Mathematics Subject Classification: 11K38, 41A55, 41A63, 65D30, 65Y20, 68Q05, 68Q25

Numerische Verfahren für Hochdimensionale Probleme und der Fluch der Dimension

Erich Novak, Erlangen

1 Einführung

Viele hochdimensionale Probleme sind schwierig, der Rechenaufwand der naheliegenden Verfahren wächst exponentiell mit der Dimension. Dies gilt für Probleme der numerischen Integration, für Probleme der Approximation bzw. Rekonstruktion von Funktionen, für das Problem der globalen Optimierung und für viele Integralgleichungen und partielle Differentialgleichungen. Es gilt *nicht* für Probleme der konvexen Optimierung und für Systeme von gewöhnlichen Differentialgleichungen. So ist die optimale Konvergenzordnung bei der Integration von C^k-Funktionen einer Variablen n^{-k} und mit Hilfe von Produktformeln erhält man in der Dimension $d > 1$ leicht Verfahren der Ordnung $n^{-k/d}$. Diese Konvergenzordnung ist für großes d sehr schlecht aber optimal, siehe Satz 5.

Wir fragen also nach der Güte von *optimalen* Verfahren und bemerken bereits hier, daß dies nur Sinn macht in Bezug auf eine *vorgegebene* Funktionenklasse F_d von Funktionen von d Variablen: hat man über $f : [0,1]^d \to \mathbf{R}$ keine *a priori Information*, so nützt die Berechnung von noch so vielen Funktionswerten $f(x_1), \dots, f(x_n)$ nichts; jeder Wert des Integrals bleibt möglich.

Gewisse hochdimensionale Probleme sind *prinzipiell* schwierig, der Rechenaufwand von optimalen Verfahren wächst exponentiell mit der Dimension. Ob ein Problem dem *Fluch der Dimension* unterliegt, hängt von den betrachteten Funktionenklassen ab, da gewisse hochdimensionale Probleme durchaus effizient gelöst werden können. Der Begriff der *optimalen Konvergenzordnung* spielt eine Rolle, ist aber für sich genommen zu schwach, um das Phänomen des Fluchs der Dimension (ein Ausdruck, der auf [6] zurückgeht) zu erfassen.

Wir werden einige klassische Ergebnisse aus der Theorie der optimalen numerischen Methoden vorstellen. Die meisten Ergebnisse sind jedoch aus den letzten Jahren und manche sind noch unveröffentlicht. Diese Arbeit versteht sich als eine Einführung. Wir beschränken uns auf solche Ergebnisse, die sich leicht beschreiben lassen. Beweise werden nur skizziert oder ganz weggelassen. Insbesondere möchte ich folgende Fragen diskutieren:

- Was ist ein hochdimensionales Problem und warum sind solche Probleme wichtig?

- Was ist ein numerisches Verfahren? Können wir Aussagen über *alle* numerischen Verfahren machen? Was ist ein optimales Verfahren?

- Wie beweist man *untere Schranken* für beliebige numerische Verfahren?

Wir werden Probleme kennenlernen, für die man exponentielle untere Schranken beweisen kann und andere, für die es polynomiale Algorithmen gibt. Ein solcher *Existenzbeweis* für polynomiale Algorithmen kann allerdings abstrakt oder konstruktiv sein, so daß wir drei verschiedene Situationen unterscheiden:

- Probleme, die dem Fluch der Dimension unterliegen (exponentielle untere Schranken)

- Probleme, die dem Fluch der Dimension nicht unterliegen (*Existenz* von polynomialen Algorithmen)

- Probleme, die dem Fluch der Dimension nicht unterliegen (*Konstruktion* von polynomialen Algorithmen).

Ein Problem rutscht von der zweiten in die dritte Klasse, wenn es gelingt, entsprechende Algorithmen zu konstruieren. Man muß auch zugeben, daß die Unterscheidung zwischen abstrakten und konstruktiven Existenzbeweisen fließend verläuft: nämlich dann, wenn die Konstruktion einer Methode zwar bloß endlich viele Rechenschritte erfordert, wenn aber gleichzeitig dieses endlich so groß ist, daß sich die Konstruktion praktisch nicht durchführen läßt. Deswegen interessieren wir uns besonders für *einfache* Konstruktionen und werden an geeigneter Stelle sagen, was wir darunter verstehen, siehe Bemerkung 21. Bei den klassischen Algorithmen, die auf Tensorprodukten beruhen, wächst die Rechenzeit exponentiell mit der Dimension. Höchstens mit neuen Algorithmen gelingt es, die Rechenzeit polynomial in der Dimension zu beschränken.

Insbesondere für den Beweis von unteren Schranken muß man das Berechenbarkeitsmodell spezifizieren. In der Numerischen Mathematik tendiert man dazu, das *real number model* zu benützen: gerechnet wird mit reellen Zahlen und man abstrahiert zunächst von Rundungsfehlern. Um allgemeine numerische Probleme (mit Funktionen f aus einer vorgegebenen Klasse F_d) behandeln zu können, gehen wir von einer *Orakelmaschine* aus. Das Orakel liefert uns Funktionswerte oder allgemeiner $L(f)$ für gewisse lineare Funktionale $L \in \Lambda_d$. Dazu mehr in Abschnitt 3.

Im Unterschied zur diskreten Komplexitätstheorie, bei der von vielen Problemen nur *vermutet* wird, daß sie schwer sind (im Sinne der Existenz exponentieller unterer Schranken), können wir bei vielen stetigen Problemen solche unteren Schranken *beweisen*.

Ein Reiz der vorgestellten Theorie besteht in den zahlreichen Zusammenhängen zu klassischen Fragen der Approximationstheorie, Stochastik und

Zahlentheorie[1]. Wer sich nur von einem unmittelbaren Nutzen überzeugen lassen will, möchte an die vielen Anwendungen denken, von denen wir im nächsten Abschnitt nur ein paar wenige Revue passieren lassen.

2 Hochdimensionale numerische Probleme

Bei einem numerischen Problem (im Sinne dieser Arbeit) möchte man eine Näherung von $S_d(f)$ berechnen. Dabei ist $S_d : F_d \to G_d$ eine Abbildung, F_d eine Menge von Funktionen von d (reellen) Variablen und G_d ein normierter Raum. Als Näherung berechnen wir Linearkombinationen

$$(1) \qquad g = \sum_{k=1}^{m} c_k \, g_{i_k}$$

oder, formaler, einen Output der Form

$$\mathrm{out}(f) = [i_1, c_1, i_2, c_2, \ldots, i_m, c_m],$$

mit $i_k \in \mathbf{N}$ und $c_k \in \mathbf{R}$, den wir mit dem Vektor g identifizieren. Der Fehler ist dann $e = \|S_d(f) - g\|$ und wir werden annehmen, daß ein g berechnet werden soll mit $e \leq \varepsilon$. Dabei ist $\varepsilon > 0$ die vorgegebene *Fehlertoleranz*.

Von einem *hochdimensionalen Problem* spricht man dann, wenn d groß ist. Wir werden sehen, daß die Rechenkosten häufig exponentiell mit dieser Anzahl d von Variablen wachsen. Daher sind Probleme mit $d = 5$ oder $d = 10$ oft schon sehr schwer und können als hochdimensional gelten. Wir beschränken uns auf wenige Beispiele.

- Die *Schrödinger-Gleichung* für $m > 1$ Teilchen ergibt ein $d = 3m$-dimensionales Problem.
- Die Berechnung von *hochdimensionalen Integralen* und von *Pfadintegralen* (z.B. von Integralen bezüglich des Wiener-Maßes) spielt in vielen Anwendungen eine Rolle, insbesondere in der Quantenphysik und in der Finanzmathematik. Die Lösung vieler Differentialgleichungen kann mit Hilfe von Pfadintegralen angegeben werden, Ausgangspunkt ist die klassische *Feynman-Kac-Formel*. Ähnliche Probleme ergeben sich bei der *Simulation von stochastischen Prozessen*. Siehe [23, 40, 67, 81, 84, 87].
- In der *globalen Optimierung* ist das Minimum einer Funktion gesucht. Solche Probleme gibt es in vielen Anwendungen, etwa in der Mustererkennung und der Bildverarbeitung, siehe [94], oder bei der Modellierung

[1] Insbesondere der Begriff der *Diskrepanz* aus der Theorie der Gleichverteilung ist ein Begriff mit vielen Bezügen zu den unterschiedlichsten Gebieten der Mathematik: Sobolev-Räume, Zufallsfelder, Kombinatorik, algebraische Geometrie, numerische Integration und Komplexitätstheorie. Entsprechende Literatur werden wir später nennen.

und Vorhersage der Geometrie von Eiweißmolekülen, siehe [41]. Sogenannte "Abkühlungsstrategien" (simulated annealing), "genetische" Algorithmen, Glättungstechniken und andere stochastische Algorithmen wurden entwickelt und teilweise mit großem Erfolg eingesetzt, siehe [9] und [71]. Die Arbeiten [44, 49] enthalten Fehlerabschätzungen für klassische Funktionenräume.

Es ist *nicht* so, daß die Dimensionalität eines Problems eine mathematische Invariante ist. Schließlich kann man jede stetige Funktion $f : [0,1]^d \to \mathbf{R}$ nach dem *Superpositionssatz von Kolmogorov* in der Form

$$(2) \qquad f(x_1, \dots, x_d) = \sum_{i=1}^{2d+1} g(\lambda_1 \varphi_i(x_1) + \dots + \lambda_d \varphi_i(x_d))$$

mit stetigen Funktionen g und φ_i von einer Variablen (wobei nur g von f abhängt) darstellen. Siehe Kapitel 17 von [35] für eine Darstellung des 13. Hilbertschen Problems und seiner Geschichte. In unserem Zusammenhang ist ein Problem gegeben durch (S_d, F_d, G_d, Λ_d), nicht durch d allein. Die Funktionenklassen F_d beschreiben häufig eine gewisse Glattheit, die sich bei der Darstellung (2) nicht von f auf g überträgt.

3 Numerische Verfahren

Das übliche Vorgehen der Numerischen Mathematik besteht darin, daß immer neue Algorithmen vorgeschlagen werden, die mit den bekannten Verfahren verglichen werden. Maßstab sind sowohl theoretische Eigenschaften als auch numerische Ergebnisse. Dieser Prozeß führt zu besseren Algorithmen und zu oberen Schranken für den Fehler bzw. den nötigen Rechenaufwand. Man benötigt auch untere Schranken und ein präzises Berechenbarkeitsmodell, wenn man gewisse sehr natürliche Fragen beantworten will:

- Wann sind adaptive Verfahren besser als *alle* nichtadaptiven?
- Wann sind *Monte-Carlo-Verfahren* (d.h. randomisierte Algorithmen) besser als *alle* deterministischen Methoden?
- Welche hochdimensionalen Probleme sind *prinzipiell* schwierig?
- Brauchen wir *unstetige* Vergleichsoperationen zur Lösung eines vorgegebenen Problems?

Wir verwenden das einfachste Berechenbarkeitsmodell der Arithmetik, indem wir die vier Grundrechenarten (mit reellen Zahlen) zulassen und dazu noch Sprünge im Programm, die vom Ergebnis eines Vergleichs der Form ($a \le b$)? abhängen. Solche Maschinen nennt man *algebraische RAMs* (random access machines). Aufgrund der wichtigen Arbeit [7] spricht man auch vom BSS-Modell. Für einen Überblick siehe [8] und [38]. Insbesondere wegen der (unstetigen) Vergleichsoperation, die im Modell exakt ausgeführt wird, sind reelle Maschinen sicherlich nur eine Abstraktion der Wirklichkeit und nicht immer angemessen. Dies wird in den Arbeiten [91] und [98] diskutiert. Alternativ könnte man auch das klassische bit number model (z.B. die Turing-Maschine) zugrundelegen, siehe [10, 90].

Wir benötigen (zusätzlich zur algebraischen RAM) ein *Orakel* zur Berechnung von Funktionswerten. Input für viele numerische Berechnungen sind ja nicht n-Tupel von reellen Zahlen (wie in der algebraischen Komplexitätstheorie, siehe [12]), sondern Funktionen. Wir nehmen an, daß einzelne Funktionswerte (oder allgemeiner Werte von linearen Funktionalen wie etwa Fourier- oder Wavelet-Koeffizienten, wobei die Funktionale aus einer vorgegebenen Klasse Λ_d sind) berechnet werden können. Damit erhält man die folgende Liste der möglichen Operationen einer *algebraischen RAM mit Orakel*:

- Input: $f \in F_d$ und reelle Parameter $\alpha_1, \ldots, \alpha_k$
- Arithmetische Operationen: die vier Grundrechenarten (mit reellen Zahlen)
- Bedingte Sprünge, die vom Ergebnis eines Vergleichs $(a \le b)$? abhängen
- Aufruf des Orakels: berechne $L(f)$ für ein $L \in \Lambda_d$
- Im Fall von randomisierten Algorithmen: Aufruf eines Zufallszahlengenerators RAN $\in [0, 1]$ oder eines Generators für zufällige Bits COIN $\in \{0, 1\}$.
- Output: $\mathrm{out}(f) = [i_1, c_1, i_2, c_2, \ldots, i_m, c_m]$, wird mit $g \in G_d$ gemäß (1) identifiziert.

Dabei kann, je nach Aufgabenstellung, die *Basisfolge* $(g_i)_{i \in \mathbf{N}}$ in G_d fest vorgegeben sein oder frei vom Designer des Algorithmus wählbar sein.

Dieses Berechenbarkeitsmodell wird seit Jahrzehnten mehr oder weniger explizit benützt, aber selten präzise beschrieben. Auch wir verzichten hier auf diese Formalitäten und verweisen den interessierten Leser auf [45]. Weitere Aspekte dieses Modells werden in [57–59] untersucht. Will man auch parallele Algorithmen untersuchen, so muß man das Modell erweitern, siehe [27].

Die *Kosten* einer Berechnung kann man einfach dadurch messen, daß man die Anzahl der Operationen zählt und eine (geeignet gewichtete) Summe betrachtet. Häufig ignoriert man Input- und einfachste Kopierbefehle und setzt die Kosten 1 für die arithmetischen Operationen und die Vergleichsoperation fest. Orakelaufrufe entsprechen oft komplizierten Berechnungen oder gar physikalischen Messungen, die viel teurer sein können als einfache arithmetische Operationen. Daher legen wir ihre Kosten als $\mathbf{c} > 0$ fest und stellen uns meistens vor, daß $\mathbf{c}$ viel größer als 1 ist.

Gelegentlich sind die Rechenkosten irrelevant, weil sie nur einmal anfallen, und man möchte die Anzahl der auszurechnenden reellen Zahlen minimieren. So ein Fall, der einer *Datenkompression* entspricht, kann leicht durch eine entsprechende Definition der Kosten modelliert werden, siehe [58].

Wir studieren in dieser Übersicht das *worst case setting*, d.h. den maximalen Fehler von Algorithmen und die maximalen Kosten. Man kann auch das *average case setting* betrachten und zum Beispiel die optimale Rekonstruktion von *zufälligen Funktionen* untersuchen. Dieses Problem wird auch in der *Statistik* studiert. Manche Ergebnisse gelten ähnlich im worst case wie im average case setting. Teilweise gibt es allerdings sehr erhebliche Unterschiede, siehe [68] für eine umfassende Übersicht.

Das folgende typische Beispiel mag der Motivation dienen. Gegeben sei eine Operatorgleichung

$$(Au)(x) = f(x) \quad \text{für alle } x \in \Omega$$

auf einer beschränkten Menge $\Omega \subset \mathbf{R}^d$, etwa eine Integralgleichung oder eine Randwertaufgabe für eine partielle Differentialgleichung. Wir nehmen an, daß A linear und bijektiv ist und daß wir Funktionswerte von $f \in F_d$ berechnen können (oder irgendwie gegeben haben). Wir wollen die Lösung u ausrechnen und erlauben einen Fehler ε, gemessen in der Norm eines Raumes G_d.

Dieses Problem ist ein Beispiel für die folgende Aufgabe: Berechne eine Approximation von $S_d(f)$, wobei der Operator

$$S_d : X_d \to G_d$$

linear ist und $f \in F_d \subset X_d$, es gilt $S_d = A^{-1}$. Oft ist F_d eine Kugel in einem Sobolev-Raum X_d und G_d ist auch ein Sobolevraum, im einfachsten Fall etwa $G_d = L_2(\Omega)$. Häufig werden S_d und $f \in F_d$ mit einer finiten Elementmethode diskretisiert.[2]

Entsprechend beruhen numerische Methoden auf einer *a priori (oder globalen) Information* $f \in F_d$ und der *berechneten (oder lokalen) Information*

$$N(f) = (L_1(f), L_2(f), \ldots, L_n(f)) \in \mathbf{R}^n, \tag{3}$$

mit stetigen linearen Funktionalen $L_i \in \Lambda_d$, entsprechend unserer Annahme also

$$N(f) = (f(t_1), f(t_2), \ldots, f(t_n)).$$

Ein Informationsoperator $N : F_d \to \mathbf{R}^n$ heißt *nichtadaptiv*, falls die Funktionale L_i bzw. die *Stützstellen* t_i fest gewählt sind, d.h. nicht von $f \in F_d$ bzw. den vorher evtl. schon berechneten Werten $f(t_1), \ldots, f(t_{i-1})$ abhängen. Allgemeiner sind *adaptive Verfahren*, bei denen erst im Lauf der Rechnung festgelegt wird,

[2] Die *Lösung von Operatorgleichungen* wird im folgenden keine große Rolle spielen. Ich möchte nur verweisen auf [11, 18, 20, 24, 26, 39, 61, 83, 92, 93]. Die Komplexität von Operatorgleichungen ist im wesentlichen bestimmt durch die n-te Weite (im Sinne von Gelfand) der Lösungsmenge.

welche Stützstellen noch dazugenommen werden sollen. Man kann hoffen, im Verlauf der Berechnung von $L_1(f), \ldots, L_{i-1}(f)$ über f zu lernen, um den Fehler durch eine geschickte Wahl von L_i zu verkleinern.

Der *Fehler* einer Abbildung $\varphi \circ N$ mit N wie in (3) und $\varphi : \mathbf{R}^n \to G_d$ ist gegeben durch

$$e(\varphi \circ N, f) = \|S_d(f) - \varphi(N(f))\|,$$

der *maximale Fehler* ist

$$e(\varphi \circ N, F_d) = \sup_{f \in F_d} \|S_d(f) - \varphi(N(f))\|.$$

Wir definieren nun die ε-*Komplexität* eines Problems in einer Weise, die unabhängig von Skalierungsfaktoren ist. Damit meinen wir folgendes. Für die meisten Probleme gibt es einen trivialen Algorithmus. Technisch gesehen ist es der beste Algorithmus, der keinerlei Information über $f \in F_d$ benützt, also die beste konstante Abbildung $f \mapsto c \in G_d$. Der Fehler dieses Algorithmus ist

$$e_0 := \inf_{c \in G_d} \sup_{f \in F_d} \|S_d(f) - c\|$$

und stimmt überein mit dem Radius der Lösungsmenge $S_d(F_d)$. Ist S_d linear und F_d die Einheitskugel eines Raumes X_d, so ist e_0 einfach die Norm von S_d. Wir betrachten den maximalen Fehler relativ zu e_0 und gelangen so zu der folgenden Definition.

Definition 1. *Die ε-Komplexität eines Problems ist definiert durch die minimalen (worst case) Kosten eines Algorithmus $\varphi \circ N$ in der Klasse aller Algorithmen mit*

$$e(\varphi \circ N, F_d) \le \varepsilon\, e_0.$$

Die ε-Komplexität hängt ab von S_d, F_d, G_d und Λ_d, außerdem von der Folge der g_i. Deshalb schreiben wir $\mathrm{comp}\,(\varepsilon, S_d, F_d, G_d, \Lambda_d, (g_i)_{i \in \mathbf{N}})$. *Wenn die g_i frei gewählt werden können, erhalten wir*

$$\mathrm{comp}\,(\varepsilon, S_d, F_d, G_d, \Lambda_d) = \inf_{(g_i)_{i \in \mathbf{N}}} \mathrm{comp}\,(\varepsilon, S_d, F_d, G_d, \Lambda_d, (g_i)_{i \in \mathbf{N}}). \qquad (4) \qquad \square$$

Wir betrachten nun für F_d immer eine Klasse von Funktionen und für Λ_d die Klasse der Punktauswertungen, $L(f) = f(t)$ für alle f. Auch (S_d, F_d, G_d) wird häufig aus dem Zusammenhang ersichtlich sein und wir betrachten die durch (4) gegebene Komplexität. Sie hängt dann nur noch von d und ε ab und wird $\mathrm{comp}\,(\varepsilon, d)$ geschrieben.

In (3) betrachten wir eine Information mit *fester Kardinalität*. Allgemeinere Algorithmen, bei denen die Kardinalität $n = n(f)$ von einer *Stoppregel* abhängt, sind für die in dieser Arbeit studierten Probleme nicht besser. Daher werden sie in dieser Übersicht keine Rolle spielen.

Nichtadaptive Verfahren lassen sich leichter parallelisieren. Sie lassen sich auch leichter analysieren. Deswegen ist der folgende Satz, der Ergebnisse von Smolyak, Bakhvalov, Gal, Micchelli, Traub und Woźniakowski zusammenfaßt, bemerkenswert. Er behandelt eine *Standardsituation* der Numerik, nämlich Fehlerabschätzungen für *lineare Probleme* mit Hilfe einer *Norm von* f. Er zeigt insbesondere, daß adaptive Methoden für lineare Probleme auf symmetrischen und konvexen Mengen kaum besser sind als nichtadaptive, siehe [2, 46, 83].

Häufig genügt es sogar, lineare Verfahren zu betrachten. Zum Beweis dieser Tatsache benützt man Ergebnisse aus der Theorie der s-Zahlen oder n-ten Weiten, siehe [36, 63, 64, 65].

Satz 2. *Gegeben sei ein linearer Operator* $S_d : X_d \to G_d$ *und* F_d *sei die Einheitskugel von* X_d. *Sei* $\varphi \circ N$ *ein beliebiges adaptives Verfahren, das für* $f = 0 \in F_d$ *die Information*

$$N^0(f) = [L_1^0(f), L_2^0(f), \dots, L_n^0(f)]$$

benützt. Dann gibt es ein nichtadaptives Verfahren der Form $\varphi^* \circ N^0$ *mit*

$$e(\varphi^* \circ N^0, F_d) \le 2\, e(\varphi \circ N, F_d).$$

Daher ist der Fehler von adaptiven Verfahren mindestens halb so groß wie der von (optimalen) nichtadaptiven. Ist überdies X_d *ein Hilbertraum oder ist* G_d *ein Raum mit der metric extension property*[3], *so kann man auf den Faktor zwei verzichten und überdies* $\varphi^* \circ N^0$ *linear wählen,*

$$\varphi^* \circ N^0(f) = \sum_{i=1}^{n} a_i\, L_i^0(f). \qquad \square$$

In der Situation von Satz 2 gibt es Beispiele, bei denen adaptive Verfahren (etwas) besser sind als nichtadaptive, siehe [33]. Alle Voraussetzungen des Satzes sind wesentlich. So nehmen wir an, daß F_d konvex und symmetrisch ist. Die Menge F_d beschreibt die *a priori Information* über das Problem. Oft weiß man, daß f gewisse Glattheitseigenschaften hat und dieses Wissen läßt sich in der Form $f \in F_d$ ausdrücken. Wenn das a priori Wissen zu einer Menge führt, die nicht konvex oder nicht symmetrisch ist, kann man Satz 2 nicht anwenden. Adaptive Verfahren können dann sehr viel besser sein als nichtadaptive. Dasselbe gilt für *nichtlineare* Probleme, siehe die Übersicht [46].

Die Idee hinter dem Beweis von Satz 2 ist, daß eine nichtadaptive Information, die gut ist für $0 \in F_d$, auch für jedes andere $f \in F_d$ gut ist. Dies gilt für *jedes* lineare Problem und *jede* Norm. Allerdings gibt dieses Ergebnis keinen Hinweis darauf, *welche* nichtadaptive Information optimal ist. Insbesondere ist es *nicht* so, daß naheliegende Informationen (wie äquidistante Gitterpunkte oder uniforme Zerlegungen) immer gut sind. Tatsächlich gibt es viele wichtige

[3] zum Beispiel der Raum der beschränkten Funktionen auf einer Menge mit der sup-Norm

Beispiele, bei denen äquidistante Stützpunkte sehr schlecht sind. Wir betonen diese Tatsache, weil in der Literatur gelegentlich äquidistante Stützstellen (oder eine uniforme Zerlegung) verglichen werden mit einer guten adaptiven Methode und dann fälschlich geschlossen wird, daß eine adaptive Methode besser ist als *alle* nichtadaptiven Methoden.

Bemerkung 3. Für viele Probleme gibt es optimale Algorithmen, die linear sind, siehe Satz 2. Dann ist die Komplexität im wesentlichen bestimmt durch die minimale Anzahl n von Orakelaufrufen. Wir definieren $\mathrm{card}\,(\varepsilon, d)$ durch

$$\min\{n \mid \text{es gibt } \varphi : \mathbf{R}^n \to G_d \text{ und } N \text{ wie (3) mit } e(\varphi \circ N, F_d) \le \varepsilon\, e_0\}\,.$$

Sind lineare Algorithmen optimal, so gilt

$$\mathbf{c} \cdot \mathrm{card}\,(\varepsilon, d) \le \mathrm{comp}\,(\varepsilon, d) \le (\mathbf{c} + 2) \cdot \mathrm{card}\,(\varepsilon, d).$$

Deshalb wird es im folgenden genügen, die Größen $\mathrm{card}\,(\varepsilon, d)$ zu studieren und man kann diese Zahlen mit der Komplexität (nahezu) *identifizieren.* □

4 Untere Schranken

Das wichtigste Problem der theoretischen Informatik (nach Ansicht von Steve Smale eines der drei wichtigsten Probleme der Mathematik überhaupt, siehe [76][4]) bezieht sich auf die Klasse der Probleme, für die ein *polynomialer Algorithmus* existiert. Solche Probleme nennt man häufig *tractable* im Gegensatz zu den *intractable problems*, für die kein polynomialer Algorithmus existiert. Wir benützen die folgende Definition.

Definition 4. *Ein Problem, gegeben durch eine Folge* $(S_d, F_d, G_d, \Lambda_d)$, *heißt tractable, falls*

$$\mathrm{comp}\,(\varepsilon, d) \le K \cdot \varepsilon^{-q} \cdot d^p$$

mit gewissen Konstanten $K, q, p > 0$*; sonst heißt es intractable.* □

Diese Definition lehnt sich an die übliche Begriffsbildung für diskrete Probleme an, man beachte aber, daß wir das *real number model* zugrundelegen.[5] Im Bereich der Numerischen Mathematik (in Bezug auf das real number

[4] "In fact, included are what I believe to be the three greatest open problems of mathematics: the Riemann Hypothesis, the Poincaré Conjecture, and 'Does P = NP?'"

[5] Manche Probleme lassen sich im diskreten Berechnungsmodell der Turing-Maschine und im real number model untersuchen. Welches Modell angemessen ist, hängt von der Anwendung ab. Ein Problem kann im einen Modell tractable sein und im anderen intractable, da sich die Voraussetzungen bezüglich Rechenkosten und Problemgröße erheblich unterscheiden. Bekanntlich ist das Problem der Linearen Optimierung im Turing-Modell tractable. Ob dies auch im real number model gilt, ist ein wichtiges offenes Problem.

model mit Orakel) geht sie auf Woźniakowski zurück, siehe [96, 97]. Verschiedene klassische Probleme sind intractable, es gilt aber $\mathrm{comp}(\varepsilon, d) \leq C_d \cdot \varepsilon^{-q_d}$ für jedes d. Man sagt dann auch, daß so ein Problem dem *Fluch der Dimension* unterliegt.

Ist man sich erst einmal über den Begriff des Algorithmus im klaren, dann liegt es auf der Hand, daß man verschiedene untere Schranken beweisen kann, indem man die verschiedenen Rechenoperationen betrachtet:

- Man kann die Anzahl der benötigten Aufrufe des Orakels untersuchen und erhält damit Schranken für die *Informationskosten*.
- Man kann kombinatorische oder arithmetische Überlegungen anstellen und die Anzahl der benötigten Rechenoperationen untersuchen, d.h. die *arithmetischen Kosten*.
- Topologische Betrachtungen ermöglichen eine Abschätzung der Anzahl der nötigen Vergleichsoperationen, man spricht auch von der *topologischen Komplexität*. Wir werden darauf in dieser Arbeit nicht weiter eingehen und verweisen nur auf [28, 57, 74, 85].
- Mit Hilfe gewisser Begriffe der Approximationstheorie (*beste m-Term Approximation*, siehe [16]) kann man abschätzen, wieviele reelle Zahlen ausgerechnet werden müssen. Man kann hier von der *Output-Komplexität* sprechen, siehe [58].

Welche Abschätzungsmethode zu den besten unteren Schranken führt, hängt vom Problem ab. Überraschend häufig sind die Gesamtkosten von optimalen Algorithmen nur wenig größer als die Informationskosten; mit Satz 2 und Bemerkung 3 haben wir eine wichtige Klasse von Problemen mit dieser Eigenschaft bereits kennengelernt. Daher hat sich für diese Untersuchungen neben dem Begriff der *analytischen Komplexitätstheorie* auch der Begriff *Information-Based Complexity* eingebürgert, siehe [83] und die neue Einführung [84].

5 Probleme, die dem Fluch der Dimension unterliegen

In gewissen Fällen folgt aus klassischen Ergebnissen über die optimale Konvergenzordnung in Verbindung mit Satz 2 bereits, daß ein vorgelegtes Problem dem Fluch der Dimension unterliegt. Das folgende Ergebnis von Bakhvalov (aus dem Jahr 1959, siehe [1]) ist recht einfach zu beweisen, zumindest der Fall $d = 1$ eignet sich ausgezeichnet für eine einführende Veranstaltung in die Numerik.[6]

[6] Bisher spielen untere Schranken (nicht nur) in der Ausbildung kaum eine Rolle.

Satz 5. *Seien $k \in \mathbf{N}$ und X_d der Raum $C^k([0,1]^d)$ mit der Norm*

$$\|f\| = \max_{|\alpha| \le k} \|D^\alpha f\|_\infty.$$

Wir betrachten das Integrationsproblem

$$S_d(f) = \int_{[0,1]^d} f(x)\,dx$$

auf der Einheitskugel F_d von X_d. Dann gibt es positive Konstanten $c_{k,d}$ und $C_{k,d}$ mit

$$c_{k,d} \cdot \varepsilon^{-d/k} \le \mathrm{comp}\,(\varepsilon, d) \le C_{k,d} \cdot \varepsilon^{-d/k}. \qquad \square$$

Beweisidee: Aus Satz 2 folgt, daß es genügt, Quadraturformeln

$$Q_n(f) = \sum_{i=1}^{n} a_i\, f(t_i) \tag{5}$$

zu betrachten. Für die oberen Schranken reicht es, Produktformeln (etwa Tensorprodukte der Gauß-Formeln oder der Clenshaw-Curtis-Formeln) zu wählen. Für die unteren Schranken betrachte man ein beliebiges Q_n mit $2n \approx \ell^d$, wobei $\ell \in \mathbf{N}$, und die entsprechende Zerlegung von $[0,1]^d$ in ℓ^d kleine Würfel mit Seitenlänge ℓ^{-1}. Dann enthalten mindestens $\approx n$ der kleinen Würfel keine Stützstelle und damit kann $Q_n(f)$ auch nicht vom Verhalten von $f \in F_d$ auf diesen Teilwürfeln abhängen. Man kann nun zwei Funktionen $f_1, f_2 \in F_d$ konstruieren, die sich nur auf den Teilwürfeln ohne Stützstellen unterscheiden und für die

$$S_d(f_1) - S_d(f_2) \ge 2\,\varepsilon$$

gilt. Wegen $Q_n(f_1) = Q_n(f_2)$ folgt

$$e(Q_n, F_d) \ge \varepsilon$$

und daraus die Behauptung, siehe z.B. [44] für die Details. $\square$

Bemerkung 6. Wir wollen diesen typischen Satz und das Beweisschema kommentieren. Zunächst folgt, daß das Problem intractable ist im Sinne der Definition 4. Für jedes d erhält man die Rate $\mathrm{comp}\,(\varepsilon, d) \asymp \varepsilon^{-d/k}$ in einer konstruktiven Weise, man kann entsprechende Produktformeln explizit hinschreiben. Allerdings benötigen (nichttriviale) Produktformeln mindestens 2^d Stützstellen, so daß man schon bei $d = 40$ kaum noch von einer *sinnvollen Konstruktion* sprechen kann.

Bei genauerer Analyse wird man feststellen, daß sich die Konstanten $C_{k,d}$ von den Konstanten $c_{k,d}$ sehr unterscheiden, ein Verhältnis

$$\frac{C_{k,d}}{c_{k,d}} \approx 5^{d-1}$$

ist schon für kleine k realistisch. Daraus folgt, daß Satz 5 zwar die Kosten bei *festen* k und d für $\varepsilon \to 0$ gut wiedergibt, aber kaum etwas aussagt über die Frage, wie $\mathrm{comp}(\varepsilon, d)$ mit der *Dimension* d wächst. In vielen Anwendungen spielt der Grenzübergang $\varepsilon \to 0$ kaum eine Rolle, häufig reicht eine bescheidene Genauigkeit. Die klassischen Fehlerabschätzungen erweisen sich also als recht unbefriedigend; wir wollen wissen, wie $\mathrm{comp}(\varepsilon, d)$ von ε *und* von d abhängt. Dieses Problem studieren wir im folgenden für Funktionenklassen, die eng mit verschiedenen Begriffen der Diskrepanz gekoppelt sind. $\square$

Mit der *Koksma-Hlawka-Ungleichung* kann man den Fehler $e(Q_n, F_d)$ für gewisse F_d durch die Diskrepanz ausdrücken, siehe [29, 42]. Für Punkte $t_1, \ldots, t_n \in [0, 1]^d$ ist die L_2-*Diskrepanz* gegeben durch

$$\mathrm{disc}_2(\{t_1, \ldots, t_n\}) = \left(\int_{[0,1]^d} \left(x_1 \cdots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right)^2 dx \right)^{1/2}.$$

Dabei bezeichnet $1_{[0,x)}$ die Funktion, die 1 ist auf dem Quader $\{y \mid 0 \le y_i < x_i,\ i = 1, 2, \ldots, d\}$ und 0 außerhalb. Die L_2-Diskrepanz drückt aus, wie gut man im quadratischen Mittel das Volumen eines Quaders $[0, x) \subset [0, 1]^d$ durch das Verhältnis $\frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i)$ der in ihm enthaltenen Punkte schätzen kann.

Wir wollen auch allgemeine Quadraturformeln (5) betrachten und definieren ihre L_2-Diskrepanz als

$$\mathrm{disc}_2(Q_n) = \left(\int_{[0,1]^d} \left(x_1 \cdots x_d - \sum_{i=1}^{n} a_i\, 1_{[0,x)}(t_i) \right)^2 dx \right)^{1/2}.$$

Als nächstes definieren wir spezielle Sobolevräume X_d. Für $d = 1$ sei

$$X_1 = \{ f : [0,1] \to \mathbf{R} \mid f(1) = 0,\ f \text{ ist abs. stetig, } f' \in L_2 \}$$

mit der Norm $\|f\| = \|f'\|_{L_2}$. Für $d > 1$ betrachten wir das d-fache *Tensorprodukt*

$$X_d = X_1 \otimes \ldots \otimes X_1 \qquad \text{mit} \qquad \|f\| = \|f^{(1,1,\ldots,1)}\|_{L_2}. \tag{6}$$

Wie immer sei F_d die Einheitskugel von X_d. Dann folgt aus der Koksma-Hlawka-Ungleichung, daß

$$\mathrm{disc}_2(Q_n) = e(Q_n, F_d).$$

Außerdem ist $\mathrm{disc}_2(Q_n)$ auch gleich dem durchschnittlichen Fehler von Q_n bezüglich des Maßes, das vom Brownschen Blatt auf $C([0,1]^d)$ induziert wird, siehe [95]. Berühmte Sätze von Roth und Frolov [21, 69, 70] besagen, daß

$$\inf_{t_i} \mathrm{disc}_2(\{t_1, \ldots, t_n\}) \asymp n^{-1} \cdot (\log n)^{(d-1)/2}.$$

Insbesondere ist die Konvergenzordnung kaum abhängig von d, für das Integrationsproblem $S_d(f) = \int_{[0,1]^d} f(x)\,dx$ auf F_d gilt

$$\text{(7)} \qquad \operatorname{card}(\varepsilon, d) \le C_{d,\delta} \cdot \varepsilon^{-1-\delta}, \qquad \delta > 0.$$

Zum Begriff der Diskrepanz siehe [4, 5, 17, 34, 42, 43] und Bemerkung 12.

Aufgrund von (7) könnte man vermuten, daß auch die Komplexität nur schwach von d abhängt und insbesondere, daß ein Fluch der Dimension hier nicht vorliegt. Diese Vermutung ist falsch, wie in [99] gezeigt wird. Es gilt nämlich folgende *exponentielle* untere Schranke für die L_2-Diskrepanz (mit gleichen Gewichten $a_i = 1/n$) oder allgemeiner für Quadraturformeln Q_n mit positiven Gewichten.

Satz 7. *Betrachte das Integrationsproblem $S_d(f) = \int_{[0,1]^d} f(x)\,dx$ auf den durch (6) definierten Räumen. Für beliebige $d \in \mathbf{N}$ und Quadraturformeln Q_n mit positiven Gewichten a_i gilt*

$$\text{(8)} \qquad e(Q_n, F_d) \ge e_0 \cdot \left(1 - \left(\frac{8}{9}\right)^d n\right)_+^{1/2}. \qquad \square$$

Zum Beweis benützt man Techniken für Hilberträume mit reproduzierendem Kern, wobei nur wesentlich ist, daß der Kern

$$K(x,t) = \prod_{\ell=1}^{d} (1 - \max(x_\ell, t_\ell))$$

von X_d (punktweise) positiv ist. Diese Techniken führen auch für sehr kleine Räume von analytischen Funktionen zu einem ähnlichen Ergebnis. In [48] wird ein zu Satz 7 analoges Ergebnis für den Raum

$$X_d^\infty = \{f : [0,1]^d \to \mathbf{R} \mid \|f\|^2 = \sum_{\alpha \in \mathbf{N}_0^d} \|D^\alpha f\|_2^2 < \infty\}$$

bewiesen. Man beachte, daß für X_d^∞ *jede* Konvergenzordnung erreichbar ist, es gilt

$$\operatorname{card}(\varepsilon, d) \le C_{d,k} \cdot \varepsilon^{-1/k}$$

für jedes $k > 0$. Dennoch gilt wieder eine exponentielle untere Schranke der Form (8).

Bemerkung 8. Diese Ergebnisse zeigen, daß der Begriff der *optimalen Konvergenzordnung* zu schwach ist, um das Phänomen des Fluchs der Dimension zu erfassen.

Der Beweis von Satz 7 funktioniert nur für Quadraturformeln mit *positiven* Gewichten. Untere Schranken, die nur für gewisse Methoden gelten, sehen wir als unbefriedigend an, es sind noch keine Schranken für die Komplexität. Tatsächlich benützen gewisse hocheffiziente Quadraturformeln (vom Smolyak-Typ, siehe Abschnitt 7) auch negative Gewichte und es war zunächst unklar, ob mit diesen Formeln polynomiale Algorithmen möglich sind. Dieses Problem wurde in [60] gelöst:

Satz 9. *Betrachte wieder das Integrationsproblem* $S_d(f) = \int_{[0,1]^d} f(x)\,dx$ *auf den durch (6) definierten Räumen. Dann gilt:*

$$\operatorname{card}(\varepsilon, d) \quad \textit{ist nicht beschränkt durch} \quad C\,\varepsilon^{-\gamma}\,d^{\beta},$$

d.h. die L_2*-Diskrepanz ist nicht tractable im Sinne von Definition 4.* □

Der Beweis beruht darauf, daß der Tensorprodukt-Hilbertraum X_d eine Zerlegbarkeitseigenschaft besitzt. Der Raum X_1 besitzt zwei nichttriviale orthogonale Unterräume $X_{(0)}$ und $X_{(1)}$, so daß

$$f(x) = 0 \quad \text{für} \quad x \notin D_{(0)}, \quad f \in X_{(0)},$$

und

$$f(x) = 0 \quad \text{für} \quad x \notin D_{(1)}, \quad f \in X_{(1)},$$

mit disjunkten Mengen $D_{(0)}$ und $D_{(1)}$ positiven Maßes. Viele Hilberträume haben diese Eigenschaft und entsprechend allgemein ist dieser Zugang. Allerdings ist auch klar, daß Hilberträume von analytischen Funktionen diese Eigenschaft *nicht* besitzen. Deshalb ist noch unklar, ob z.B. für die Räume X_d^∞ der Fluch der Dimension gilt.

6 Ein abstrakter Existenzsatz: polynomiale Schranken für die Stern-Diskrepanz

Wir betrachten jetzt die Stern-Diskrepanz

$$\operatorname{disc}^*(\{t_1, \ldots, t_n\}) = \sup_{x \in [0,1]^d} \left| x_1 \cdots x_d - \frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right|,$$

wobei $t_1, \ldots, t_n \in [0,1]^d$, und entsprechend die Stern-Diskrepanz $\operatorname{disc}^*(Q_n)$ von beliebigen Quadraturformeln Q_n. Mit Hilfe der Koksma-Hlawka-Ungleichung läßt sich dieser Begriff wieder als Fehler von Q_n interpretieren,

$$\operatorname{disc}^*(Q_n) = e(Q_n, F_d),$$

allerdings muß dazu der Sobolevraum X_d (und damit seine Einheitskugel F_d) modifiziert werden. Wir betrachten jetzt für $d = 1$

$$X_1 = \{f : [0,1] \to \mathbf{R} \mid f(1) = 0,\ f \text{ ist abs. stetig},\ f' \in L_1\}$$

mit der Norm $\|f\| = \|f'\|_{L_1}$. Für $d > 1$ betrachten wir das d-fache Tensorprodukt

$$X_d = X_1 \otimes \ldots \otimes X_1 \qquad \text{mit} \qquad \|f\| = \|f^{(1,1,\ldots,1)}\|_{L_1}. \tag{9}$$

Es ist bekannt, daß

$$\inf_{t_i} \operatorname{disc}^*(\{t_1, \ldots, t_n\}) \le C_d \cdot n^{-1} \cdot (\log n)^{(d-1)} \tag{10}$$

und man vermutet, daß diese Ordnung optimal ist. Insbesondere ist die Konvergenzordnung wieder nahezu unabhängig von d, für das Integrationsproblem $S_d(f) = \int_{[0,1]^d} f(x)\,dx$ auf F_d gilt

$$\operatorname{card}(\varepsilon, d) \le C_{d,\delta} \cdot \varepsilon^{-1-\delta}, \qquad \delta > 0. \tag{11}$$

Aufgrund von Satz 2 gilt

$$\operatorname{card}(\varepsilon, d) = \inf\{n \mid \operatorname{disc}^*(Q_n) \le \varepsilon \ \text{ für ein } Q_n\},$$

tatsächlich gelten unsere oberen Schranken von Satz 13 auch für die Stern-Diskrepanz mit gleichen Gewichten $1/n$.

Bemerkung 10. Mancher Leser wird an dieser Stelle verwundert sein. Satz 9 besagt, daß das Problem der L_2-Diskrepanz *nicht* tractable ist. Dabei geht es um die Frage, wie gut man *im Durchschnitt* das Volumen eines Quaders $[0, x) \subset [0, 1]^d$ durch das Verhältnis $\sum_{i=1}^n 1_{[0,x)}(t_i)$ der in ihm enthaltenen Punkte schätzen kann.

In diesem Abschnitt fragen wir, ob die Stern-Diskrepanz tractable ist. Dabei geht es um die Frage, wie gut man *im worst case* das Volumen eines Quaders $[0, x) \subset [0, 1]^d$ durch das Verhältnis $\sum_{i=1}^n 1_{[0,x)}(t_i)$ der in ihm enthaltenen Punkte schätzen kann. Dies scheint schwieriger zu sein und man wird vielleicht erwarten, daß dieses Problem ebenfalls dem Fluch der Dimension unterliegt.

Es ist aber so, daß wir in Definition 1 nicht *absolute* Fehler betrachten, sondern wir fordern bei der ε-Komplexität, daß der Fehler beschränkt ist durch $\varepsilon\, e_0$. Nun ist zwar bei der Stern-Diskrepanz $e_0 = 1$ für alle d, so daß dieser Normierungsfaktor hier nicht wesentlich ist. Bei der L_2-Diskrepanz gilt aber $e_0 = 3^{-d/2}$, anschaulich: die meisten Teilquader von $[0, 1]^d$ sind sehr klein, daher ist die Schätzung "das Volumen beträgt 0" bei hoher Dimension *im Durchschnitt* ziemlich gut.

Würde man den Normierungsfaktor bei der L_2-Diskrepanz weglassen, dann wäre das Problem tractable, siehe die Arbeiten [37, 66, 86, 88]. □

Bemerkung 11. Gibt es $n = 3^d$ Punkte mit $\operatorname{disc}^*(\{t_1, \ldots, t_n\}) \le 1/3$, gilt also

$$\operatorname{card}(1/3, d) \le 3^d \ ?$$

Dies war vor kurzem noch unbekannt, siehe [34, 82]. Wir werden sehen (Satz 13, Bemerkung 17), daß card $(1/3, d)$ viel langsamer wächst als 3^d. Wieder zeigt sich, daß günstige Aussagen über die Konvergenzordnung wie (11) wenig bedeuten für die Komplexität in Abhängigkeit von der Dimension.

Man mag überrascht sein über die Tatsache, daß es so schwer ist, den (bescheidenen) Fehler von $\varepsilon = 1/3$ zu erreichen und kann sich zunächst fragen, wie viele Punkte man braucht, wenn man mit einem äquidistanten Gitter arbeitet – sicher die naheliegendste Konfiguration überhaupt. Seien also $d, \ell \in \mathbf{N}$, wir betrachten die $n = \ell^d$ Punkte $G_{d,\ell}$ mit Koordinaten der Form k/ℓ, wobei $k = 0, 1, \ldots, \ell - 1$. Dann gilt

$$\mathrm{disc}^*(G_{d,\ell}) = 1 - \left(1 - \frac{1}{\ell}\right)^d,$$

siehe [42]. Um also für ein (festes) ε mit $0 < \varepsilon < 1$ eine Stern-Diskrepanz $\mathrm{disc}^*(G_{d,\ell}) \le \varepsilon$ zu erzielen, braucht man ungefähr

$$n_G(\varepsilon, d) \approx \left(\frac{1}{-\log(1-\varepsilon)} d\right)^d$$

Punkte, man wähle $\ell \approx (-\log(1-\varepsilon))^{-1} d$. Diese Zahl wächst also sogar superexponentiell in d, für $\varepsilon = 1/3$ erhält man

$$n_G(1/3, d) \approx (2{,}47\, d)^d. \qquad \square$$

Bemerkung 12. Ich kann und will hier nicht die bekannten Konstruktionen von Punktmengen mit kleiner Stern-Diskrepanz diskutieren und verweise auf [17, 34, 42, 43]. Aber ich möchte wenigstens die einfachsten Punktmengen angeben, mit denen man die Schranke (10) zeigen kann. Sei $b \ge 2$ eine natürliche Zahl und

$$i = \sum_{j=0}^{\infty} d_j(i)\, b^j, \qquad d_j(i) \in \{0, 1, \ldots, b-1\}$$

die Entwicklung von $i \in \mathbf{N}$ bezüglich der Basis b. Dann ist die *Van der Corput-Folge* $(\gamma_b(i))_{i \in \mathbf{N}}$ definiert durch

$$\gamma_b(i) = \sum_{j=0}^{\infty} d_j(i)\, b^{-j-1}.$$

Aus naheliegenden Gründen spricht man auch von der "Spiegelung am Dezimalkomma". Seien nun $b_1, \ldots, b_{d-1}$ die ersten $d-1$ Primzahlen und sei $n \in \mathbf{N}$ gegeben. Dann ist die *d-dimensionale Hammersley-Folge der Länge n* gegeben durch

$$t_i = (\gamma_{b_1}(i), \ldots, \gamma_{b_{d-1}}(i), i/n), \qquad i = 1, \ldots, n.$$

Hammersley hat 1960 mit diesen Punkten die Schranke (10) bewiesen. $\square$

Das folgende Resultat aus [56] ist (zumindest für die Autoren der Arbeit) überraschend.

Satz 13. *Die Stern-Diskrepanz ist tractable, es gelten obere Schranken der Form*[7]

$$\inf\{n \mid \operatorname{disc}^*(t_1,\dots,t_n) \le \varepsilon \ \ \textit{für gewisse } t_i\} \le C_\delta \cdot \varepsilon^{-2-\delta} \cdot d^2$$

für jedes $\delta > 0$. □

Beweisidee: Der Beweis zerfällt in zwei Teile:

1. Teil: Analog zur L_2-Diskrepanz betrachten wir für gerades p die entsprechende L_p-Diskrepanz,

$$\operatorname{disc}_p(\{t_1,\dots,t_n\}) = \left(\int_{[0,1]^d} \left(x_1 \cdots x_d - \frac{1}{n}\sum_{i=1}^n 1_{[0,x)}(t_i)\right)^p dx\right)^{1/p}.$$

Wir berechnen das p-te Mittel von $\operatorname{disc}_p(\{t_1,\dots,t_n\})$, d.h. das Integral

$$\operatorname{av}_p(n,d) = \left(\int_{[0,1]^{nd}} \operatorname{disc}_p(t_1,\dots,t_n)^p\, dt\right)^{1/p}, \qquad t = (t_1,\dots,t_n)$$

und erhalten nach einigen Umformungen

$$\operatorname{av}_p(n,d)^p = \sum_{j=0}^p \binom{p}{j} (-n)^{-j} \sum_{k=0}^j (k+p-j+1)^{-d} \cdot \#(j,k,n). \tag{12}$$

Hierbei sei $\#(j,k,n)$ die Anzahl der Abbildungen von $\{1,\dots,j\}$ nach $\{1,\dots,n\}$ mit der Eigenschaft, daß das Bild genau k Elemente enthält. Diese Zahlen sind in der Kombinatorik gut bekannt, man kann sie auch mit den Stirling-Zahlen $s(k,i)$ erster Art und den Stirling-Zahlen $S(j,k)$ zweiter Art ausdrücken, es gilt

$$\#(j,k,n) = \sum_{i=0}^k s(k,i)\, S(j,k)\, n^i = \sum_{i=0}^k \binom{n}{k}\binom{k}{i}(-1)^i (k-i)^j.$$

Durch Umordnen der Terme folgt, daß $\operatorname{av}_p(n,d)^p$ ein Polynom in n^{-1} ist,

$$\operatorname{av}_p(n,d)^p = \sum_{r=0}^p C(r,p,d) \cdot n^{-r},$$

[7] Stefan Heinrich hat (im Juni 1999) die obere Schranke noch verbessert, seine Schranke ist von der Form $C\,\varepsilon^{-2} d\,(\log d - \log \varepsilon)$.

wobei $C(r,p,d)$ gegeben ist durch den Ausdruck

$$(-1)^r \sum_{i=0}^{p-r} \binom{p}{r+i} (-1)^i \sum_{k=i}^{i+r} (p+1-r+k-i)^{-d} s(k,i)\, S(i+r,k).$$

Aus den Definitionen folgt, daß $C(p,p,d)$ verschwindet. Wir zeigen sogar, daß

$$C(r,p,d) = 0 \quad \text{für alle} \quad r = 0,\ldots,p/2-1.$$

Dazu benötigen wir das folgende Lemma.

Lemma 14. *Es sei $p = 2m$ eine ganze natürliche Zahl. Dann gilt für $r = 0,\ldots,m-1$ und $j = 0,\ldots,r$*

$$\sum_{k=0}^{p-r+j} \binom{p}{r+k-j} (-1)^k\, s(k,k-j)\, S(k-j+r,k) = 0.$$

Obwohl diese Aussage einen rein kombinatorischen Charakter hat, haben wir keinen einfachen (Induktions-) Beweis gefunden. Wir beweisen das Lemma mit Hilfe des Satzes vom iterierten Logarithmus in der Fassung von [32]. Damit kann man dann die folgende obere Schranke für $\mathrm{av}_p(n,d)$ beweisen.

Lemma 15. *Für gerades p und $d \in \mathbf{N}$ gilt*

$$\mathrm{av}_p(n,d)^p = \sum_{r=p/2}^{p-1} C(r,p,d) \cdot n^{-r}.$$

Außerdem gilt

$$|C(r,p,d)| \le \frac{(r+1)(4p)^p}{(p+1-r)^d},$$

und

$$\mathrm{av}_p(n,d) \le 4\sqrt{2}\, p\,(1+p/2)^{-d/p}\, n^{-1/2} \left(\sum_{i=0}^{p/2-1} n^{-i} \left(\frac{1+p/2}{1+p/2\,-\,i} \right)^d \right)^{1/p}.$$

2. Teil: Im allgemeinen ist $\|f\|_\infty$ nicht nach oben beschränkt durch eine Funktion von $\|f\|_p$. Man kann sich aber überlegen, daß im Fall der Diskrepanzfunktion solche Ungleichungen gelten, es ergibt sich folgendes.

Zunächst betrachten wir eine Diskrepanz mit optimalem Gewicht, wir definieren für $(t_1,\ldots,t_n) \in [0,h]^{dn}$

$$d_\infty(t_1,\ldots,t_n) = \inf_{c>0} \sup_{x\in[0,1]^d} \left| x_1\cdots x_d - c \cdot \sum_{i=1}^{n} 1_{[0,x)}(t_i) \right|.$$

Wir beweisen Schranken für

$$\operatorname{disc}(n,d) := \inf_{(t_1,\ldots,t_n)\in[0,1]^{nd}} d_\infty(t_1,\ldots,t_n)$$

und zeigen, daß die optimalen Gewichte c anstelle von $1/n$, die wir aus beweistechnischen Gründen benötigen, nur eine kleine Verbesserung bringen. Definiere

$$A_p^d(\varepsilon) = \int_{[(1-\varepsilon)^{1/d}h,\, h)} (\varepsilon \cdot h^d - h^d + y_1 \cdots y_d)^p \, dy$$

und

$$B_p^d(\varepsilon) = \int_{[h,1)} (1 - y_1 \cdots y_d)^p \, dy,$$

wobei $h = (1+\varepsilon)^{-1/d}$. Mit diesen Größen erhält man folgendes.

Lemma 16. *Sei $d \in \mathbf{N}$ und $\varepsilon = \operatorname{disc}(n,d) \in (0,1)$ und $h = (1+\varepsilon)^{-1/d}$. Dann gilt für alle geraden p*

$$\operatorname{av}_p(n,d) \geq \min(A_p^d(\varepsilon), B_p^d(\varepsilon))^{1/p}.$$

Jetzt muß man noch untere Schranken für die Größen $A_p^d(\varepsilon)$ und $B_p^d(\varepsilon)$ zeigen und dann p optimal (in Abhängigkeit von ε und d) wählen, um den Beweis von Satz 13 zuende zu führen. □

Bemerkung 17. Die Beweisskizze zeigt, daß wir die *probabilistische Methode*, d.h. Mittelwertargumente, benutzen. Dieses Beweismittel liefert hier sehr gute Ergebnisse. Es ist jetzt eine große Herausforderung, ähnlich gute obere Schranken mit *konstruktiven* Methoden zu erhalten.

Wir haben numerische Experimente gemacht, wobei wir die exakte Formel (12) für $\operatorname{av}_p(n,d)$ und Lemma 16 benützt haben. Für $\varepsilon = 1/3$ erhält man z.B.

$$\begin{aligned}
\operatorname{card}(1/3,\ 5) &\leq 148,\\
\operatorname{card}(1/3,10) &\leq 310,\\
\operatorname{card}(1/3,20) &\leq 637,\\
\operatorname{card}(1/3,40) &\leq 1293.
\end{aligned}$$

Diese Ergebnisse sind erstaunlich wenn man bedenkt, daß vor kurzer Zeit noch unklar war, ob 3^d Punkte für einen Fehler $\varepsilon = 1/3$ ausreichen. □

Bemerkung 18. Man kann jetzt fragen, ob $\operatorname{card}(\varepsilon, d)$ für festes ε überhaupt von der Dimension d abhängt. In der Arbeit [56] werden auch untere Schranken bewiesen. Für $\varepsilon \geq 1/2$ genügen die zwei Punkte $t_1 = (0,\ldots,0)$ und $t_2 = (1,\ldots,1)$, da

$$\operatorname{disc}^*(\{t_1,t_2\}) = 1/2.$$

Für jedes $\varepsilon < 1/2$ gilt dagegen

$$\operatorname{card}(\varepsilon, d) \geq C_\varepsilon \log d,$$

d.h. die Anzahl der Punkte muß mit d wachsen, und zwar mindestens wie $\log d$ und höchstens wie $d \log d$. $\square$

Bemerkung 19. In den Anwendungen hat man es oft mit Funktionen f zu tun, die von vielen Variablen abhängen, d ist groß. Häufig ist es jedoch so, daß f von manchen Variablen "stark" abhängt und von anderen nur "schwach". Klassische Verfahren können so eine a priori Information kaum ausnützen, es gibt höchstens die Möglichkeit, manche Variablen völlig zu ignorieren. Besser ist es, gewichtete Normen zu studieren, siehe die Arbeiten [30, 47, 60, 73, 89, 99].

Wir wollen hier nur eine *gewichtete Version der L_2-Diskrepanz* diskutieren, analog zu den Räumen (6) und zum Satz 7. Wir betrachten Funktionen mit d Variablen, das Gewicht der i-ten Variablen sei w_i, es gelte

$$w_1 \geq w_2 \geq \ldots > 0.$$

Für $d = 1$ betrachten wir

$$X_1 = \{f : [0,1] \to \mathbf{R} \mid f(1) = 0,\ f \text{ ist abs. stetig},\ f' \in L_2\}$$

mit der Norm

$$\|f\|_w^2 = f(0)^2 + w^{-1}\, \|f'\|_{L_2}^2$$

und bezeichnen diesen Raum mit $X_{1,w}$. Für $d > 1$ betrachten wir dann das Tensorprodukt

$$X_d = X_{1,w_1} \otimes \ldots \otimes X_{1,w_d}$$

mit der Norm

$$\|f_1 \otimes \ldots \otimes f_d\|_{X_d} = \|f_1\|_{X_{1,w_1}} \cdots \|f_d\|_{X_{1,w_d}}.$$

Es sei wieder F_d die Einheitskugel von X_d und wir betrachten das Integrationsproblem

$$S_d(f) = \int_{[0,1]^d} f(x)\, dx.$$

Wir wissen schon (Satz 9), daß dieses Problem im Fall *gleicher Gewichte* intractable ist. Allgemeiner gilt:

- Es ist tractable genau dann, wenn $\limsup_{d\to\infty} \sum_{j=1}^d w_j / \log d < \infty$, siehe [60, 73].

- Für die Komplexität gilt eine dimensionsunabhängige obere Schranke der Form

  $$\mathrm{card}\,(\varepsilon, d) \le C \cdot \varepsilon^{-2}$$

  genau dann, wenn $\sum_{j=1}^{\infty} w_j < \infty$, siehe [60, 73].

- Gilt $\sum_{j=1}^{\infty} w_j^{1/2} < \infty$, so gilt eine obere Schranke der Form

  $$\mathrm{card}\,(\varepsilon, d) \le C_\delta \cdot \varepsilon^{-1-\delta}, \qquad \text{für } \delta > 0,$$

  siehe [30]. Alle bisher genannten oberen Schranken werden mit der *probabilistischen Methode* bewiesen, konstruktive Methoden sind bisher nicht bekannt.

- Gilt sogar $\sum_{j=1}^{\infty} w_j^{1/3} < \infty$, so kann man die obere Schranke

  $$\mathrm{card}\,(\varepsilon, d) \le C_\delta \cdot \varepsilon^{-1-\delta}, \qquad \text{für } \delta > 0$$

  konstruktiv mit einem Verfahren vom Smolyak-Typ erhalten, siehe [89].

Bei der probabilistischen Methode wird zum Beispiel der (worst case) Fehler $e(Q_n, F_d)$ von Methoden

$$Q_n(f) = \frac{1}{n} \sum_{i=1}^{n} f(t_i)$$

gemittelt über alle $(t_1, \ldots, t_n) \in [0,1]^{nd}$, entsprechend der Berechnung von $\mathrm{av}_p(n,d)$ (für $p = 2$) im Beweis von Satz 13. Es ist dann klar, daß ein Q_n *existiert*, dessen Fehler höchstens so groß ist wie dieser Mittelwert. □

7 Konstruktion polynomialer Algorithmen

Die bekannten *Konstruktionen* von polynomialen Algorithmen beruhen auf der Methode von Smolyak (1963) beziehungsweise auf verschiedenen Modifikationen dieser Methode, siehe auch [3, 11, 13, 15, 19, 20, 22, 23, 24, 49–54, 62, 66, 67, 77, 78, 79, 81, 86, 89, 93]. Selbst diese lange Liste ist sehr unvollständig, weil ähnliche Methoden (unter Bezeichnungen wie *Blending-Methode* oder *dünne Gitter-Methode* oder *Boolesche Methode* oder auch *hyperbolic cross points method*) in verschiedenem Zusammenhang studiert wurden.

Wir wollen die Methode für die numerische Integration

$$S_d(f) = \int_{[-1,1]^d} f(x)\, dx$$

präsentieren, obwohl in den genannten Arbeiten auch allgemeinere Probleme studiert werden wie die Interpolation (und optimale Rekonstruktion) und die Lösung von Operatorgleichungen. Gegeben seien für $d = 1$ Quadraturformeln[8]

$$U^i(f) = \sum_{j=1}^{m_i} a_j^i \, f(t_j^i).$$

Dabei sei $m_{i+1} > m_i$ und zum Beispiel

(13) $\quad m_1 = 1 \quad$ und $\quad m_i = 2^{i-1} + 1 \quad$ für $i > 1$.

Für $d > 1$ sind *Tensorproduktmethoden* gegeben durch

$$(U^{i_1} \otimes \cdots \otimes U^{i_d})(f) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} a_{j_1}^{i_1} \cdots a_{j_d}^{i_d} \; f(t_{j_1}^{i_1}, \ldots, t_{j_d}^{i_d}).$$

Man benötigt zur Auswertung dieser Formel

$$n = \prod_{\ell=1}^{d} m_{i_\ell}$$

Funktionswerte. Die Methode von Smolyak ist eine geschickte Linearkombination von solchen Tensorprodukten, mit der man das Ziel verfolgt, nur wenige Funktionswerte benützen zu müssen, insbesondere wenn d groß ist. Sie ist gegeben durch

$$A(q,d) = \sum_{q-d+1 \le |i| \le q} (-1)^{q-|i|} \cdot \binom{d-1}{q-|i|} \cdot (U^{i_1} \otimes \cdots \otimes U^{i_d}),$$

wobei $q \geq d$. Wählt man speziell die Clenshaw-Curtis-Formeln

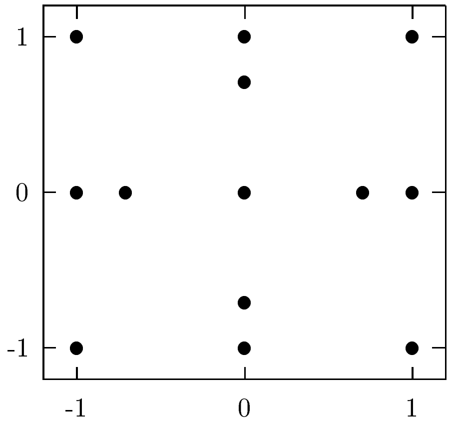
$$U^i(f) = \sum_{j=1}^{m_i} a_j^i \, f(t_j^i)$$

mit (13) und den Stützstellen

$$t_j^i = -\cos \frac{\pi(j-1)}{m_i - 1}$$

(bzw. $t_1^1 = 0$), so gelangt man zu der Methode, die in [50] und in verschiedenen anderen Arbeiten untersucht wurde. Man benützt hier für $d = 1$ nicht-äquidistante Stützstellen und die Stützstellen von U^i werden von U^{i+1} mitbenützt. Beide Eigenschaften wirken sich günstig auf die Fehlerabschätzungen und auf die Stabilität der Methode aus. Wir zeigen die Stützstellen von $A(4, 2)$.

[8] Die a_j^i sind also reelle Zahlen. Wenn man Interpolationsprobleme oder Operatorgleichungen lösen will, dann kann man ähnlich vorgehen, die a_j^i sind dann Funktionen.



$A(4,2)$, 13 Stützstellen

Satz 20. *Die $A(q,d)$ sind für viele Klassen von Funktionen fast optimal, siehe [50, 51]. Für die in (6) und in (9) definierten Räume gilt*

$$e(A(q,d),F_d) \preceq n^{-1} \cdot (\log n)^{2(d-1)},$$

wobei n die Anzahl der Stützstellen von $A(q,d)$ ist. Die Gewichte von $A(q,d)$ sind relativ klein, es gilt $\|A(q,d)\|_\infty \preceq (\log n)^{d-1}$. Die Formel $A(k+d,d)$ ist exakt für Polynome vom Grad $2k+1$ und benötigt für großes d nur $n \approx 2^k\, d^k/k!$ Stützstellen, siehe [52]. □

Bemerkung 21. Diese Methode ist *konstruktiv* in einem sehr starken Sinn. Die Gesamtrechenzeit ist in jeder Dimension proportional zur Anzahl der Stützstellen, die zusätzlichen Rechnungen (z.B. zur Berechnung der Binomialkoeffizienten) fallen kaum ins Gewicht.

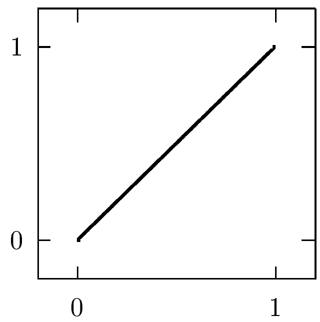
Es gibt auch "konstruktive" obere Schranken, zum Beispiel konstruktive Versionen des Satzes von Tchakalov[9], die so aufwendig sind, daß sie (ab etwa $d = 4$) praktisch nicht durchgeführt werden können. □

Bemerkung 22. Explizite Fehlerabschätzungen (die keine unbekannten Konstanten enthalten) für Smolyak-Konstruktionen werden in [86] bewiesen. Der *Witz* an diesen Konstruktionen und Fehlerabschätzungen besteht darin, daß man lediglich gute Methoden (und Fehlerabschätzungen) für den eindimensionalen Fall $d = 1$ benötigt. Alles andere liefert dann die Smolyak-Konstruktion

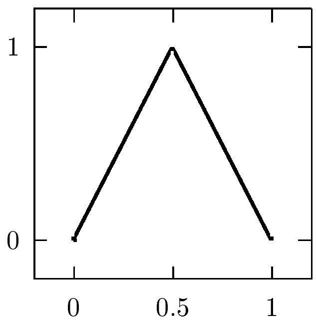
[9] Dieser Satz besagt folgendes. Sei $V(d,k)$ der Raum der Polynome in d Variablen vom Grad maximal k. Dann existiert eine Quadraturformel mit positiven Gewichten, die auf $V(d,k)$ exakt ist, wobei die Anzahl der Stützstellen höchstens so groß ist wie die Dimension von $V(d,k)$. Für eine konstruktive Version siehe [14].

und die entsprechenden *allgemeingültigen* Fehlerabschätzungen für *Tensorproduktprobleme*. In verschiedenen Arbeiten wird die Methode modifiziert, siehe die eingangs aufgezählten Arbeiten. □

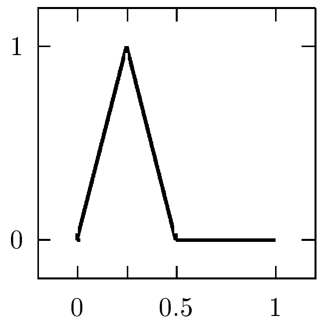
Zum Abschluß dieser Arbeit diskutieren wir die Berechnung von *Wiener-Integralen*, wiederum mit einer Methode, die Ideen von Smolyak aufgreift und modifiziert, siehe [54, 79]. Wir beginnen mit der *Lévy-Ciesielski-Zerlegung* des Wiener-Maßes mit Hilfe von *Schauder- oder Hut-Funktionen*.



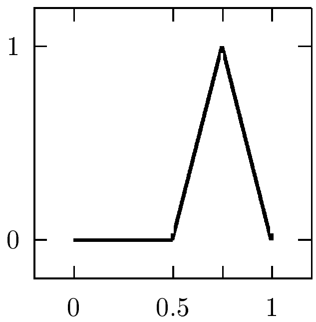
$S_{1,1}$



$S_{2,1}$



$S_{3,1}$



$S_{3,2}$

Die Schauder-Funktionen $S_{j,k}$ sind definiert für $k = 1, \ldots, m(j)$, wobei $m(j) = 2^{j-2}$ für $j \geq 2$ und $m(1) = 1$. Betrachte standardnormalverteilte unabhängige Zufallsvariablen $Y_{j,k}$ und Varianzen $\sigma^2(1) = 1,\ \ \sigma^2(j) = 2^{-j}$ für $j \geq 2$. Dann sei für $d = 2^{s-1}$

$$W^{(d)} = \sum_{j=1}^{s} \sigma(j) \sum_{k=1}^{m(j)} Y_{j,k} \cdot S_{j,k}$$

die s-te Lévy-Ciesielski-Approximation an das Wiener-Maß. Für $s \to \infty$ würde man schwache Konvergenz gegen das Wiener-Maß erhalten. Wir approximieren nun Wiener-Integrale

$$E(f) = \int_{C[0,1]} f(x)\, dW(x)$$

durch die endlichdimensionalen Erwartungswerte oder Integrale $E(f(W^{(d)}))$. Dabei kann man die Ideen von Bemerkung 19 benützen, weil der Einfluß der Variablen $Y_{j,k}$ mit steigendem j fällt.

Wir benützen eine Modifikation der Smolyak-Methode von der Form

$$A(q,d) = \sum_{q-d+1 \le |i| \le q} (-1)^{q-|i|} \cdot \binom{d-1}{q-|i|} \cdot (U^{i_1-\beta_1} \otimes \cdots \otimes U^{i_d-\beta_d}),$$

wobei die $\beta_j \in \mathbf{N}_0$ so gewählt werden, daß die Anzahl der Stützstellen von $U^{i-\beta_j}$ ungefähr $\sigma(j)$ mal so groß ist wie die Anzahl der Stützstellen von U^i. Schließlich definieren wir Quadraturformeln für das Wiener-Maß durch

$$A(s)(f) = A(s-1+d,\, d)(f \circ W^{(d)}).$$

Bemerkung 23. Man berechnet Wiener-Integrale meist mit *Monte-Carlo-Methoden* oder auch mit zahlentheoretischen Methoden, sog. *Quasi-Monte-Carlo-Methoden*, bei denen Folgen kleiner Diskrepanz benützt werden. Diese Methoden sind recht zuverlässig, aber langsam. Insbesondere können sie die (evtl. vorhandene) Glattheit des Integranden nur schlecht ausnützen.

Bei genügend glatten Integranden ist die neue Methode überlegen, in [79] werden *worst case* Fehlerabschätzungen für eine Klasse von analytischen Funktionen bewiesen. Wir zeigen Rechenergebnisse von Achim Steinbauer (Erlangen) für den Test-Integranden

$$f(x) = \cos\Big(\int_0^1 x(t)\, dt\Big).$$

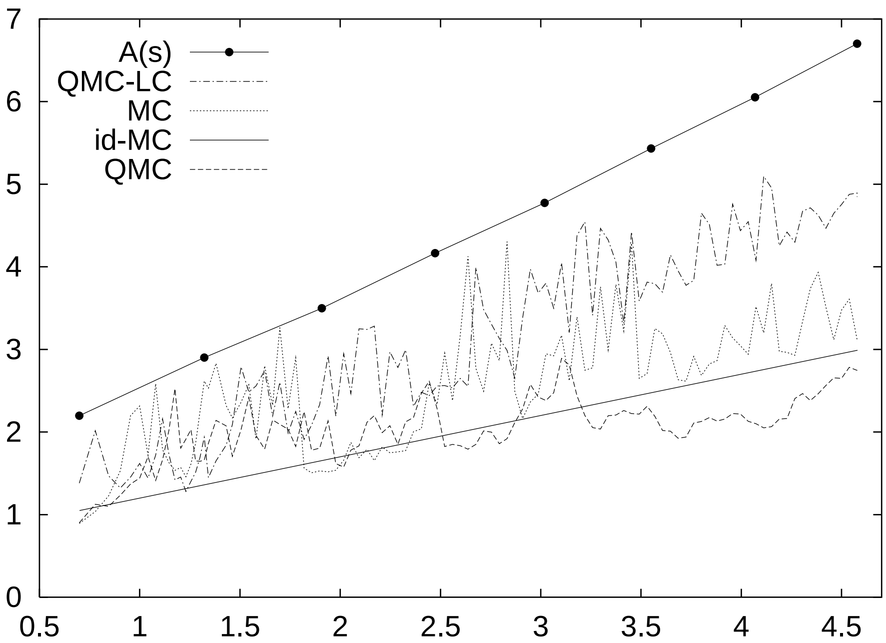
Verglichen wird das neue Verfahren $A(s)$ mit der Monte-Carlo-Methode (MC) und der idealisierten Monte-Carlo-Methode (durchgezogene Linie)[10], außerdem mit einer Quasi-Monte-Carlo-Methode (QMC), bei der mit Hilfe der Sobol-Folge Pfade in naiver Weise erzeugt werden. Schließlich bezeichnet QMC-LC eine Quasi-Monte-Carlo-Methode, bei der die Pfade mit Hilfe der Lévy-Ciesielski-Zerlegung erzeugt werden, siehe auch [40]. Gezeigt wird jeweils ein "typisches" Rechenergebnis. Das tatsächliche Ergebnis hängt bei den Methoden MC, QMC und QMC-LC vom Startwert bei den Zufallszahlen ab, es kann daher auch besser oder schlechter sein.

[10] Hier wird als Fehler die Streuung aufgetragen.

Abgetragen ist die Anzahl der gültigen Ziffern, d.h. die Größe

$$-\log_{10} \frac{|S(f) - S_n(f)|}{|S(f)|}$$

im Verhältnis zu $\log_{10} n$, wobei n die Anzahl der Pfade (=Stützstellen) ist, $S(f)$ der wahre Erwartungswert und $S_n(f)$ der berechnete Näherungswert. □



8 Schlußbemerkung

Ich habe manche Probleme nur angedeutet und oft nur wichtige Spezialfälle von allgemeineren Resultaten genannt. Für Interessierte habe ich deshalb viele Anregungen für die weitere Lektüre gegeben. Besonders hinweisen möchte ich noch auf die aktuelle Einführung [84]. Sie enthält sehr viele motivierende Beispiele, Anregungen und offene Probleme, allerdings kaum Beweise.

Ich danke meinen Freunden und Kollegen Klaus Meer, Knut Petras, Klaus Ritter, Achim Steinbauer und Henryk Woźniakowski. Sie haben eine erste Version der Arbeit gelesen und Hinweise zur Verbesserung gegeben. Besonders danke ich Herrn Triebel für die Ermunterung zu dieser Arbeit.

Literatur

[1] N. S. Bakhvalov (1959): On approximate computation of integrals. Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem. **4**, 3–18. [Russisch]

[2] N. S. Bakhvalov (1971): On the optimality of linear methods for operator approximation in convex classes of functions. USSR Comput. Maths. Math. Phys. **11**, 244–249.

[3] V. Barthelmann, E. Novak, K. Ritter (1998): High dimensional polynomial interpolation on sparse grids. Adv. in Comput. Math., erscheint.

[4] J. Beck, W. Chen (1987): *Irregularities of Distribution.* Cambridge University Press.

[5] J. Beck, V. T. Sós (1995): Discrepancy theory. In: Handbook of Combinatorics, R. Graham, M. Grötschel, L. Lovász, eds., Elsevier Science.

[6] R. Bellman (1957): *Dynamic Programming.* Princeton University Press, Princeton NJ.

[7] L. Blum, M. Shub, S. Smale (1989): On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines. Bull. of the AMS **21**, 1–46.

[8] L. Blum, F. Cucker, M. Shub, S. Smale (1998): *Complexity and Real Computation.* Springer, New York.

[9] C. G. E. Boender, H. E. Romeijn (1995): Stochastic methods. In: Handbook of global optimization. R. Horst, P. M. Pardalos (eds.), Kluwer, pp. 829–869.

[10] V. Brattka, P. Hertling (1998): Feasible real random access maschines. J. Complexity **14**, 490–526.

[11] H.-J. Bungartz, M. Griebel (1999): A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives. J. Complexity, erscheint.

[12] P. Bürgisser, M. Clausen, A. Shokrollahi (1998): *Algebraic Complexity Theory.* Springer-Verlag.

[13] R. Cools, E. Novak, K. Ritter (1999): Smolyak's construction of cubature formulas of arbitrary trigonometric degree. Computing **62**, 147–162.

[14] P. J. Davis (1967): A construction of nonnegative approximate quadratures. Math. Comp. **21**, 578–582.

[15] F.-J. Delvos, W. Schempp (1989): *Boolean Methods in Interpolation and Approximation.* Pitman Research Notes in Mathematics Series **230**, Longman, Essex.

[16] R. A. DeVore (1998): Nonlinear approximation. Acta Numerica **7**, 51–150.

[17] M. Drmota, R. F. Tichy (1997): *Sequences, Discrepancies and Applications.* Lecture Notes in Mathematics **1651**, Springer.

[18] E. G. D'yakonov (1996): *Optimization in Solving Elliptic Problems.* CRC Press, Boca Raton.

[19] K. Frank, S. Heinrich (1996): Computing discrepancies of Smolyak quadrature rules. J. Complexity **12**, 287–314.

[20] K. Frank, S. Heinrich, S. Pereverzev (1996): Information complexity of multivariate Fredholm integral equations in Sobolev classes. J. Complexity **12** (1996), 17–34.

[21] K. K. Frolov (1980): Upper bounds on the discrepancy in L_p, $2 \leq p < \infty$. Dokl. Akad. Nauk SSSR, **252**, 805–807.

[22] A. C. Genz (1986): Fully symmetric interpolatory rules for multiple integrals. SIAM J. Numer. Anal. **23**, 1273–1283.

[23] Th. Gerstner, M. Griebel (1998): Numerical integration using sparse grids. Num. Algorithms **18**, 209–232.

[24] M. Griebel, M. Schneider, Ch. Zenger (1992): A combination technique for the solution of sparse grid problems. In: Iterative Methods in Linear Algebra, Beauwens R. and de Groen P. (eds.), Elsevier, North-Holland, pp. 263–281.

[25] S. Heinrich (1993): Random approximation in numerical analysis. In Proc. of the Functional Analysis Conf., Essen 1991, Lecture Notes in Pure and Applied Math. (K. D. Bierstedt et al., eds.), vol 150, Marcel Dekker, New York, pp. 123–171.

[26] S. Heinrich (1998): Monte Carlo complexity of global solution of integral equations. J. Complexity **14**, 151–175.

[27] S. Heinrich, J.-D. Kern (1991): Parallel information-based complexity. J. Complexity **7**, 339–370.

[28] P. Hertling (1996): Topological complexity with continuous operations. J. Complexity **12**, 315–338.

[29] F. J. Hickernell (1998): A generalized discrepancy and quadrature error bound. Math. Comp. **67**, 299–322.

[30] F. J. Hickernell, H. Woźniakowski (1999): Integration and approximation in arbitrary dimension. Preprint.

[31] C. Katscher, E. Novak, K. Petras (1996): Quadrature formulas for multivariate convex functions. J. Complexity **12**, 5–16.

[32] J. Kiefer (1961): On large deviations of the empiric d.f. of vector chance variables and a law of the interated logarithm. Pacific J. Math. **11**, 649–660.

[33] M. A. Kon, E. Novak (1990): The adaption problem for approximating linear operators. Bull. of the AMS **23**, 159–165.

[34] G. Larcher (1998): Digital point sets: analysis and application. In: P. Hellekalek, G. Larcher, eds., Lecture Notes in Statistics **138**, 167–222. Springer Verlag.

[35] G. G. Lorentz, M. v. Golitschek, Y. Makovoz (1996): *Constructive Approximation. Advanced Problems.* Springer Grundlehren 304, Springer Verlag.

[36] P. Mathé (1990): s-numbers in information-based complexity, J. Complexity **6**, 41–66.

[37] J. Matoušek (1998): The exponent of discrepancy is at least 1.0669. J. Complexity **14**, 448–453.

[38] K. Meer, C. Michaux (1997): A survey on real structural complexity theory. Bull. Belg. Math. Soc. **4**, 113–148.

[39] J. M. Melenk (1999): On n-widths for elliptic problems. Preprint der ETH-Zürich.

[40] W. J. Morokoff (1998): Generating quasi-random paths for stochastic processes. SIAM Review **40**, 765–788.

[41] A. Neumayer (1997): Molecular modeling of proteins and mathematical prediction of protein structure. SIAM Review **39**, 407–460.

[42] H. Niederreiter (1992): *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia.

[43] H. Niederreiter, C. Xing (1998): Nets, (t, s)-sequences, and algebraic geometry. In: P. Hellekalek, G. Larcher, eds., Lecture Notes in Statistics **138**, 267–302. Springer Verlag.

[44] E. Novak (1988): *Deterministic and Stochastic Error Bounds in Numerical Analysis.* Lecture Notes in Mathematics **1349**, Springer.

[45] E. Novak (1995): The real number model in numerical analysis. J. Complexity **11**, 57–73.

[46] E. Novak (1996): On the power of adaption. J. Complexity **12**, 199–237.

[47] E. Novak (1999): Intractability results for positive quadrature formulas and extremal problems for trigonometric polynomials. J. Complexity, erscheint.

[48] E. Novak (1999): Is there a curse of dimension for integration? Eleventh Int. Conf. on Domain Decomposition Methods. C.-H. Lai, P. E. Bjørstad, M. Cross, O. B. Widlund, eds., erscheint.

[49] E. Novak, K. Ritter (1996): Global optimization using hyperbolic cross points. In: State of the art in global optimization: computational methods and applications. C. A. Floudas, P. M. Pardalos (eds.), Kluwer, Dordrecht, 19–33.

[50] E. Novak, K. Ritter (1996): High dimensional integration of smooth functions over cubes. Numer. Math. **75**, 79–97.

[51] E. Novak, K. Ritter (1997): The curse of dimension and a universal method for numerical integration. Multivariate approximation and Splines, G. Nürnberger, J. W. Schmidt, G. Walz (eds.), ISNM **125**, Birkhäuser, 1997, pp. 177–188.

[52] E. Novak, K. Ritter (1997): Simple cubature formulas with high polynomial exactness. Constructive Approximation, erscheint.

[53] E. Novak, K. Ritter, R. Schmitt, A. Steinbauer (1997): On an interpolatory method for high dimensional integration. J. Comput. Appl. Math., erscheint.

[54] E. Novak, K. Ritter, A. Steinbauer (1998): A multiscale method for the evaluation of Wiener integrals. Approximation Theory IX, Volume 2: C. K. Chui, L. L. Schumaker (eds.), 251–258.

[55] E. Novak, I. H. Sloan, H. Woźniakowski (1997): Tractability of tensor product linear operators. J. Complexity **13**, 387–418.

[56] E. Novak, G. W. Wasilkowski, H. Woźniakowski (1999): Polynomial upper bounds for the star-discrepancy. Preprint.

[57] E. Novak, H. Woźniakowski (1996): Topological complexity of zero finding. J. Complexity **12**, 380–400.

[58] E. Novak, H. Woźniakowski (1998): Complexity of linear problems with a fixed output basis. J. Complexity, erscheint.

[59] E. Novak, H. Woźniakowski (1999): On the cost of uniform and nonuniform algorithms. Theoretical Computer Science **219**, 301–318.

[60] E. Novak, H. Woźniakowski (1999): Intractability results for integration and discrepancy. Preprint.

[61] S. V. Pereverzev (1996): *Optimization of Methods for Approximate Solution of Operator Equations*. Nova Science, New York.

[62] K. Petras (1999): On the Smolyak cubature error for analytic functions. Preprint.

[63] A. Pietsch (1987): *Eigenvalues and s-Numbers*. Cambridge University Press. Cambridge.

[64] A. Pinkus (1985): *n-Widths in Approximation Theory*. Springer, Berlin.

[65] A. Pinkus (1986): n-widths and optimal recovery. In: Approximation Theory, C. de Boor, ed., Proc. of Symp. in Applied Math. **36**, AMS, 51–66.

[66] L. Plaskota (1999): The exponent of discrepancy of sparse grids is at least 2.1933. Preprint.

[67] L. Plaskota, G. W. Wasilkowski, H. Woźniakowski (1999): A new algorithm and worst case complexity for Feynman-Kac path integration. Preprint.

[68] K. Ritter (1999): *Average Case Analysis of Numerical Problems*. Erscheint in der Reihe Lecture Notes in Mathematics des Springer Verlags.

[69] K. F. Roth (1954): On irregularities of distributions. Mathematika **1**, 73–79.

[70] K. F. Roth (1980): On irregularities of distributions IV. Acta Arithm. **37**, 67–75.

[71] S. Schäffler (1995): Unconstrained global optimization using stochastic integral equations. Optimization **35**, 43–60.

[72] I. H. Sloan, H. Woźniakowski (1997): An intractability result for multiple integration. Math. Comp. **66**, 1119–1124.

[73] I. H. Sloan, H. Woźniakowski (1998): When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? J. Complexity, **14**, 1-33.

[74] S. Smale (1987): On the topology of algorithms. J. Complexity **3**, 81-89.

[75] S. Smale (1990): Some remarks on the foundations of numerical analysis. SIAM Review **32**, 211–220.

[76] S. Smale (1998): Mathematical problems for the next century. Math. Intell. **20**, 7–15.

[77] S. A. Smolyak (1963): Quadrature and interpolation formulas for tensor products of certain classes of functions. Sov. Math. Doklady **4**, 240–243.

[78] F. Sprengel (1997): Interpolation and wavelets on sparse Gauss-Chebyshev grids. In: Multivariate Approximation, W. Haussmann et al., eds., Math. Res. **101**, pp. 269-286, Akademie Verlag, Berlin.

[79] A. Steinbauer (1999): Quadrature formulas for the Wiener measure. J. Complexity, erscheint.

[80] V. N. Temlyakov (1994): *Approximation of Periodic Functions*. Nova Science, New York.

[81] S. Tezuka (1998): Financial applications of Monte Carlo and Quasi-Monte Carlo methods. In: Random and Quasi-random point sets. P. Hellekalek, G. Larcher (eds.), Lecture Notes in Statistics **138**, Springer-Verlag.

[82] E. Thiémard (1999): Computing bounds for the star discrepancy. Preprint.

[83] J. F. Traub, G. W. Wasilkowski, H. Woźniakowski, *Information-Based Complexity*. Academic Press, Boston, 1988.

[84] J. F. Traub, A. G. Werschulz, *Complexity and Information.* Cambridge University Press, 1998.

[85] V. A. Vassiliev (1996): Topological complexity of root-finding algorithms. In: The Mathematics of Numerical Analysis, J. Renegar, M. Shub, S. Smale (eds.), Lectures in Applied Mathematics 32, AMS, pp. 831-856.

[86] G. W. Wasilkowski, H. Woźniakowski (1995): Explicit cost bounds of algorithms for multivariate tensor product problems. J. Complexity **11**, 1–56.

[87] G. W. Wasilkowski, H. Woźniakowski (1996): On tractability of path integration. J. Math. Phys. **37**, 2071–2088.

[88] G. W. Wasilkowski, H. Woźniakowski (1997): The exponent of discrepancy is at most 1.4778.... Math. Comp. **66**, 1125–1132.

[89] G. W. Wasilkowski, H. Woźniakowski (1999): Weighted tensor-product algorithms for linear multivariate problems. J. Complexity, erscheint.

[90] K. Weihrauch (1995): A simple introduction to computable analysis. Informatik-Berichte der Fernuniversität Hagen **171-2**.

[91] K. Weihrauch (1998): A refined model of computation for continuous problems. J. Complexity **14**, 102–121.

[92] A. G. Werschulz (1991): *The Computational Complexity of Differential and Integral Equations.* Oxford University Press, Oxford.

[93] A. G. Werschulz (1996): The complexity of the Poisson problem for spaces of bounded mixed derivatives. In: The Mathematics of Numerical Analysis (J. Renegar, M. Shub, S. Smale, eds.), Lect. in Appl. Math. **32**, pp. 895–914, AMS, Providence.

[94] G. Winkler (1995): *Image Analysis, Random Fields and Dynamic Monte Carlo Methods.* Springer Verlag.

[95] H. Woźniakowski (1991): Average case complexity of multivariate integration. Bull of AMS (new series) **24**, 185-191.

[96] H. Woźniakowski (1994): Tractability and strong tractability of linear multivariate problems. J. Complexity **10** (1994), 96–128.

[97] H. Woźniakowski (1994): Tractability and strong tractability of multivariate tensor product problems. J. of Computing and Information, **4**, 1–19.

[98] H. Woźniakowski (1998): Why does information-based complexity use the real number model? Theoretical Computer Science, erscheint.

[99] H. Woźniakowski (1999): Efficiency of quasi-Monte Carlo algorithms for high dimensional integrals. Erscheint in Proceedings of MCQMC conference, Claremont, USA, 1998, H. Niederreiter and J. Spanier, eds.

Erich Novak
Mathematisches Institut
Universität Erlangen-Nürnberg Bismarckstr. 1 1/2
D-91054 Erlangen
novak@mi.uni-erlangen.de

Eingegangen 1.7.1999