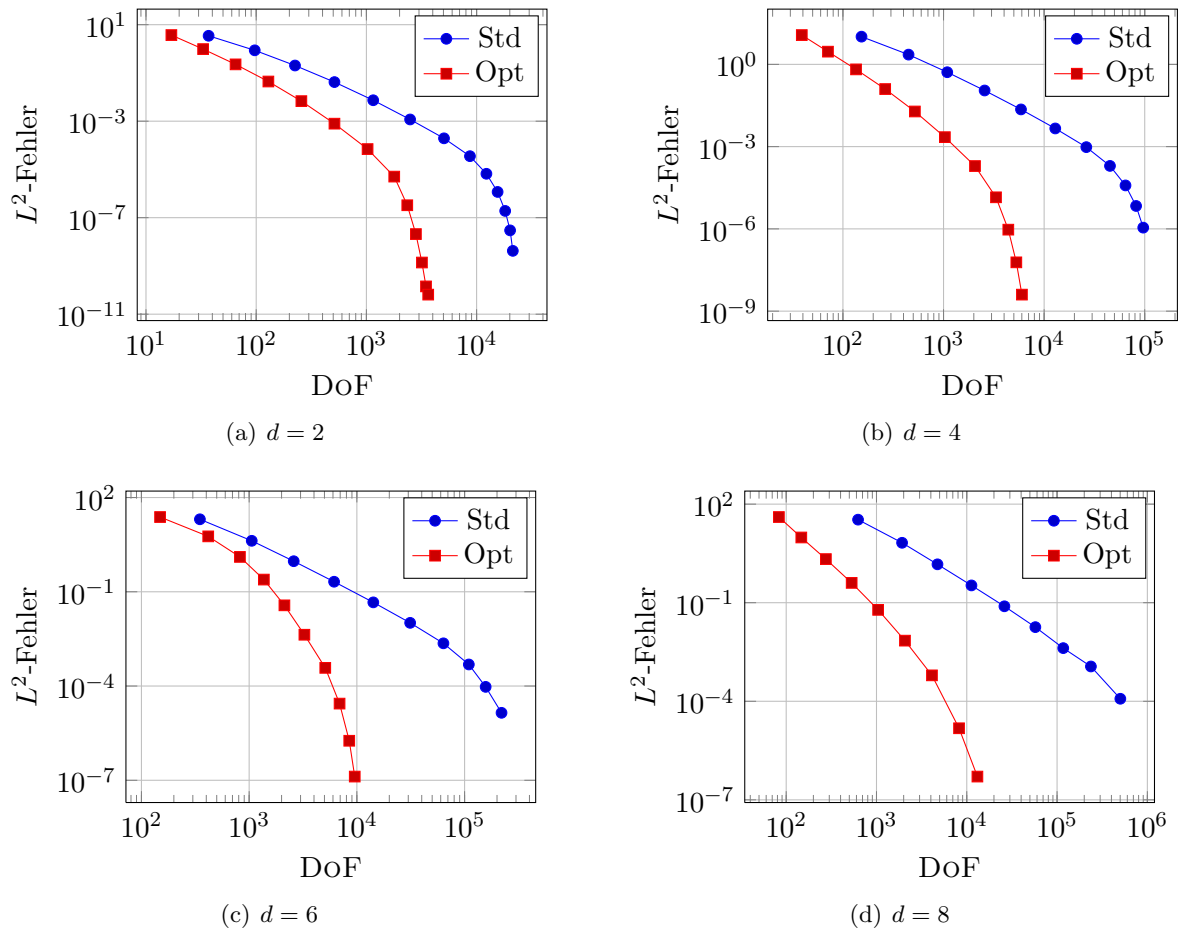
Wir interpolieren Funktionen $f : \mathbb{R}^d \to \mathbb{R}$ mit dem orts- und dimensionsadaptiven Dünngitter Algorithmus aus [Feu10], indem wir mit Verfahren II eine geeignete Drehung des Koordinatensystems suchen und dann die gedrehte Funktion

$$\hat{f}(\boldsymbol{x}) := f(\boldsymbol{Q}\boldsymbol{x})$$

interpolieren. Dazu transfomieren wir die Dünngitterbasis durch die inverse Normalverteilung Φ^{-1} auf den Ganzraum $\mathbb{R}^d$, wobei wir, um die sich dadurch ergebende Singularität zu umgehen, lediglich das Gebiet $\Phi^{-1}([\varepsilon, 1-\varepsilon]^d)$ mit $\varepsilon = 10^{-3}$ abbilden. Eine Möglichkeit zur Abschätzung des Fehlers, den ein solches Vorgehen beinhaltet findet sich in [GKS10].

Dünne Gitter in Standard- und diversen Ganzraumkoordinaten sind in den Abbildungen 3.1 und 3.5 aus Abschnitt 3.1 dargestellt.

Um nun die Güte zweier Interpolanten miteinander zu vergleichen, messen wir jeweils den approximierten (bzw. diskretisierten) $\mathcal{L}^2$-Abstand der interpolierenden Dünngitterfunktion f_h



(a) $d = 2$ (b) $d = 4$

(c) $d = 6$ (d) $d = 8$

Abb. 5.1: Konvergenzrate der Dünngitter-Approximation für das Polynom (5.3) in verschiedenen Dimensionen.

zu f an $N := 10^5$ Quasi-Monte Carlo Punkten der Sobol-Folge und vergleichen

$$\begin{aligned} E_{\text{std}}(f, f_h) &:= \frac{1}{N} \sum_{i=1}^{N} (f(\boldsymbol{x}^{(i)}) - f_h(\boldsymbol{x}^{(i)})^2 \quad \text{und} \\ E_{\text{opt}}(f, \hat{f}_h \circ \boldsymbol{Q}^t) &:= \frac{1}{N} \sum_{i=1}^{N} (f(\boldsymbol{x}^{(i)}) - \hat{f}_h(\boldsymbol{Q}^t \boldsymbol{x}^{(i)})^2 \end{aligned} \tag{5.2}$$

miteinander.

Im Folgenden betrachten wir wieder die bereits in Kapitel 4 eingeführten und diskutierten Modellfunktionen und untersuchen, inwiefern sich die Konvergenzraten bei deren Interpolation verbessern lassen.

Polynome

Wir betrachten ein Polynom der Form (4.2) für $n = 2$

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} \sum_{j=i+1}^{d} x_i \, x_j, \tag{5.3}$$

welches wir, wie oben beschrieben mit dem orts- und dimensionsadaptiven Dünngitter-Interpolationsverfahren aus [Feu10] approximieren.

In Abbildung 5.1 sind $E_{\text{std}}(f, f_h)$ und $E_{\text{opt}}(f, \hat{f}_h \circ \boldsymbol{Q}^t)$ gegen die Zahl der aufgewendeten Freiheitsgrade DoF für das Polynom (5.3) geplottet. Es ist eine exponentielle Konvergenz der dünnen Gitter, sowohl für das untransformierte Polynom, als auch für sein dimensionsoptimiertes Pendant zu erkennen. Mit steigender Dimension benötigen die dünnen Gitter jedoch eine immer größere Zahl von Freiheitsgraden, bis sie diese exponentielle Asymptotik erreichen. Die Approximationsgüte des gedrehten Polynoms wird zwar auch mit steigender Dimension etwas schlechter, allerdings steigen die Kosten um eine vorgegebene Genauigkeit zu erreichen nur noch linear in d. Dies ist auch zu erwarten, denn wie wir in 4.1 nachgewiesen haben, besitzt das entsprechend gedrehte Polynom nur noch Superpositionsdimension eins.

Ridge-Funktionen

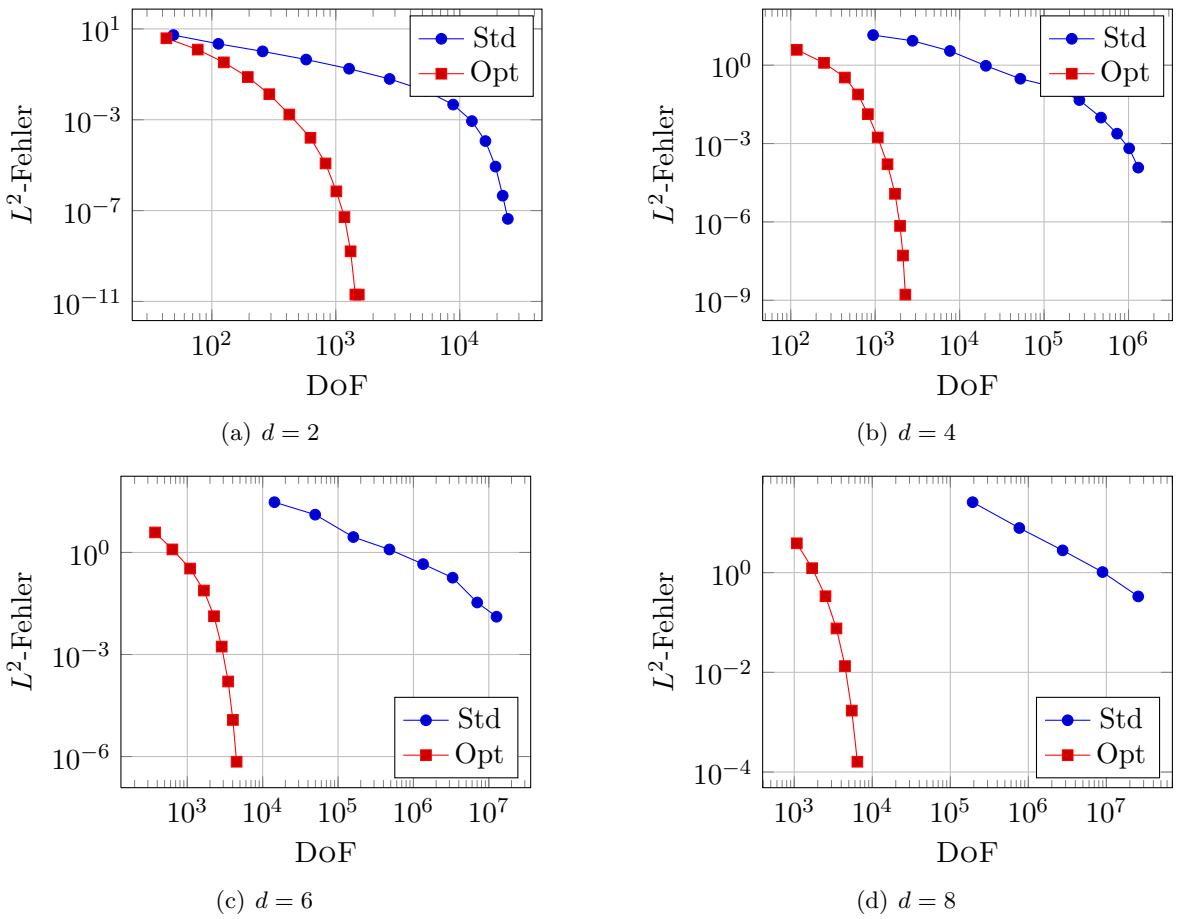
In diesem Abschnitt wollen wir untersuchen, inwiefern sich die Anwendbarkeit von dünnen Gittern auf die Interpolation von Ridge-Funktionen durch die Polynom-Methode (Verfahren II) verbessern lässt. Exemplarisch betrachten wir die bereits in Abschnitt 4.2 untersuchte Funktion

$$f(\boldsymbol{x}) = \exp\left(\frac{1}{\sqrt{d}} \sum_{i=1}^{d} x_i\right), \tag{5.4}$$

welche wir wieder mit der inversen Normalverteilung auf den Einheitswürfel transformieren und dort sowohl orts- als auch dimensionsadaptiv interpolieren.

In Abbildung 5.2 ist das Konvergenzverhalten für verschiedene Dimensionen d dargestellt. Für den zwei- und vierdimensionalen Fall ist auch hier die exponentielle Konvergenzrate noch gut zu erkennen. Für die nicht-optimierte Funktion („Std“) stellt sich diese Asymptotik in den Fällen $d = 6$ und $d = 8$ jedoch innerhalb der von uns betrachteten 10^7 Freiheitsgrade nicht mehr ein.

Richtet man (5.4) jedoch durch Verfahren II achsenparallel („Opt“) aus, so erhält man eine von der Dimension d beinahe unabhängige exponentielle Konvergenzrate. Dies ist nach den Erkenntnissen aus 4.2 auch zu erwarten, da Ridge-Funktionen – unabhängig von ihrer nominellen Dimension – durch unser Verfahren auf eine eindimensionale Funktion reduziert werden können.



(a) $d = 2$ (b) $d = 4$

(c) $d = 6$ (d) $d = 8$

Abb. 5.2: Konvergenzrate der Dünngitter-Approximation für die Ridge-Funktion (5.4) in verschiedenen Dimensionen.

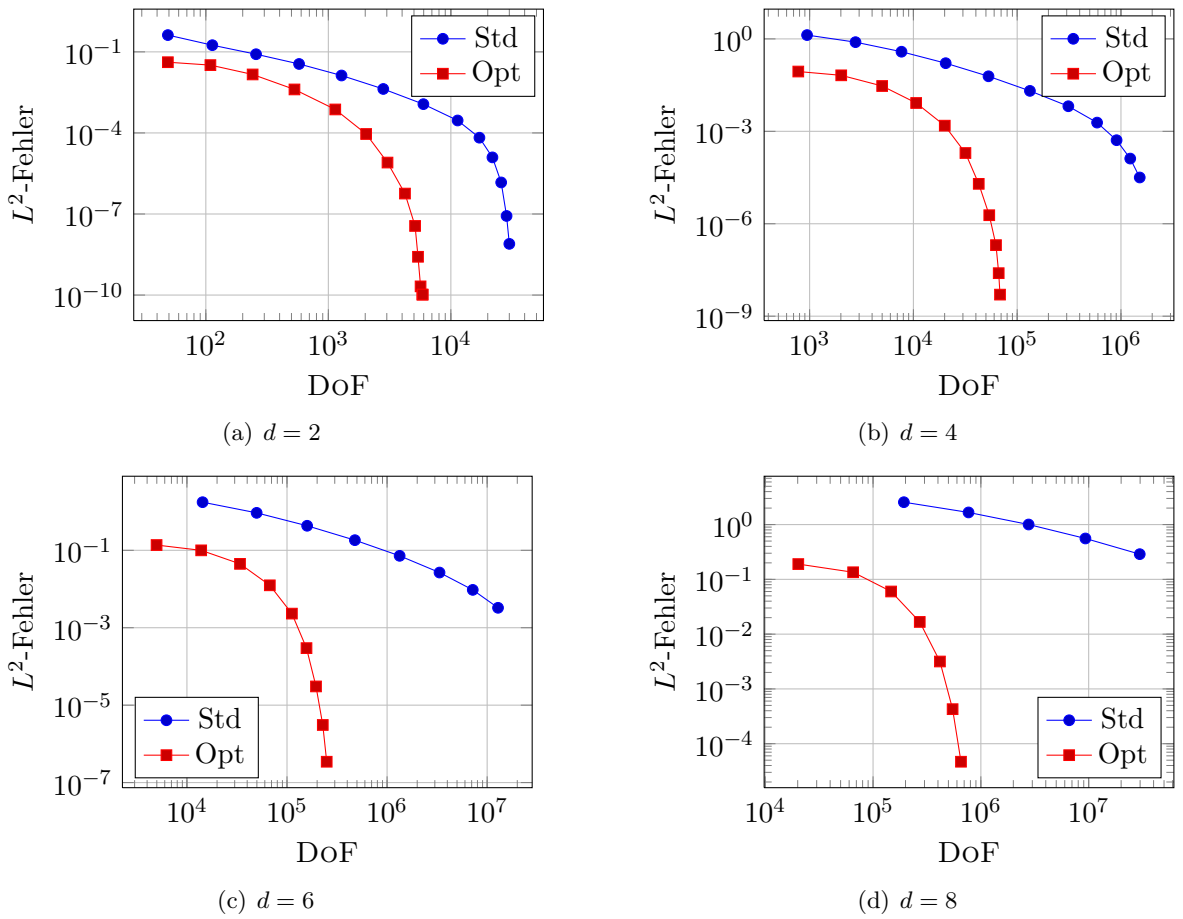
Verallgemeinerte Ridge Funktionen

An dieser Stelle betrachten wir wieder verallgemeinerte Ridge-Funktionen aus 4.2 als Beispiel für effektiv-hochdimensionale Funktionen, die sich durch eine nicht-triviale Drehung auf eine Funktion mit Superpositionsdimension eins reduzieren lassen. In Abbildung 5.4 sind die Konvergenzeigenschaften für stückweise-lineare, orts- und dimensionsadaptive dünne Gitter in verschiedenen Dimensionen dargestellt. Als Testproblem wurde die Funktion (4.9)

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} \sin(2\boldsymbol{w}_i^t \boldsymbol{x}) \tag{5.5}$$

verwendet.

Auch hier ist für den Fall („Std“) in zwei und vier Dimensionen die Asymptotik der Dünngitter-Konvergenz noch erkennbar – in sechs und acht Dimensionen kommt man ohne eine geeignete



(a) $d = 2$ (b) $d = 4$

(c) $d = 6$ (d) $d = 8$

Abb. 5.3: Verallgemeinerte Ridge-Funktionen (5.5) in $d = 8$ Dimensionen.

Drehung jedoch kaum noch mit einer vertretbaren Zahl von Freiheitsgraden in einen aktzeptablen Fehlerbereich.

Spiralförmige Funktion

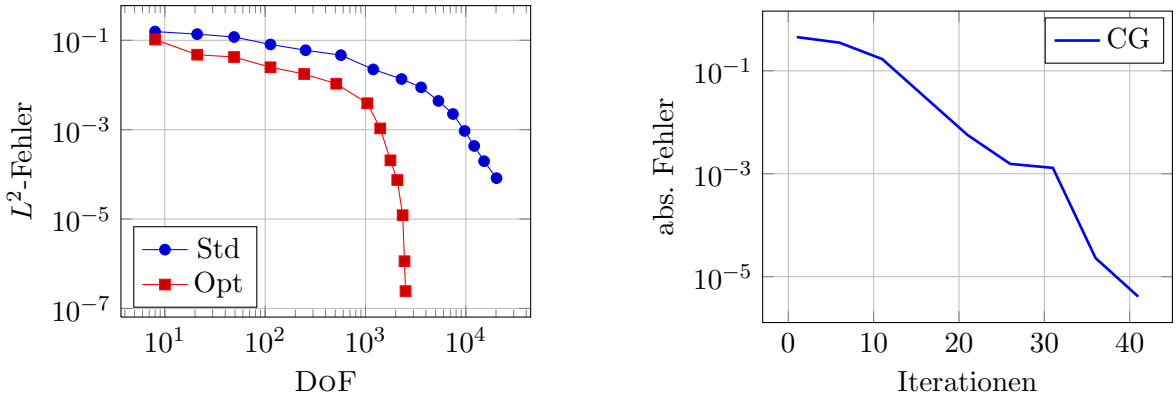
Auch die Transformation mit stückweiseweise-orthogonalen Abbildungen aus Abschnitt 3.1.3 mit dem in 3.8.1 beschriebenen Verfahren III wollen wir anhand der Interpolation mit einem dünnen Gitter untersuchen.

Dazu betrachten wir wieder die Funktion

$$f(x_1, x_2) = \exp\left(-\left(\cos(x^2 + y^2)(x + y) + \sin(x^2 + y^2)(x - y)\right)^2\right), \tag{5.6}$$

die in Abbildung 4.7(a) dargestellt ist und interpolieren sie mit dem üblichen Verfahren sowohl orts- als auch dimensionsadaptiv.

In Abbildung 5.4(a) ist der approximierte $\mathcal{L}^2$-Fehler sowohl für f, als auch für $f \circ \hat{\boldsymbol{Q}}$ zu sehen.



(a) Adaptive Interpolation

(b) CG-Konvergenz

Abb. 5.4: Links ist die verbesserte Konvergenz dünner Gittter durch die stückweise-orthogonale Transformation dargestellt. Rechts die Konvergenzgeschwindigkeit des CG-Verfahrens um diese Transformation zu erzeugen.

Fehlermessung in der Maximumsnorm

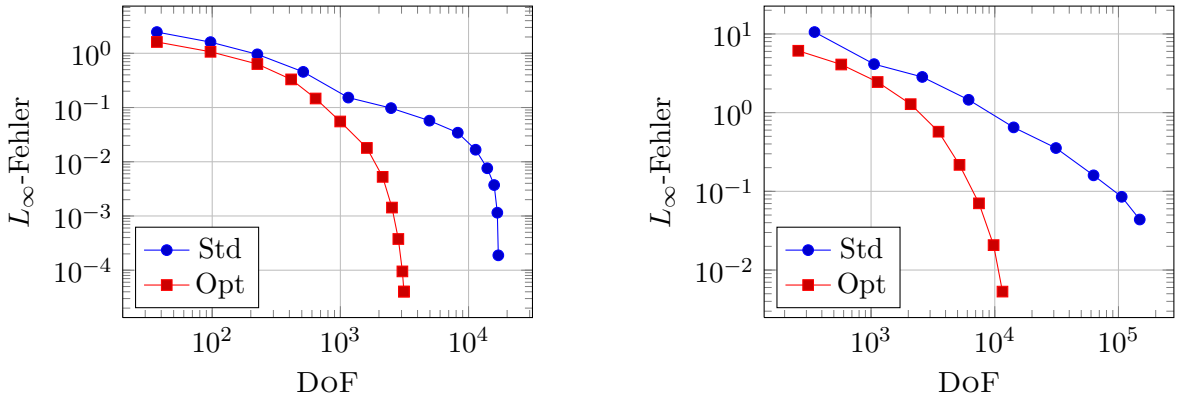
Bisher haben wir lediglich die $\mathcal{L}^2$-Norm zur Messung des Interpolationsfehler betrachtet. Um diesen Abschnitt abzurunden, betrachten wir im Folgenden noch den L_∞-Fehler der Polynom-Modellfunktion (5.3) und der verallgemeinerten Ridge-Funktion (5.5). Diesen messen wir wieder an 10^5 Quasi-Monte Carlo Punkten der Sobol Folge.

In Abbildung 5.5 ist der absolute L_∞-Fehler der orts- und dimensionsadaptiven Interpolation der Testfunktion (5.3) gegen die Zahl der benötigten Freiheitsgrade geplottet.

Genau wie in Abbildung 5.1 für den $\mathcal{L}^2$-Fehler wird auch der Fehler in der Maximumsnorm durch die Drehung des Koordinatensystems deutlich reduziert.

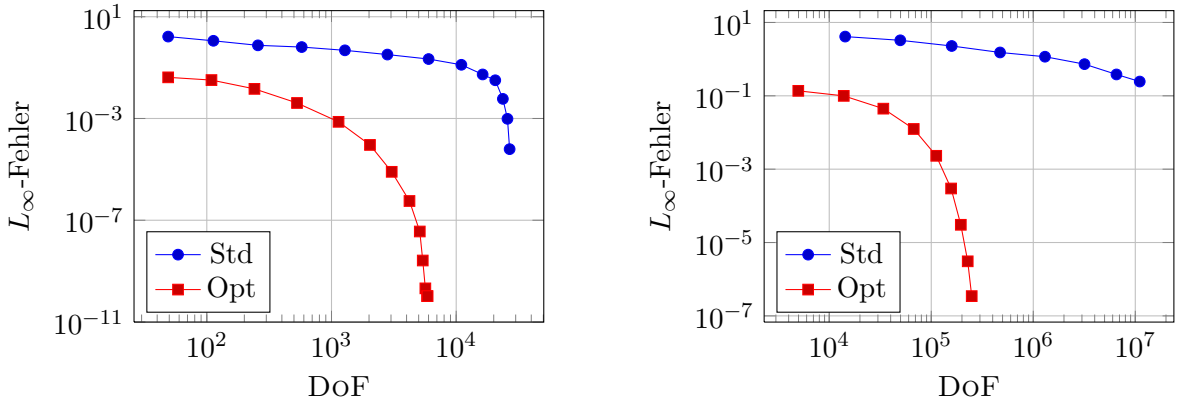
Das gleiche gilt für den den maximalen Interpolationsfehler bei der Testfunktion (5.5). In Abbildung 5.3 ist der $\mathcal{L}^2$-Fehler, in Abbildung 5.6 der Fehler in der Maximumsnorm dargestellt.

Abschließend können wir also bemerken, dass sich sowohl für die $\mathcal{L}^2$-, als auch für die Maximumsnorm für die von uns betrachteten Testprobleme durch die Reduktion der effektiven Dimension eine signifikante Verbesserung der Konvergenz dünner Gitter ergibt.



(a) $d = 2$ (b) $d = 6$

Abb. 5.5: Interpolationsfehler in der Maximumsnorm für die Testfunktion (5.3).



(a) $d = 2$ (b) $d = 6$

Abb. 5.6: Interpolationsfehler in der Maximumsnorm für die Testfunktion (5.5).

5.2 Hochdimensionale Integration

In diesem Abschnitt werden wir untersuchen, inwiefern sich die Konvergenzeigenschaften der Quasi-Monte Carlo und Dünngitter-Quadratur Verfahren durch eine geeignete Drehung des Koordinatensystems verbessern lassen. Denn während die Konvergenzrate von Monte Carlo Verfahren (MC) unabhängig vom Koordinatensystem nur von der Varianz der betrachteten Funktion abhängt, lässt sich für Quasi-Monte Carlo Methoden ([WF03, WS03, SWW04] und dünne Gitter ([Hol08]) zeigen, dass deren tatsächliche Konvergenzrate hochgradig von der effektiven Dimension des Integranden abhängig ist.

Inbesondere dimensionsadaptive dünne Gitter [GG03, Hol08] sind in der Lage eine niedrige effektive Dimension effizient auszunutzen, indem das Gitter nur entlang der relevanten Dimensionsrichtungen verfeinert wird.

Im Folgenden werden wir numerische Quadraturverfahren der Form

$$\int f \, d\mu \approx \sum_{i=1}^{N} w_i f(\boldsymbol{x}^{(i)}) \tag{5.7}$$

betrachten, wobei Quadraturgewichte w_i und eine Folge von Stützstellen $\boldsymbol{x}^{(i)}$ für $i = 1, \ldots, N$ gegeben seien. Diese definieren eine d-dimensionale Quadraturregel auf $[0,1]^d$ oder $\mathbb{R}^d$.

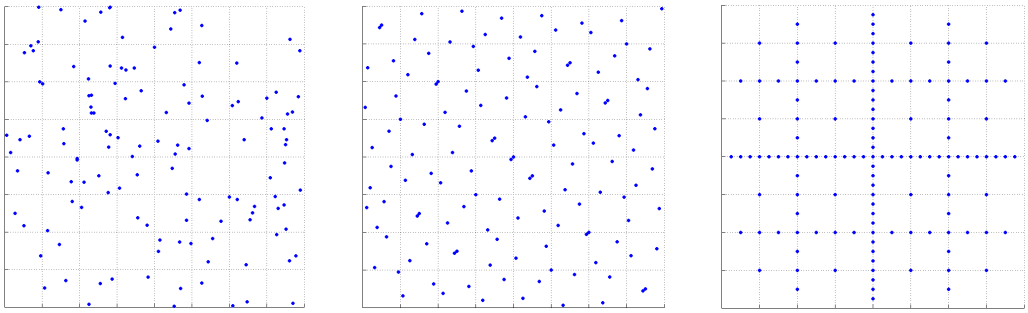
Die Quadratur mit Quasi-Monte Carlo Folgen und dünnen Gittern sind Spezialfälle eines solchen Verfahrens, welche wir in den beiden folgenden Unterabschnitten zusammenfassend beschreiben werden.

5.2.1 Quasi-Monte Carlo Methoden

Bei Quasi-Monte Carlo Verfahren (QMC) gilt, genau wie bei Monte Carlo Methoden (MC), für alle Gewichte $w_i = \frac{1}{N}$. Anstatt von Pseudo-Zufallszahlen wird jedoch eine deterministische Folge von Punkten verwendet - sogenannte Nieder-Diskrepanz Folgen. Diese besitzen für kleine und mittlere Dimensionalitäten bessere Verteilungseigenschaften als MC und daher für Funktionen mit beschränkter Variation auch eine bessere Konvergenzrate von $\mathcal{O}(\frac{\log(N)^d}{N})$ anstatt $\mathcal{O}(\frac{1}{\sqrt{N}})$.

Ein zweidimensionales Beispiel mit je 128 MC- und QMC-Punkten ist in der aus [Hol08] entnommenen Abbildung 5.7 dargestellt. Die „gleichmäßigere“ Verteilung der QMC-Punkte ist deutlich zu erkennen. Um die Qualität dieser „Gleichmäßigkeit“ in der Verteilung zu messen, wird der Begriff der *Diskrepanz* verwendet. Hiervon existiert in der Literatur ein Vielzahl von Varianten, etwa in [Hic98, NW01, NW07]. Ein Überblick findet sich im Anhang von [Hol08].

Häufig gebrauchte QMC-Verfahren sind die Halton-, Faure- und Sobol- Folgen, deren Anwendung auf Probleme aus dem Bereich der Finanzmathematik in [Gla04, Wat07] näher erläutert wird.



(a) MC (b) Quasi-MC (c) DG

Abb. 5.7: Gitterpunkte der Monte Carlo, Quasi-Monte Carlo (Sobol Folge) und Dünngitter (Trapezregel) Methoden in zwei Dimensionen

5.2.2 Dünngitter Integration

Dünne Gitter (eng. *sparse grids*) können für beliebige Produktmengen $\Omega^{(d)} = \bigotimes \Omega_i$ definiert werden. Wir betrachten hier den Fall $\Omega_i = [0,1]$ und $\Omega_i = \mathbb{R}$.

Ausgehend von einer Folge von eindimensionalen Quadraturregeln auf Ω_i wird nach dem Tensor-Prinzip eine d-dimensionale Quadraturformel konstruiert.

Für eine monoton-steigende Folge natürlicher Zahlen m_k, $k \in \mathbb{N}$, definieren wir für eindimensionale Funktionen $f : [0,1] \to \mathbb{R}$ die Folge univariater Quadraturformeln

$$U_{m_k} f := \sum_{i=1}^{m_k} w_{i,k}\, f(x_{i,k}). \tag{5.8}$$

mit m_k Punkten $x_{i,k}$ und Gewichten $w_{i,k}$, so dass

$$\lim_{k\to\infty} U_{m_k} f = \int f \, d\mu$$

gilt.

Beispiele für eine solche Regel U_{m_k} sind die Trapezregel und die Gauß-Quadratur (etwa mit Legendre-Polynomen) auf $[0,1]$, die Gauß-Laguerre Quadratur für $[0,\infty)$ und die Gauß-Hermite Quadratur auf $\mathbb{R}$.

Mit $m_1 = 1$ definieren wir für $k \geq 1$ die Differenzen-Formeln

$$\Delta_k := U_{m_k} - U_{m_{k-1}} \quad \text{mit } U_{m_0} := 0. \tag{5.9}$$

Für einen d-dimensionalen Multiindex $\mathbf{k} \in \mathbb{N}^d$ mit $k_j > 0$ für alle $j \in \mathcal{D}$ sei

$$\boldsymbol{\Delta}_{\mathbf{k}} f := (\Delta_{k_1} \otimes \ldots \otimes \Delta_{k_d})\, f. \tag{5.10}$$

Damit können wir nun das Integral multivariater Funktionen $f : [0,1]^d \to \mathbb{R}$ durch eine

Teleskop-Summe

$$\int f \, d\mu = \sum_{\mathbf{k} \in \mathbb{N}^d} \boldsymbol{\Delta}_{\mathbf{k}} f, \tag{5.11}$$

welche alle Kombinationen von Produkten der eindimensionalen Differenzen-Formeln durchläuft, approximieren.

Die klassischen dünnen Gitter nach Smolyak (siehe [Smo63, GG98]) zu einem gegebenen Level $\ell \in \mathbb{N}$ sind nun durch

$$\mathrm{SG}_\ell f := \sum_{|\mathbf{k}|_1 \leq \ell + d - 1} \boldsymbol{\Delta}_{\mathbf{k}} f, \tag{5.12}$$

gegeben, wobei $|\mathbf{k}|_1 := \sum_{j=1}^{d} k_j$ die 1-Norm des Multiindex $\mathbf{k}$ bezeiche.

Verwendet man in (5.12) statt dessen die Maximumsnorm $|\boldsymbol{k}|_\infty := \max\{k_1, \ldots, k_d\}$, so erhält man ein normales Produktverfahren.

Lässt man anstelle der Indexmenge $\{\mathbf{k} : |\mathbf{k}|_1 \leq \ell + d - 1\}$ beliebige $\mathcal{C}$ zu, welche die Bedingung

$$\mathbf{k} \in \mathcal{C} \text{ und } \mathbf{l} \leq \mathbf{k} \Rightarrow \mathbf{l} \in \mathcal{C}$$

erfüllen, so kann man durch eine geeignete adaptive Konstruktion von $\mathcal{C}$ Gitterstrukturen erzeugen, die sich der dimensionalen Struktur des Integranden anpassen. Dieses Konzept der *dimensionsadaptiven Dünngitterquadratur* stammt aus [GG03] und wird ausführlich in [GH10b] und [Hol08] diskutiert.

Durch eine geeignete Wahl der eindimensionalen Quadraturregeln sind die Dünngitter-Methoden in der Lage, die Glattheit des Integranden auszunutzen und auf diese Weise hohe Konvergenzraten zu erzielen. Durch die „dünne“ Struktur ihrer Gitter sind sie in der Lage, den Fluch der Dimension zu brechen.

Die Funktionenklasse, für die das möglich ist, sind gerade die Funktionen mit beschränkten gemischten Ableitungen vom Grad s, also

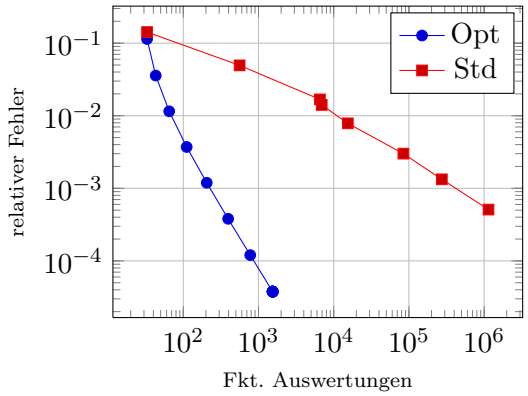
$$H^s([0,1]^d) := \left\{ f : [0,1]^d \to \mathbb{R} : \max_{|\mathbf{k}|_\infty \leq s} \left\| \frac{\partial^{|\mathbf{k}|_1} f}{\partial {x_1}^{k_1} \ldots \partial {x_d}^{k_d}} \right\|_\infty < \infty \right\}.$$

Der Fehler der klassischen Dünngitter Quadratur kann für alle $f \in H^s([0,1]^d)$ durch

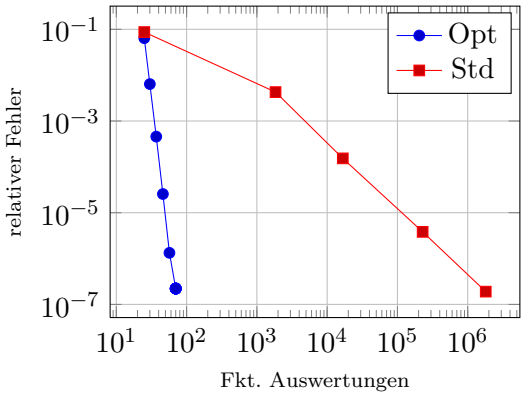
$$\varepsilon(n) = O(n^{-s} (\log n)^{(d-1)(s-1)}), \tag{5.13}$$

abgeschätzt werden, d.h. die Konvergenzrate ist bis auf einen logarithmischen Faktor von der Dimension d aunabhängig.

Basierend auf den eindimensionalen Quadraturregeln, lassen sich viele verschiedene Typen von Dünngitter Quadraturformeln konstruieren. *Clenshaw-Curtis*-Folgen werden in [NR96], *Gauß-Patterson* und *Gauß-Legendre* in [GG98], Gauß-Integration mit den *Gauß-Hermite* and *Genz-Keister* Regeln in [Nah05, NRS98] und [HW08] eingeführt. In dieser Arbeit werden wir uns



(a) Gauß-Legendre



(b) Gauß-Hermite

Abb. 5.8: Ridge-Funktion $\exp(\boldsymbol{w}^t\boldsymbol{x})$, $d = 8$. „Std“ bezeichnet die untransformierte Funktion, „Opt“ die optimierte Variante.

jedoch auf die durch die Gauß-Legendre und Clenshaw-Curtis Regeln definierten Gitter im Einheitswürfel $[0,1]^d$, und auf die Gauß-Hermite dünnen Gitter im Ganzraum $\mathbb{R}^d$ einschränken.

5.2.3 Numerische Versuche

Um zu zeigen, dass die von uns entwickelten Verfahren den Fehler, bzw. die Kosten für hochdimensionale Quadraturprobleme substanziell verringern können, betrachten wir Funktionen $f : \mathbb{R}^d \to \mathbb{R}$, die wir bezüglich des Gauß-Maßes integrieren.

Solche Integrale treten bei der Berechnung des Erwartungswertes von Funktionen, welche auf Gauß-Prozessen definiert sind, auf – etwa beim Option-Pricing nach dem Black-Scholes Modell, aber auch in vielen Bereichen der Physik, wo das Verhalten von Teilchen durch normalverteilte Zufallsvariablen modelliert wird.

Im Wesentlichen wollen wir in diesem Abschnitt Funktionen betrachten, die wir bereits in Kapitel 4 hinsichtlich ihrer effektiven Dimension untersucht haben.

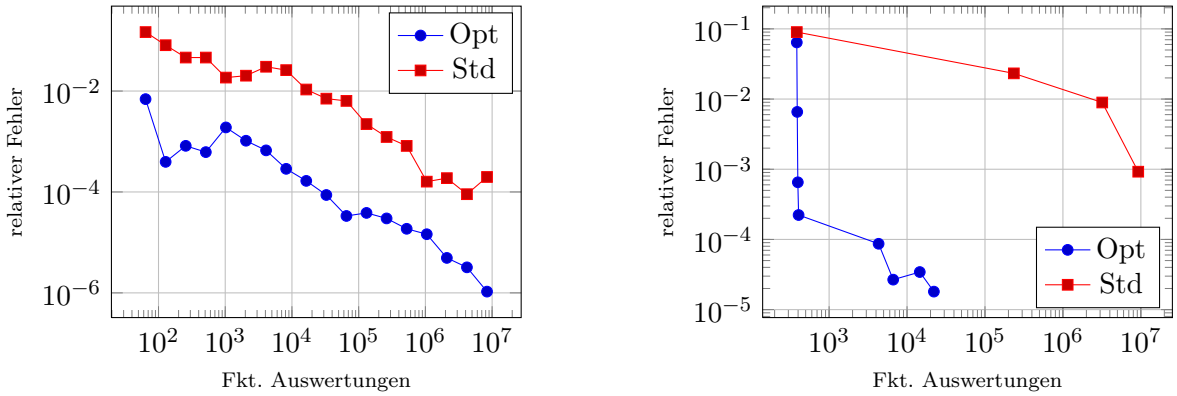
Ridge-Funktionen

Als Beispiel für eine Ridge-Funktion betrachten wir wieder die in Abschnitt 4.2 diskutierte Funktion

$$f(\boldsymbol{x}) = \exp(\boldsymbol{w}^t\boldsymbol{x}), \quad \boldsymbol{w} = \frac{1}{\sqrt{d}}(1,\dots,1)^t,$$

die wir bezüglich des Gauß-Maßes über dem Ganzraum $\mathbb{R}^d$ integrieren wollen.

In Abbildung 5.8 ist zu sehen, dass sowohl die Gauß-Legendre, als auch die Hermite dünnen Gitter substanziell von der niedrigen effektiven Dimension des gedrehten Integranden profitieren.



(a) Qmc (Sobol)

(b) Gauß-Hermite

Abb. 5.9: Ridge-Funktion $\exp(\boldsymbol{w}^t\boldsymbol{x})$ in $d = 128$ Dimensionen.

Doch obwohl beide über dieselbe polynomielle Exaktheit von $2n - 1$ verfügen (siehe [Hol08]), schneidet die Hermite-Quadratur deutlich besser ab. Dies ist darauf zurückzuführen, dass bei der singulären Transformation auf den Einheitswürfel diese polynomielle Exaktheit verloren geht. So ist etwa eine lineare Funktion auf $\mathbb{R}$ nach Transformation mit der inversen Normalverteilung als Funktion auf $(0, 1)$ nicht mehr linear.

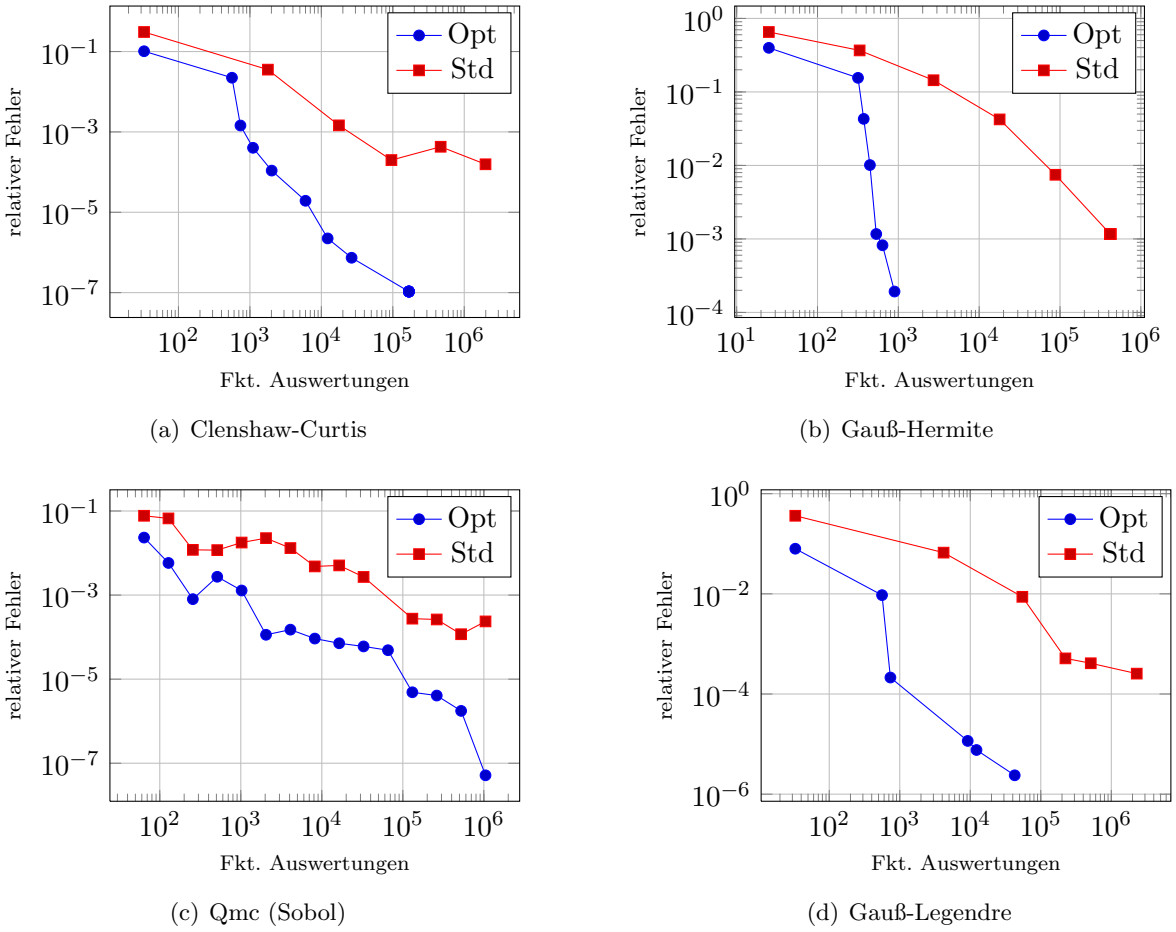
In Abbildung 5.9 ist zu sehen, dass auch Quasi-Monte Carlo Verfahren deutlich von der reduzierten Trunkationsdimension profitieren. Interessant ist an dieser Stelle auch das Verhalten der Hermite-DG, die beim 128-dimensionalen Problem in Standardkoordinaten erst nach ca. 10^6 Funktionsauswertungen in den Bereich kommen, indem ihr asymptotisches Konvergenzverhalten sichtbar wird. Für den transformierten Integranden flacht die Konvergenz bei einem relativen Fehler von 10^{-4} jedoch deutlich ab. Dies ist darauf zurück zu führen, dass Verfahren II die optimale Drehung natürlich nur nährungsweise bestimmt, was in hohen Dimensionen dazu führen kann, dass sich zwar ein Anteil von 0.9999 der Varianz in der ersten Dimension befindet, der „Rest“ der Funktion jedoch noch immer hochdimensional ist. Wenn man von der effektiven Dimension eines Problems spricht, sollte dies also immer im Kontext der gewünschten Genauigkeit der Lösung dieses Problems geschehen. Eine Aussage wie „die effektive Trunkationsdimension zum Paramter $\alpha = 0.9999$ ist 1“, sollte also kritisch gesehen werden, falls die Funktion mit einer höheren Genauigkeit als α integriert werden muss.

Verallgemeinerte Ridge-Funktionen

Als Beispiel für verallgemeinerte Ridge-Funktionen (siehe 4.2) betrachten wieder die Funktion

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} \sin(2\boldsymbol{w}_i^t\boldsymbol{x}),$$

welche wir ebenfalls bezüglich des Gauß-Maßes auf dem $\mathbb{R}^d$ integrieren wollen.



(a) Clenshaw-Curtis

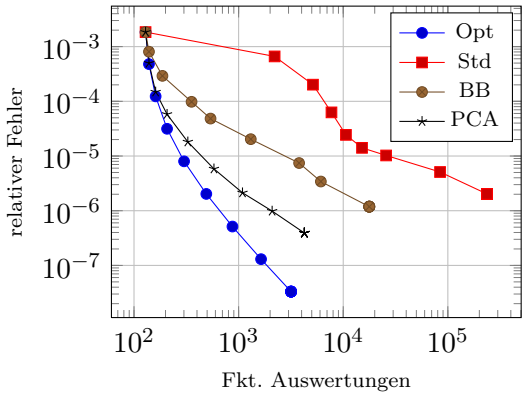
(b) Gauß-Hermite

(c) Qmc (Sobol)

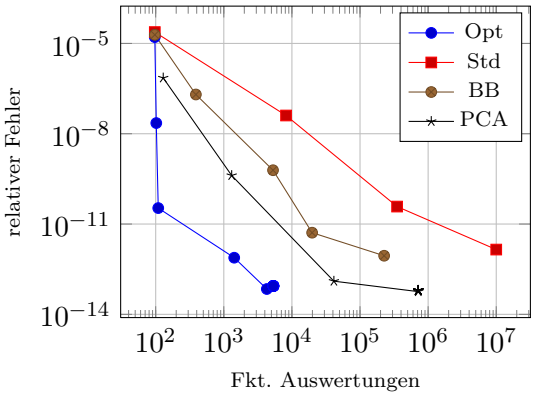
(d) Gauß-Legendre

Abb. 5.10: Verallgemeinerte Ridge-Funktionen mit $g_i(x) = \sin(2x)$, in $d = 8$ Dimensionen.

In Abbildung 5.10 sind die Konvergenzeigenschaften der Clenshaw-Curtis, Gauß-Legendre, Gauß-Hermite und Quasi-Montecarlo Methoden dargestellt. Alle vier betrachteten Verfahren profitieren deutlich von der niedrigen effektiven Dimension des Integranden, welcher nach der Transformation – wie wir in Abschnitt 4.2 dargelegt haben – eine effektive Superpositionsdimension von 1 besitzt.



(a) Clenshaw-Curtis



(b) Gauß-Hermite

Abb. 5.11: Geometrische asiatische Option, $K = 0$, $d = 32$.

Asiatische Optionen

Als praktische Anwendung betrachten wir wieder asiatische Optionen nach dem Black-Scholes Modell, dessen Paramter wir wie in Abschnitt 4.3 wählen.

Für einen Ausübungspreis $K = 0$ ergibt sich ein überrall differenzierbarer Integrand, wodurch die Anwendung von dünnen Gittern sinnvoll wird. In Abbildung 5.11 sind die Konvergenzeigenschaften der Clenshaw-Curtis und die Hermite-Dünnen Gitter zu verschiedenen Pfadkonstruktionsmethoden dargestellt.

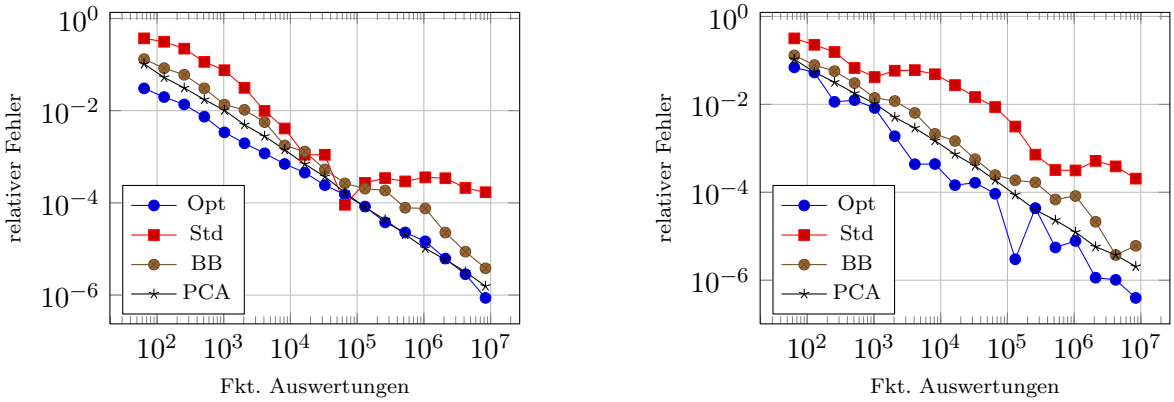
Deutlich ist die generelle Überlegenheit der Ganzraum-Integration gegenüber der Transformation auf den Einheitswürfel erkennbar. Der relative Quadraturfehler der Hermite-Integration ist selbst für den untransformierten Integranden (Random Walk) um den Faktor 10^5 kleiner als bei Clenshaw-Curtis.

Ebenfalls sehr deutlich zu erkennen ist die Relevanz einer Transformation des Integranden. So bringen im Falle der Hermite-Dünnen Gitter die Brownsche-Brücke Konstruktion eine 10^3 mal höhere Genauigkeit, die PCA sogar einen Faktor 10^4.

Den größten Effekt zeigt jedoch die Polynommethode, welche eine Transformation auf eine „fast“ eindimensionale Funktion ergibt (in diesem Fall ist sie zur LT äquivalent) und daher mit lediglich 142 Quadraturpunkten eine relative Genauigkeit von $6.3 \cdot 10^{-10}$ erreicht. Für eine größere Genauigkeit flacht die Kurve jedoch wieder ab, was wir hier auch auf den im Abschnitt über Ridge-Funktionen beschriebenen Effekt zurückführen.

In Abbildung 5.12 betrachten wir dann eine arithmetische Option mit Ausübungspreis $K = 120$. Durch den Knick können dünne Gitter hier nicht mehr praktikabel eingesetzt werden, weswegen wir lediglich die Quasi-Monte Carlo Quadratur mit dünnen Gittern untersuchen.

Für den Fall $d = 32$ ist zwar zu erkennen, dass eine gute Pfadkonstruktion die Integration beschleunigt, ein deutlicher Untschied zwischen diesen Methoden ist jedoch nicht erkennbar.



(a) Qmc, $d = 32$

(b) Qmc, $d = 128$

Abb. 5.12: Arithmetische asiatische Option, $K = 120$, Quasi-Monte Carlo für $d = 32$ und $d = 128$.

Auch in $d = 128$ Dimensionen ist der Unterschied keineswegs so deutlich wie bei den dünnen Gittern, aber immerhin ist zu erkennen, dass unsere Polynommethode ($n = 1$) etwas besser ist als die PCA, welche wiederum die Brownsche Brücke knapp übertrifft.

Dieser Effekt ist auf die in Abschnitt 2.2.3 beschriebene ANOVA-Zerlegung des Quasi-Monte Carlo-Fehlers zurückzuführen, welche effektive Dimension und Diskrepanz zueinander in Beziehung setzt. Da die Uniformität der Quasi-Monte Carlo Folgen in höheren Dimension abnimmt, wirken sich große Variationsbeiträge der hochdimensionalen ANOVA-Terme negativ auf die Rate von QMC aus, die daher für Funktionen ohne geringe effektive Dimension deutlich unter das theoretische Optimum N^{-1} fallen kann.

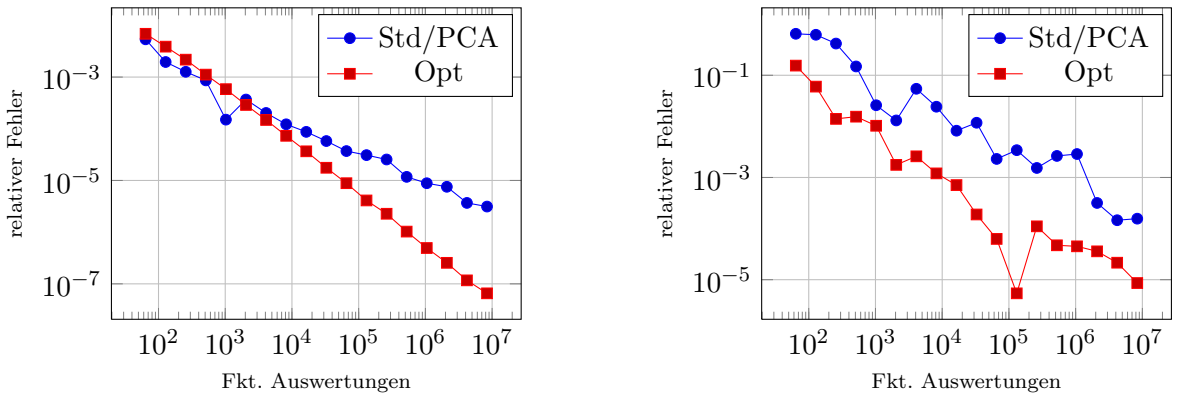
Basket-Optionen

Auch bei der Integration der in Abschnitt 4.3 eingeführten Basket-Optionen profitieren Quasi-Monte Carlo Methoden von der reduzierten effektiven Dimension.

In Abbildung 5.13 sind die QMC-Konvergenzraten für unkorrelierte, 16-dimensionale Basket-Optionen zu den Ausübungspreisen $K = 105$ und $K = 110$ dargestellt.

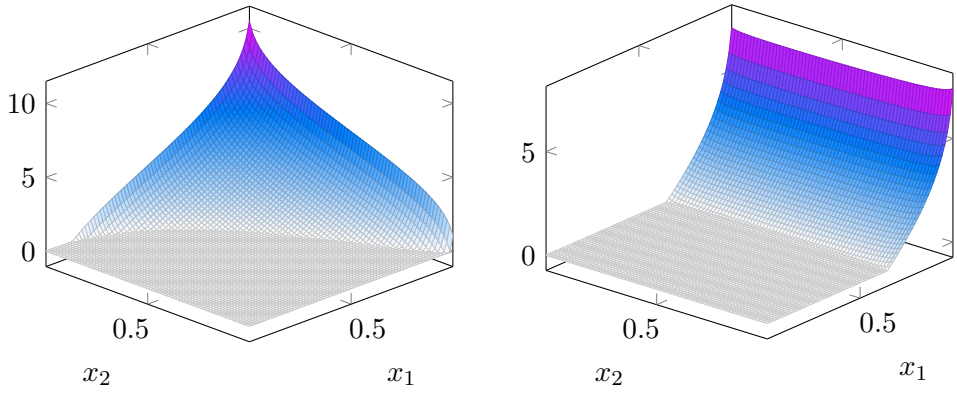
In beiden Fällen ist die PCA äquivalent zum Standard-Integranden, da die Kovarianzmatrix bereits Diagonalstruktur besitzt (siehe Abbildung 5.14).

Der Fall $K = 110$ ist wieder ein Beispiel für ein Problem, bei dem die LT und DM nicht ohne Weiteres angewendet werden können, da am Entwicklungspunkt die Differentiale erster und zweiter Ordnung identisch Null sind.



(a) $K = 105$ (b) $K = 110$

Abb. 5.13: Quasi-Monte Carlo (Sobol) für unkorrelierte euopäische Basket Optionen zu verschienden Ausübungspreisen K in $d = 16$ Dimensionen.



(a) PCA (b) Opt (Poly1)

Abb. 5.14: Unkorrelierte Basket-Option zum Ausübungspreis $K = 110$ – im Gegensatz zur Polynom-Methode kann die PCA den Integranden nicht achsenparallel ausrichten.

5.3 Regression mit Dünnen Gittern

In diesem Abschnitt werden wir nachweisen, die von uns entwickelten Verfahren zur Reduktion der effektiven Dimension auch im Bereich des maschinellen Lernens, also der Rekonstruktion eines funktionalen Zusammenhangs zwischen vorgegebenen Datenpunkten $\boldsymbol{x}$ und einer Antwortvariablen y, eine Anwendung finden können.

Dazu betrachten wir das auf dem Dünngitter-Konzept basierende multivariate Regressionsverfahren aus [Gar04], für das wir exemplarisch anhand von zwei Modell-Datensätzen die Anwendbarkeit der Dimensionsreduktionsmethoden darlegen werden.

Dessen grundsätzliche Ideen werden wir im Folgenden noch einmal kurz zusammenfassen – weitere Informationen zur Theorie und Hintergründen finden sich in [Gar04, Boh10].

5.3.1 Multivariate Regression

Im d-dimensionalen Merkmalsraum seien M Datenpunkte $\boldsymbol{x}^{(i)}$ und zugehörige Werte einer Antwortvariablen $y^{(i)}$ gegeben

$$S := \{(\boldsymbol{x}^{(i)}, y^{(i)}) \in \mathbb{R}^d \times \mathbb{R}\}_{i=1}^{M}$$

Gesucht ist eine Abbildung f_h, die den funktionalen Zusammenhang zwischen den Attributen und der Antwortvariablen möglichst gut vorhgersagen soll

$$y \approx f_h(\boldsymbol{x}).$$

Maschinelle Lernverfahren bestimmen innerhalb der Parameter eines vorgegebenen Funktionenraumes V die beste Approximation an die Daten, in der ℓ_2-Norm ergibt sich also das Problem

$$\arg\min_{f \in V} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (y^{(i)} - f_h(\boldsymbol{x}^{(i)}))^2}.$$

Derartige Approximationsprobleme sind jedoch im Allgemeinen *schlecht gestellt*, d.h. sie besitzen keine eindeutige Lösung – insbesondere dann, wenn die Zahl der Datenpunkte im Vergleich zur Zahl der Basisfunktionen gering ist, wodurch die Gefahr des *Overfitting* – also einer übermäßigen Anpassung an die Trainingsdaten, besteht.

Daher wird zur Beurteilung einer geeigneten Approximierenden f_h nicht nur der ℓ_2-Fehler, sondern auch den Wert eines sogeannten *Regularisierungstermes* herangezogen was auf einProblem der Form

$$\arg\min_{f \in V} \sqrt{\frac{1}{M} \sum_{i=1}^{M} (y^{(i)} - f(\boldsymbol{x}^{(i)}))^2} + \lambda \|\mathcal{A}(f)\|_{\mathcal{L}^2} \tag{5.14}$$

führt, wobei wir mit $\mathcal{A}$ einen beschränkten, linearen Operator auf V und mit λ einen Parameter zur Gewichtung dieses Regularisierungsfunktionales einführen.

Ist $\mathcal{B}$ eine abzählbare Basis von V, so kann man jede Funktion $f_h \in V$ in dieser Basis darstellen

$$f_(\boldsymbol{x}) = \sum_{\psi_j \in \mathcal{B}} \kappa_j \psi_j(\boldsymbol{x}). \tag{5.15}$$

Setzt man diese Darstellung in (5.14) ein, so führt dies auf das quadratische Problem

$$\arg\min_{f_h \in V} \frac{1}{M} \sum_{i=1}^{M} \left(y^{(i)} - \sum_{\psi_j \in \mathcal{B}} \kappa_j \psi_j(\boldsymbol{x}^{(i)}) \right)^2 + \lambda \sum_{\psi_k \in \mathcal{B}} \sum_{\psi_j \in \mathcal{B}} \kappa_j \kappa_k \, (\mathcal{A}\psi_j, \mathcal{A}\psi_k)_{\mathcal{L}^2}. \tag{5.16}$$

Durch Differentiation ist die Minimierung von (5.16) nun äquivalent zur Lösung eines linearen Gleichungssystems:

Für alle Elemente ψ_j einer N-elementigen Basis $\mathcal{B}$ definieren wir die Datenmatrix $\boldsymbol{B} \in \mathbb{R}^{M \times N}$ durch

$$\boldsymbol{B}_{i,j} = \psi_j(\boldsymbol{x}^{(i)}), \quad i = 1, \ldots, M.$$

Die positiv semi-definite Regularisierungsmatrix $\boldsymbol{C}$ definieren wir als

$$\boldsymbol{C} = (\mathcal{A}\psi_j, \mathcal{A}\psi_k)_{\mathcal{L}^2}, \quad i, j = 1, \ldots, N.$$

Damit erhalten wir aus (5.16) das lineare System

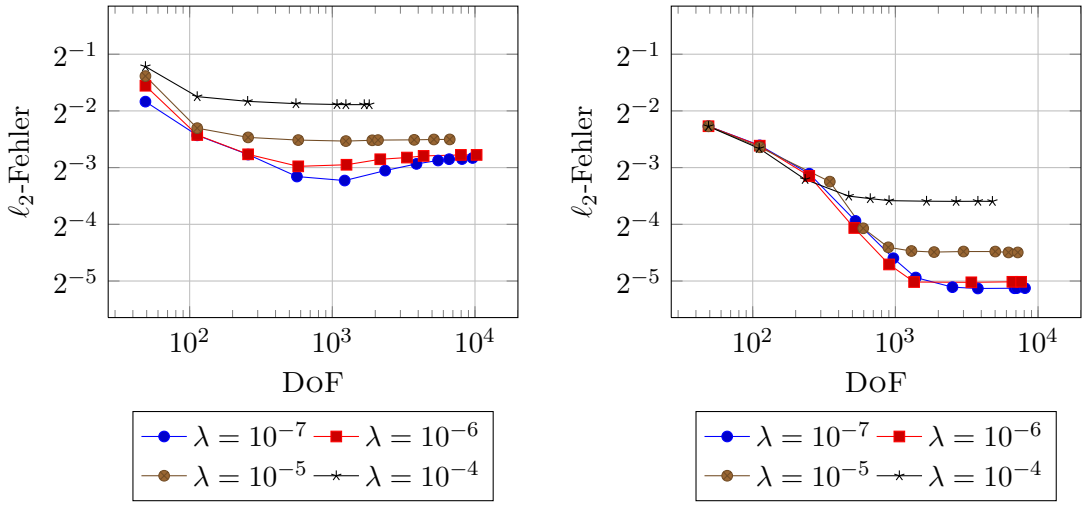
$$(\boldsymbol{B}^t \boldsymbol{B} + \lambda M \cdot \boldsymbol{C})\, \boldsymbol{\kappa} = \boldsymbol{B}^t \boldsymbol{y}, \tag{5.17}$$

Wählt man als Basiselemente ψ_j die Dünngitterbasis aus Abschnitt 5.1.1, so entspricht die Lösung von (5.17) also den Koeffizienten einer Dünngitterfunktion, welche die Daten aus S „möglichst gut“ approximiert. (Weiterführende Informationen zu Hintergründen und Theorie des oben beschriebenen Verfahrens finden sich in [Gar04, Boh10].

5.3.2 Numerische Versuche

Um Verfahren II auf die Regression mit dünnen Gittern anzuwenden werden wir folgendermaßen vorgehen:

1. Verwende Algorithmus 7 um die Daten $(\boldsymbol{x}^{(i)}, y^{(i)})_{i=1}^{M}$ mit einem Polynom P_n zu approximieren.
2. Verwende Verfahren I(a oder b) um die optimale Drehung $\boldsymbol{Q}$ von P_n zu berechnen.
3. Erzeuge daraus einen neuen Datensatz $(\boldsymbol{Q}\boldsymbol{x}^{(i)}, y^{(i)})_{i=1}^{M}$.
4. Approximiere diesen Datensatz mit dem in [Gar04, Boh10] beschriebenen Dünngitter-Verfahren.



(a) $d = 2$, Std

(b) $d = 2$, Opt

Abb. 5.15: Ridge-Datensatz (5.18) in $d = 2$ Dimensionen.

Wir betrachten die bereits im Abschnitt 4.2 untersuchten Funktionen

$$f_1(\boldsymbol{x}) = \sin(2 \cdot \boldsymbol{w}^t \boldsymbol{x}) \tag{5.18}$$

und

$$f_2(\boldsymbol{x}) = \sum_{i=1}^{d} \sin(2 \cdot \boldsymbol{w}_i^t \boldsymbol{x}), \tag{5.19}$$

von denen wir jeweils $M = 9000$ normalverteilte Punkte $\boldsymbol{x}^{(i)} \in \mathbb{R}^d$ und die zugehörige Antwortvariable $f(\boldsymbol{x}^{(i)}) =: y^{(i)}$ als Trainigsdatensatz samplen um daraus die Funktion selbst zu rekonstruieren.

Dazu verwenden wir das orts- und dimensionsadaptive Verfahren aus [Boh10], welches eine Dünngitterapproximation f liefert. Als Regularisierungsterm verwenden wir die H^1_{mix}-Norm.

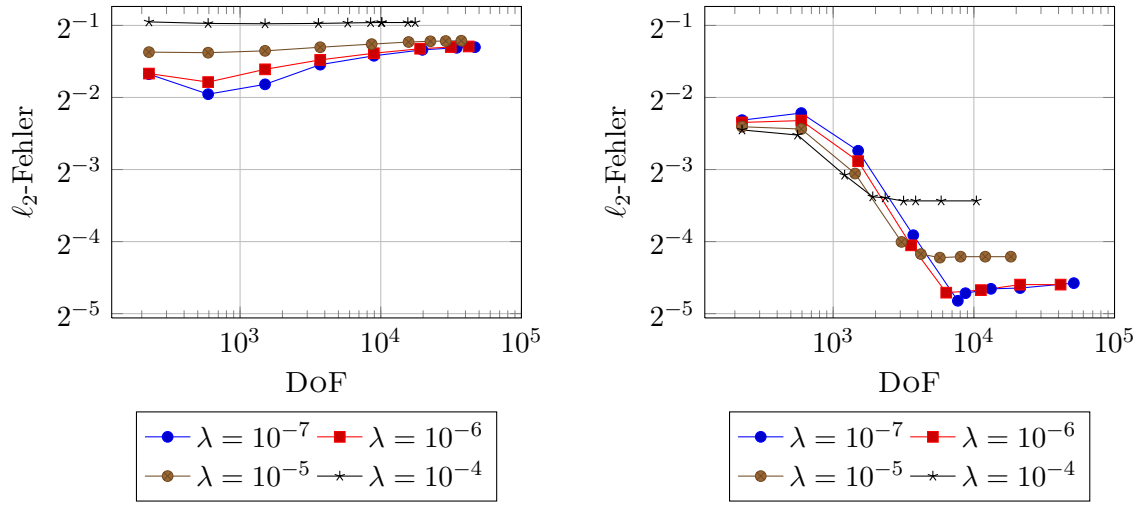
Wir benutzen nun weitere $\tilde{M} = 1000$, zufällig gewählte Punkte als so genannten Testdatensatz, auf dem wir den ℓ_2-Abstand

$$\frac{1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} (y^{(i)} - f_h(\boldsymbol{x}^{(i)}))^2$$

als das Fehlermaß für unsere Versuche definieren.

In den Abbildungen 5.15 und 5.16 sind die Erkennungsraten auf dem Testdatensatz in der ℓ_2-Norm für verschiedene Parameter $\lambda \in \{10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}\}$ in zwei und drei Dimensionen dargestellt.

Man sieht, dass der optimierte Datensatz bei weniger Freiheitsgraden deutlich besser gelernt werden kann, als das nicht-transformierte Problem. Dies ist darauf zurückzuführen, dass durch



(a) $d = 3$, Std

(b) $d = 3$, Opt

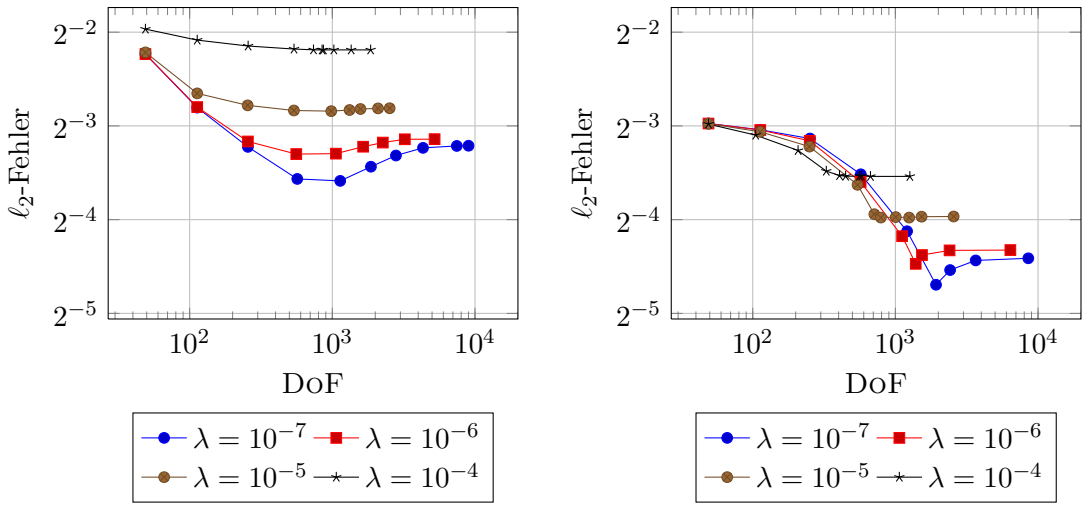
Abb. 5.16: Ridge-Datensatz (5.18) in $d = 3$ Dimensionen.

die Drehung ein Großteil der Varianz der Daten achsenparallel ausgerichtet werden kann. Da bei einem dünnen Gitter auf den Achsen, bzw. allgemeiner in den niedrigdimensionalen Teilräumen mehr Basisfunktionen lokalisiert sind, kann eine höhere Erkennungsrate erzielt werden.

Der gleiche Effekt spielt auch bei der Regression des von der Funktion (5.19) gesampleten Datensatzes eine Rolle. In den Abbildungen 5.17 und 5.18 sind die Erkennungsraten auf dem Testdatensatz in zwei und drei Dimensionen dargestellt. Auch hier ist zu sehen, dass der optimierte Datensatz deutlich besser gelernt werden kann.

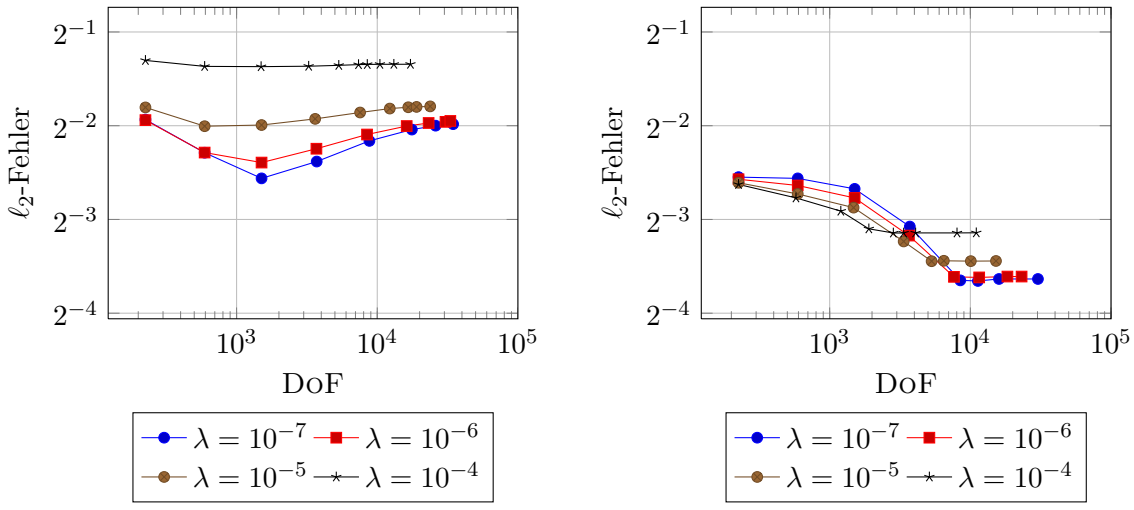
Bei beiden Testproblemen sind auch die für die Regression typischen Probleme zu beobachten. Während bei einer geringen Zahl von Freiheitsgraden ein niedriger Regularisierungsparamteter λ zu besseren Ergebnissen führt, stellen sich mit steigender Zahl der Freiheitsgrade die üblichen Overfitting-Effekte ein, d.h. die Dünngitter-Approximierende passt sich übermässig an die Trainingsdaten an – erzielt auf den Testdaten (die den dargestellten Konvergenzplots zugrunde liegen) jedoch immer schlechtere Ergebnisse.

Wir weisen nocheinmal darauf hin, dass wir hier nur synthetisch erzeugte Datensätze untersucht haben. Ob die Anwendung der hier vorgestellten Dimensionsreduktionsmethoden auf realen Datensätzen ähnlich gute Ergebnisse erzielt, wird Gegenstand zukünftiger Arbeiten sein.



(a) $d = 2$, Std (b) $d = 2$, Opt

Abb. 5.17: Test-Datensatz (5.19) in $d = 2$ Dimensionen.



(a) $d = 3$, Std (b) $d = 3$, Opt

Abb. 5.18: Test-Datensatz (5.19) in $d = 3$ Dimensionen.

6 Zusammenfassung und Ausblick

Wie bereits in der Einleitung bemerkt, ist die Wahl eines guten Koordinatensystems zur Beschreibung eines multivariaten Problems ein absolut entscheidender Faktor bei der Suche und Darstellung seiner Lösung.

Diese Erkenntnis ist nicht neu – in vielen Bereichen verwendet man Nichtstandard-Koordinaten, welche jedoch in der Regel schon aus der Modellierung stammen.

In den letzten Jahren haben sich jedoch in verschiedenen Bereichen Ansätze etabliert, solche Koordinatensysteme automatisiert zu bestimmen. Sehr populär im Bereich der Quadratur sind etwa die LT [IT04] und die DM [Mor98], doch auch bei der Lösung von PDEs zur Optionspreisbewertung finden sich solche Ideen wieder [Rei04] – meist werden die dabei verwendeten Konzepte jedoch nur in geringer Allgemeinheit und lediglich im Kontext des konkreten Problems betrachtet.

Obwohl wir den Schwerpunkt auf orthogonale Transformationen legen und damit noch immer in der Tradition der oben genannten Methoden stehen, stellt diese Arbeit den ersten uns bekannten Versuch dar, einen allgemeinen Formalismus zur Reduktion der effektiven Dimension multivariater Probleme und deren numerischer Behandlung in einer größeren Allgemeinheit zu schaffen und diese Konzepte auch außerhalb der Quadratur anzuwenden.

Den Weg dorthin werden wir abschließend kurz zusammenfassen und einen Ausblick auf Verbesserungsmöglichkeiten und vielversprechende Anwendungen geben.

Resultate dieser Arbeit

Wir zeigten, dass die ANOVA-Zerlegung das sinnvolle Werkzeug zur Erfassung verborgener Niederdimensionalität multivariater Funktionen darstellt, indem wir bewiesen, dass sie bezüglich der $\mathcal{L}^2$-Norm die beste niederdimensionale Approximation ist.

Zur Quantifizierung dieser Niederdimensionalität diskutierten wir die Begriffe Superpositions-, Trunkations- und Mittlere Dimension und stellten neuartige Zusammenhänge zwischen ihnen her. Zudem führten wir einen erweiterten Begriff von effektiver Dimension ein, der eine Verallgemeinerung der Mittleren Dimension darstellt und einer gewichteten Summe über die in einer problemabhängigen Norm gemessenen ANOVA-Terme entspricht. Damit lässt sich ein direkter Bezug zu den Fehlerabschätzungen dünner Gitter und Quasi-Monte Carlo Methoden herstellen.

Wir formalisierten die Reduktion der effektiven Dimension einer reellwertigen $\mathcal{L}^2$-Funktion f durch die Minimierung eines Funktionals $\mathfrak{M}_f$ auf einer Teilmenge $\Phi \subset \mathbf{Diff}$ der Menge aller Diffeomorphismen und diskutierten sinnvolle Einschränkungen an Φ.

Wir leiteten ein dazu äquivalentes Maximierungsproblem $\widehat{\mathfrak{M}}_f$ her, das bei einer geeigneten Wahl der zugrundeliegenden ANOVA-Zerlegung und Norm die gleichen Extremstellen besitzt wie $\mathfrak{M}_f$, jedoch die Vernachlässigung hoher ANOVA-Terme gestattet und somit die numerische Behandlung des Problemes erst ermöglicht.

Da sich für den Fall der speziellen orthogonalen Gruppe $\Phi = \mathbf{SO}(d)$ signifikante Vereinfachungen ergaben, beschränkten wir uns im weiteren auf diesen Fall.

Wir fassten $\mathbf{SO}(d)$ als differenzierbare Mannigfaltigkeit auf und entwickelten Ansätze, um $\widehat{\mathfrak{M}}$ über dieser zu optimieren. Dazu verwendeten wir nicht den konventionellen Zugang als Lagrange-Problem sondern moderne Konzepte aus der Differentialgeometrie, welche lokale Informationen über die Krümmung der Mannigfaltigkeit ausnutzen. Zu diesem Zweck verwendeten wir lokale Parametrisierungen der Mannigfaltigkeiten $\mathbf{SO}(d)$ und $\mathbf{St}(p, d)$ über ihren Tangentialräumen, für die wir eine effiziente Numerik entwickelten.

Die bei jeder Funktionalauswertung von $\widehat{\mathfrak{M}}_f$ auftretenden Integrale diskretisierten wir je nach Glattheit von f mit dünnen Gittern oder Quasi-Monte Carlo Methoden.

Für den Fall, dass f ein homogenes Polynom ist, leiteten wir eine analytische Darstellung des Funktionals $\widehat{\mathfrak{M}}_f$ her und diskutierten verschiedene Möglichkeiten, diese auf allgemeine $\mathcal{L}^2$-Funktionen anzuwenden. Dazu projizierten wir die zugrundeliegende Funktion in entsprechende Polynomräume, wo die optimale Koordinatentransformationen dann auf analytischem Wege deutlich schneller berechnet werden konnten. Den dazu notwendigen Projektionsoperator realisierten wir durch eine Dünngitter-Diskretisierung und der anschließenden Lösung eines gewichteten Least-Squares Problems.

Zur Verifikation dieser Vorgehensweise, wiesen wir anhand von ausgewählten Beispielen nach, dass unsere Ansätze in der Lage sind, die theoretisch optimale Transformation aufzufinden.

Außerdem betteten wir existierende Theorie in das von uns entwickelte Framework ein, indem wir zeigten, dass die Lineare Transformation (LT) und die Diagonal Methode (DM) einen Spezialfall unserer Polynommethode darstellen, wenn man anstelle der orthogonalen $\mathcal{L}^2$-Projektion eine abgeschnittene Taylorreihe als Projektionsoperator verwendet. Ferner bewiesen wir, dass LT und DM für Ridge-Funktionen stets die optimale Drehung auf eine achsenparallele Funktion liefern.

Wir stellten LT und DM unserer Polynomprojektionsmethode gegenüber, wobei deutlich wurde, dass LT und DM numerisch zwar weniger kostspielig, dafür jedoch auch weniger allgemein einsetzbar sind, da sie gewisse Differenzierbarkeitsanforderungen besitzen.

Anhand von sowohl synthetischen, als auch aus dem Bereich der Finanzmathematik entnommenen Funktionen untersuchten wir das Potential der in dieser Arbeit behandelten Dimensionsreduktionsmethoden und stellten fest, dass sich damit signifikante Verringerungen der effektiven Dimension erzielen lassen.

Daran anknüpfend untersuchten wir, inwiefern sich dies auf die Lösung hochdimensionaler Probleme anwenden lässt, wobei wir im Bereich der Interpolation und Regression mit dünnen Gittern anhand diverser Modellfunktionen feststellten, dass sich substanzielle Verbesserungen für die Konvergenz dieser Methoden ergeben können. Den Nachweis, dass unser Verfahren

in diesen Bereichen auch praktische Relevanz besitzt, gilt es in weiterführenden Arbeiten zu erbringen.

Im Bereich der Quadratur untersuchten wir neben Modellfunktionen hingegen auch praxisrelevante Probleme aus dem Bereich der Optionspreisbewertung anhand der pfadabhängigen Asiatischen und Basket-Optionen. Wir demonstrieren die Überlegenheit dünner Gitter und Quasi-Monte Carlo Folgen in dimensionsoptimierten Koordinaten. Insbesondere die Quadratur im Ganzraum durch Hermite-Polynome profitiert davon massiv, wenn der Integrand zusätzlich über eine genügende Glattheit verfügt.

Abschließend stellen wir fest, dass die Wichigkeit eines Vorverarbeitungsschrittes zur Reduktion der effektiven Dimension zwar im Bereich der hochdimensionalen Integration in den letzten Jahren sehr an Bedeutung gewonnen hat – in anderen Bereichen, wo hochdimensionale Probleme auftreten aber noch großes Potential besteht, bestehende Verfahren durch eine entsprechende „Vorkonditionierung“ des Problems zu verbessern. Einige Möglichkeiten sowie sinnvolle Erweiterungen unseres Ansatzes werden wir im Folgenden erläutern.

Ausblick

Anwendung auf partielle Differentialgleichungen

In [LO08, Rei04] werden lineare Transformationen zur Lösung von d-dimensionalen Diffusionsgleichungen zum Einpreisen von Basket-Optionen verwendet. Ähnlich wie bei der Integration wird eine Hauptkomponentenanalyse (PCA) der Kovarianzmatrix des Baskets verwendet, um eine geeignete Transformation zu berechnen.

Wie wir in dieser Arbeit jedoch dargelegt haben, ist die Diagonalisierung der Kovarianzmatrix zwar bezüglich des zugrundeliegenden Prozesses optimal – durch die Miteinbeziehung der Struktur der darauf definierten (Auszahlungs-)Funktion kann jedoch eine substanzielle Verbesserung erzielt werden.

Im Kontext der dimensionsadaptiven Kombinationstechnik erscheint eine Anwendung der in dieser Arbeit vorgestellten Konzepte daher vielversprechend.

Anwendung auf Dimensionsweise-Methoden

Bislang haben wir Drehungen $\boldsymbol{Q} : \mathbb{R}^d \to \mathbb{R}^d$ betrachtet, welche die gesamte Funktion (bzw. das gesamte Koordinatensystem) drehen. In [GH10b, Hol08] wird eine Methode vorgeschlagen, mehrdimensionale Funktionen durch ihre einzelnen ANOVA-Terme darzustellen und diese getrennt voneinander (also auf unterschiedlichen Gittern) zu integrieren:

$$\int f(\boldsymbol{x})\, d\mu(\boldsymbol{x}) = \sum_{\mathbf{u}\in\mathcal{D}} \int f_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})\, d\mu_{\mathbf{u}}(\boldsymbol{x})$$

Mit geringfügigen Anpassungen des Verfahrens aus [GH10b] ist es möglich, *jeden* multivariaten ANOVA-Term noch einmal *zusätzlich* separat zu drehen.

Wir wollen das Potential dieses Ansatzes anhand eines einfachen Beispiel demonstrieren, bei dem eine Funktion, die nominell und *auch effektiv* zweidimensional ist, in eine Summe eindimensionaler Funktionen zerfällt.

Beispiel 6.1: Auf $\mathbb{R}^2$ sei die Funktion

$$f(x_1, x_2) := \exp(x_1) + \exp(x_2) + x_1 \cdot x_2$$

gegeben. Der größte Anteil der Varianz liegt auf den beiden Achsen, das heißt die Funktion liegt bereits „optimal“ im Koordinatensystem – eine weitere Drehung zur Reduktion ihrer effektiven Dimension bringt also nichts.

Mit einem dimensionsweisen Verfahren (Ankerpunkt $\boldsymbol{a} = \mathbf{0}$) kann man nun $f_1(x_1) = \exp(x_1) - 1$ und $f_2(x_2) = \exp(x_2) - 1$ jeweils als eindimensionale Funktionen integrieren und lediglich den Term $f_{12}(x_1, x_2) = x_1 \cdot x_2$ als zweidimensionales Problem behandeln.

Für den ANOVA-Term $f_{12}(x_1, x_2)$ gilt dann weiter

$$f_{12}(\boldsymbol{Q}\boldsymbol{x}) = \frac{1}{2}(x_1^2 - x_2^2), \quad \text{wobei } \boldsymbol{Q} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Damit ist die ursprünglich zweidimensionale Funktion nach einer Drehung um $\pi/4$ in eine Summe lediglich eindimensionaler Funktionen zerfallen.

Anwendung auf die Approximation mit separablen Funktionen

Während achsenorientierte Verfahren wie die dünnen Gitter die additive Struktur hochdimensionaler Funktionen erfassen, gibt es auch Ansätze separable Strukturen zu erkennen und auzunutzen. In [BGM09] wird die Approximation multivariater Funktionen, durch eine Summe separabler Funktionen, also $f(\boldsymbol{x}) = \sum_{i=1}^{k} \prod_{j=1}^{d} f_{i,j}(x_j)$ dargestellt. Auch Adaptive-Cross-Approximation (ACA) [Beb09, Beb08] zur Approximation mehrdimensionaler Tensoren entspricht einem solchen Ansatz.

Es stellt sich die Frage, inwiefern sich die Zahl der benötigten Summanden k (der sogenannte *separation rank*) bei solchen Ansätzen durch geeignete Koordinatentransformationen verringern lässt.

Beispiel 6.2: Auch hier wollen wir uns anhand eines einfachen Beispiels einen ersten Eindruck vom Potential geeigneter Koordinatentransformatioben verschaffen, indem wir die Approximation des zweidimensionalen Polynoms $f(x, y) = x^2 - y^2$ untersuchen.

Dieses hat offensichtlich mindestens einen Separationsrang von $k = 2$. Wir betrachten wieder die Drehung um $\pi/4$, womit sich die Funktion

$$f(\boldsymbol{Q}\boldsymbol{x}) = 2x \cdot y, \quad \boldsymbol{Q} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

ergibt, welche den Separationsrang $k = 1$ besitzt.

Anwendung auf beliebige Levy-Prozesse

Im Bereich der multivariaten Quadratur haben wir ausschließlich das Gauß-Maß betrachtet, welches inbesondere im Fall von Funktionen, die auf stochastischen Prozessen definiert sind, die auf einer Brownschen Bewegung beruhen, eine Rolle spielt.

Die meisten modernen Modelle basieren jedoch mittlerweile auf den allgemeineren Levy-Prozessen. In [IT] wird dargelegt, dass sich die dabei auftetenden Integrale ebenfalls durch orthogonale Transformationen hinsichtlich ihrer effektiven Dimension optimieren lassen. Dies eröffnet weitere interessante Anwendungsgebiete im Bereich der Finanzmathematik.

Wahl der ANOVA-Zerlegung und Norm

Bisher haben wir stets die ANOVA-Zerlegung von zum Lebesgue-Maß absolut-stetigen Maßen μ betrachtet. Die Beiträge zum Minimierungsfunktion wurden in der durch μ induzierten $\mathcal{L}^2$-Norm gemessen, wodurch wir die Orthogonalität der ANOVA-Zerlegung ausnutzen konnten.

Unser Ansatz lässt prinzipiell jedoch auch andere ANOVA-Zerlegungen und Normen zu, um diese zu quantifizieren. Interessant wäre es, eine Funktion f bezüglich des Dirac-Mases zu zerlegen (Anker-ANOVA) und als Norm $\|f_{\mathbf{u}}\|_* := \int_{\Omega^{(d)}} D^{\mathbf{2}} f(\boldsymbol{x})^2 \, d\boldsymbol{x}$ zu betrachten[1].

Mit komplexeren Diffeomorphismen (etwa der Hintereinanderschaltung von Drehungen und Streckungen) könnte man Koordinatentransformationen entwickeln, welche nicht nur die effektive Dimension, sondern zugleich auch die Krümmung von f verringern. Es ist zu erwarten, dass sich die Approximationsgüte von dünnen Gittern damit noch weiter verbessern lässt.

Komponentenweise Streckungen des Koordinatensystems

In den Beispielen 3.2, 3.1 und den Abbildungen 3.3, 3.4 betrachteten wir Diffeomorphismen, die komponentenweise agieren. Ihre Komplexität steigt daher nur linear in der Dimension d.

Wir konnten zeigen, dass die Erzeugung derartiger Gittergradierungen mit dem in dieser Arbeit eingeführten Formalismus für den Fall $d = 1$ eine Darstellung als Minimierungsproblem auf der Sphäre S^{N-1} besitzt, wobei N die Zahl der Freiheitsgrade der eindimensionalen Bijektion bezeichnet.

Für den Fall $d = 1$ sind solche Gradierungen jedoch eher uninteressant, da eine adaptive Gittergenerierung wesentlich schneller und effizienter zum gleichen Ziel führt. Auch in mittleren Dimensionen $d = 2, \ldots, 15$ existieren effiziente ortsadaptive Varianten der dünnen Gitter [Feu10], deren Datenstrukturen in hohen Dimensionen jedoch nicht mehr handhabbar sind. In solchen Fällen (etwa der bei der Approximation von Wahrscheinlichkeitsdichten) könnten solche a-priori optimierten Gittergradierungen also wieder interessant sein.

[1] In diesem Fall geht jedoch die Orthogonalitätseigenschaft der ANOVA-Zerlegung verloren, was die Entwicklung einer sinnvollen Theorie deutlich erschwert.

Literaturverzeichnis

[AMS08] Absil, P.-A., R. Mahony und R. Sepulchre: *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[Arm66] Armijo, Larry: *Minimization of functions having Lipschitz continuous first partial derivatives*. Pacific J. Math., 16:1–3, 1966.

[Beb08] Bebendorf, M.: *Hierarchical Matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems*, Band 63 der Reihe *Lecture Notes in Computational Science and Engineering (LNCSE)*. Springer-Verlag, 2008.

[Beb09] Bebendorf, M.: *Adaptive Cross Approximation of Multvariate Functions*. 2009. SFB Preprint 453, to appear in Constructive Approximation.

[Bel61] Bellman, R.: *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[BG04] Bungartz, Hans-Joachim und Michael Griebel: *Sparse grids*. Acta Numerica, 13:1–123, 2004.

[BGM09] Beylkin, Gregory, Jochen Garcke und Martin J. Mohlenkamp: *Multivariate Regression and Machine Learning with Sums of Separable Functions*. SIAM Journal on Scientific Computing, 31(3):1840–1857, 2009.

[Boh10] Bohn, B.: *Einbettung von Zeitreihen nach Takens' Theorem*. Diplomarbeit, Institut für Numerische Simulation, Universität Bonn, März 2010.

[CL10] Cardoso, João R. und F. Silva Leite: *Exponentials of skew-symmetric matrices and logarithms of orthogonal matrices*. Journal of Computational and Applied Mathematics, 233(11):2867–2875, 2010.

[CMO97] Caflisch, Russel E., William Morokoff und Art Owen: *Valuation of Mortgage Backed Securities Using Brownian Bridges to Reduce Effective Dimension*, 1997.

[dC92] Carmo, Manfredo P. do: *Riemannian Geometry*. Birkhauser, 1992.

[DS83] Dennis, J. E. Jr. und R. B. Schnabel: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, NJ, 1983.

[EAS+98] Edelman, Alan, Tomás A. Arias, Steven T. Smith, Tom As, A. Arias, Steven und T. Smith: *The Geometry Of Algorithms With Orthogonality Constraints*. SIAM J. Matrix Anal. Appl, 20:303–353, 1998.

[Feu10] Feuersänger, Christian: *Sparse Grid Methods for Higher Dimensional Approximation*. Doktorarbeit, 2010.

[FG09] Feuersänger, Chr. und M. Griebel: *Principal Manifold Learning by Sparse Grids.* Computing, 85(4), August 2009. Also available as INS Preprint no 0801.

[FH07] Freund, Roland W und Ronald H. W Hoppe: *Stoer/Bulirsch: Numerische Mathematik 1*, 2007.

[Fis02] Fischer, Gerd: *Lineare Algebra: Eine Einführung für Studienanfänger.* Vieweg, F, 13 Auflage, 2002.

[Fle00] Fletcher, R.: *Practical Methods of Optimization.* Wiley & Sons, New York, 2 Auflage, 2000.

[Gar04] Garcke, J.: *Maschinelles Lernen durch Funktionsrekonstruktion mit verallgemeinerten dünnen Gittern.* Doktorarbeit, Institut für Numerische Simulation, Universität Bonn, 2004.

[GG98] Gerstner, T. und M. Griebel: *Numerical Integration using Sparse Grids.* Numer. Algorithms, 18:209–232, 1998. (also as SFB 256 preprint 553, Univ. Bonn, 1998).

[GG03] Gerstner, T. und M. Griebel: *Dimension–Adaptive Tensor–Product Quadrature.* Computing, 71(1):65–87, 2003.

[GGG10] Garcke, J., T. Gerstner und M. Griebel: *Intraday Foreign Exchange Rate Forecasting using Sparse Grids.* submitted to Journal of Econometrics 2010, also available as INS Preprint No. 1006, 2010.

[GH10a] Gao, Z. und J. S. Hesthaven: *On ANOVA expansions and strategies for choosing the anchor point.* Technischer Bericht 2010-17, Scientific Computing Group, Brown University, Providence, RI, USA, April 2010.

[GH10b] Griebel, M. und M. Holtz: *Dimension-wise Integration of High-dimensional Functions with Applications to Finance.* J. Complexity, 26:455–489, 2010. Also available as INS Preprint 0809.

[GKS10] Griebel, M., F. Y. Kuo und I. H. Sloan: *The Smoothing Effect of the ANOVA Decomposition.* J. Complexity, 26:523–551, 2010.

[Gla04] Glasserman, Paul: *Monte Carlo Methods in Financial Engineering.* Springer, 2004.

[Gri06] Griebel, M.: *Sparse grids and related approximation schemes for higher dimensional problems.* In: Pardo, L., A. Pinkus, E. Suli und M.J. Todd (Herausgeber): *Foundations of Computational Mathematics (FoCM05), Santander*, Seiten 106–161. Cambridge University Press, 2006.

[GX00] Gallier, Jean und Dianna Xu: *Computing Exponentials of Skew Symmetric Matrices And Logarithms of Orthogonal Matrices.* International Journal of Robotics and Automation, 18:10–20, 2000.

[Hic98] Hickernell, Fred J.: *A generalized discrepancy and quadrature error bound.* Mathematics of Computation, 67(221):299–322, 1998.

[Hol08] Holtz, M.: *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance*. Dissertation, Institut für Numerische Simulation, Universität Bonn, 2008.

[Hoo07] Hooker, Giles: *Generalized Functional ANOVA Diagnostics for High Dimensional Functions of Dependent Variables*. JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS, 16(3):709–732, 2007.

[Hul09] Hullmann, A.: *Schnelle Varianten des Generative Topographic Mapping*. Diplomarbeit, Institut für Numerische Simulation, Universität Bonn, Dezember 2009.

[HW08] Heiss, F. und V. Winschel: *Likelihood approximation by numerical integration on sparse grids*. Journal of Econometrics, 144(1):62–80, 2008.

[IT] Imai, Junichi und Ken Seng Tan: *A Generalized Linear Transformation Method for Simulating Meixner Lévy Processes*.

[IT04] Imai, Junichi und Ken Seng Tan: *Minimizing Effective Dimension using Linear Transformation*. Monte Carlo and Quasi-Monte Carlo Methods, 2002:275–292, 2004.

[IT06] Imai, Junichi und Ken Seng Tan: *A general dimension reduction technique for derivative pricing*. The journal of computational finance, 10(2):129–155, 2006.

[Kön09] Königsberger, Konrad: *Analysis 2*. Springer, 5 Auflage, 2009.

[KSWW08] Kuo, F. Y., I. H. Sloan, G. W. Wasilkowski und H. Wozniakowski: *On decompositions of multivariate functions*. Technischer Bericht, 2008.

[Lee97] Lee, J.M.: *Riemannian Manifolds: An Introduction to Curvature*. Springer, 1997.

[LO08] Leentvaar, C. C. W. und C. W. Oosterlee: *On coordinate transformation and grid stretching for sparse grid pricing of basket options*. J. Comput. Appl. Math., 222:193–209, December 2008.

[MC96] Moskowitz, B. und R. Calfisch: *Smoothness and dimension reduction in quasi-Monte Carlo methods*. J. Math. Comp. Modeling, 23:37–54, 1996.

[Mor98] Morokoff, William J.: *Generating Quasi-Random Paths for Stochastic Processes*. SIAM Review, 40:765–788, 1998.

[Nah05] Nahm, T.: *Error Estimation and Index Refinement for Dimension-Adaptive Sparse Grid Quadrature with Applications to the Computation of Path Integrals*. Diplomarbeit, Universität Bonn, 2005.

[Nie92] Niedereiter, Harald: *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial Mathematics, 1992.

[Noc92] Nocedal, Jorge: *Theory of Algorithms for Unconstrained Optimization*, 1992.

[NR96] Novak, E. und K. Ritter: *High dimensional integration of smooth functions over cubes*. Numer. Math., 75:79–97, 1996.

[NRS98] Novak, E., K. Ritter und A. Steinbauer: *A multiscale method for the evaluation of Wiener integrals*. In: *Approximation Theory IX, Volume 2: Computational*

Aspects, C. K. Chui and L. L. Schumaker (eds.), Seiten 251–258. Vanderbilt Univ. Press, 1998.

[NW01] Novak, E. und H. Woźniakowski: *When are integration and discrepancy tractable?* In: DeVore, R. A., A. Iserles und E. Süli (Herausgeber): *Foundations of Computational Mathematics (FoCM99), Oxford*, Seiten 211–266. Cambridge University Press, 2001.

[NW07] Novak, E. und H. Woźniakowski: *Discrepancy and Multivariate IntegrationL_2*. To appear in the Roth Festschrift, 2007.

[Owe03] Owen, Art B.: *The Dimension Distribution and Quadrature Test Functions*. Statistica Sinica, 13:1–17, 2003.

[Pol97] Polak, E.: *Optimization: Algorithms and Consisten Approximations*. Springer-Verlag, New York, 1997.

[QGA10] Qi, Chunhong, Kyle A. Gallivan und P.-A. Absil: *Riemannian BFGS algorithm with applications*. In: *Recent Advances in Optimization and its Applications in Engineering*. Springer, 2010. To appear.

[RA99] Rabitz, Herschel und Ömer Aliş: *General foundations of higher-dimensional model representations*. Journal of Mathematical Chemistry, 25:197–233, 1999.

[Rei04] Reisinger, Christoph: *Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben*. Doktorarbeit, Universität Heidelberg, 2004.

[Smo63] Smolyak, S.A.: *Quadrature and interpolation formulas for tensor products of certain classes of functions*. Dokl. Akad. Nauk SSSR, 4:240–243, 1963.

[Sob01] Sobol, I.: *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*. Math. Comput. Simulation, 55:271–280, 2001.

[SW97] Sloan, Ian H. und Henryk Wozniakowski: *When are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals?* J. Complexity, 14:1–33, 1997.

[SWW04] Sloan, Ian H., Xiaoqun Wang und Henryk Woźniakowski: *Finite-order weights imply tractability of multivariate integration*. J. Complex., 20(1):46–74, 2004.

[Wan06] Wang, X.: *On the Effects of Dimension Reduction Techniques on Some High-Dimensional Problems in Finance*. Operations Research, 54(6):1063–1078, 2006.

[Wan10] Wang, Xiaoqun: *High Dimensional Model Representations in Quasi-Monte Carlo Methods For Computational Finance*. To appear, 2010.

[Wat07] Waterhouse, Ben: *New developments in the construction of lattice rules*. Doktorarbeit, University of New South Wales, 2007.

[WF03] Wang, Xiaoqun und Kai-Tai Fang: *The effective dimension and quasi-Monte Carlo integration*. J. Complex., 19(2):101–124, 2003.

[WS03] Wang, Xiaoqun und Ian H. Sloan: *Why are high-dimensional finance problems often of low effective dimension?* SIAM J. Sci. Comput, 27:159–183, 2003.