

**Simplex-Bereichsanfragen mit
 ϵ -Netzen**

Christian Volk

Algorithm Engineering Report

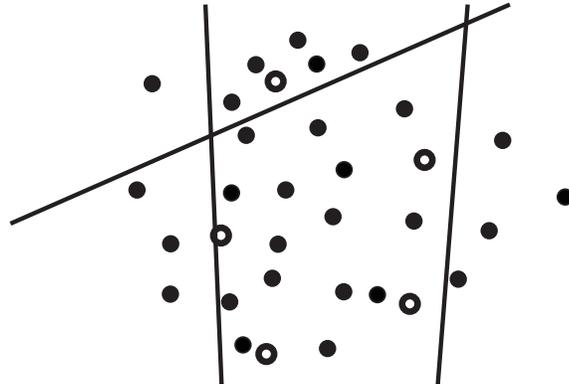
TR08-3-006

Dez. 2008

ISSN 1864-4503

Diplomarbeit zum Thema

„Simplex-Bereichsanfragen mit ε -Netzen“



zur Erlangung des akademischen Grades

Diplom-Mathematiker

vorgelegt dem Fachbereich Mathematik
der Westfälischen Wilhelms-Universität Münster

Christian Volk

30. Januar 2008

Gutachter:

Prof. Dr. Vahrenhold

Fachgebiet Grundlagen und

Vermittlung der Informatik

Lehrstuhl für Algorithm Engineering

Fakultät für Informatik

Technische Universität Dortmund

Otto-Hahn-Str. 14

44227 Dortmund

Inhaltsverzeichnis

1	Einleitung	1
1.1	Thema der Arbeit	1
1.2	Halbraum- und Simplex-Bereichsanfragen	1
1.3	Bekannte Ergebnisse zu Halbraum-Bereichsanfragen	2
1.4	Vorgehensweise zur Lösung der Problemstellung	3
1.5	Motivation der Gliederung der Diplomarbeit	6
1.6	Zusammenhang mit kombinatorischer Geometrie und Wahrscheinlichkeitsrechnung	7
2	Zufall	9
2.1	Grundlagen der Wahrscheinlichkeitsrechnung	9
2.1.1	Zufallsexperimente und Ereignisse	10
2.1.2	Wahrscheinlichkeitsräume	10
2.1.3	Zufällige, unabhängige n -Stichproben	16
2.1.4	Bernoulli-Experiment	17
2.1.5	Binomialverteilung und n -stufiges Bernoulli-Experiment	17
2.1.6	Ungleichungen für große Abweichungen (<i>tail estimates</i>)	20
2.1.7	Anwendung der Chebyshev-Ungleichung für Stichproben	21
	Theorem I: Das Stichprobentheorem	21
2.2	Randomisierte Algorithmen	22
2.2.1	Las-Vegas-Algorithmus	22
2.2.2	Monte-Carlo-Algorithmus	23
2.2.3	Nachteile randomisierter Algorithmen	23
2.3	Die probabilistische Methode	23
2.4	Derandomisierung	25
2.4.1	Bemerkung zur Praktikabilität	26
2.4.2	Anschauliche Rechtfertigung der Praktikabilität	26
2.5	Die Methode der bedingten Erwartungswerte bzw. Wahrscheinlichkeiten	27
2.5.1	Beschreibung der Methode	27

2.5.2	Zum Theorem über die maximale Wahrscheinlichkeit	28
	Theorem II über die maximale Wahrscheinlichkeit	28
2.5.3	Satz über die maximale Erwartung	29
2.5.4	Exkurs zur Erläuterung der Methode:	
	Algorithmus zum Finden von Cuts	30
3	Bereichsräume und ε-Netze	33
3.1	Bereichsräume, ε -Netze und ε -Approximationen	33
3.2	Existenz von kleinen ε -Netzen	37
3.3	Die VC-Dimension von Bereichsräumen	40
3.4	ε -Approximationen der konstanten Größe $\mathcal{O}\left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$	45
3.5	ε -Netze der konstanten Größe $\mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$	47
	Theorem III: ε-Netze der Größe $\mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$	59
4	Geometrische Bereichsräume	61
4.1	Grundlage: Arrangements von Hyperebenen	64
4.2	Grundlage: Geometrische Dualität	67
4.3	Der Bereichsraum der Halbräume des \mathbb{R}^d	68
4.4	Kombination von Bereichsräumen	70
4.5	Der Bereichsraum der k -Korridore im \mathbb{R}^d	71
4.6	Zum ε -Theorem für k -Korridore	72
	Theorem IV: Das ε-Theorem für k-Korridore	72
5	Verwendung von Korridoren	74
5.1	Geometrische Grundlage: Zonen in Arrangements	74
5.2	Formulierung des Ziels	75
5.3	Einbettung der Zone in einen geeigneten Korridor	75
5.4	Geeignete ν -Stichproben endlicher Mengen	80
	Theorem V: Das zentrale ε-Theorem	80
5.5	Zulässigkeit von Stichproben	81
6	Halbraum-Bereichsanfragen	85
6.1	Wiederholung: Partitionsbäume für Bereichsanfragen	85
6.2	Der (ε, ν) -Baum: ein Partitionsbaum mit Parametern	86
6.3	Existenz von (ε, ν) -Bäumen	93
6.4	Zur Existenz der Datenstruktur	93
	Theorem VI: Existenz der Datenstruktur	93
6.5	Plausibilitätsüberlegungen	94
6.6	Konstruktion der Datenstruktur	95
	Theorem VII: Konstruktion der Datenstruktur	96

6.7	Abschätzung der Konstanten	98
6.7.1	Abschätzung der Stichprobengröße	98
6.7.2	Abschätzung der Zellen im Arrangement	99
7	Simplex-Bereichsanfragen	100
7.1	Einleitung zur verwendeten Datenstruktur	100
7.2	Analyse der Parameter des (ε, ν) -Baumes	101
7.3	Der Anfragealgorithmus	101
7.4	zur Existenz der Datenstruktur	102
	Theorem VIII: Existenz der Datenstruktur	102
7.5	zur Konstruktion der Datenstruktur	104
	Theorem IX: Konstruktion der Datenstruktur	104
8	Ergänzungen	105
8.1	Größe und Konstruktion von ε -Netzen im Allgemeinen	105
8.2	ε -Netze für horizontale Segmente in Bezug auf vertikale Segmente	107
8.2.1	Ein Plane-Sweep-Algorithmus zur Konstruktion eines ε -Netzes	108
8.2.2	Deterministische Konstruktion	110
8.2.3	Realisierung des Konstruktionsalgorithmus	111
8.2.4	Laufzeit des Algorithmus	114
9	Zusammenfassung und Ausblick	116
A	Abschätzung der Rekursion	118
A.1	Einleitung: Formulierung des Abschätzungstheorems	118
	Theorem X: Das Abschätzungstheorem	119
A.2	Untersuchung von $Q(n)$	119
A.3	Untersuchung der Hilfsfunktion $\gamma: \mathbb{N} \rightarrow \mathbb{R}^+$	121
A.4	Untersuchung der Hilfsfunktion $f: (0, \infty) \rightarrow \mathbb{R}$	126
A.5	Beweis des Abschätzungstheorems	129
B	Beweisdiagramm	131

Kapitel 1

Einleitung

Wir beschreiben zuerst des Thema der Arbeit. Nach der Beschreibung der Problemstellung werden kurz bekannte Ergebnisse genannt. Daraufhin wird die Vorgehensweise erläutert und mit dieser die Gliederung der Diplomarbeit in Kapitel motiviert.

1.1 Thema der Arbeit

Diese Arbeit basiert auf dem Artikel *ε -Nets and Simplex Range Queries* von David Haussler und Emo Welzl [HW87]. Es wird für Halbraum- und Simplex-Bereichsanfragen im \mathbb{R}^d eine Datenstruktur mit $\mathcal{O}(n)$ Speicher und sublinearer Anfragezeit bestimmt. Die genaue asymptotische Laufzeit ist dabei für ein beliebiges vorgegebenes $\gamma > 0$:

$$T(n) = \mathcal{O}\left(n^{\beta+\gamma}\right), \beta = 1 - \frac{1}{d(d-1)+1}.$$

Die Dimension d des \mathbb{R}^d wird als Konstante betrachtet, so dass $\beta < 1$ fest ist. $\gamma > 0$ kann dagegen beliebig klein vorgegeben werden.

Für klein gewählte γ ist $T(n)$ damit sublinear. Bis auf einen beliebig kleinen polynomiellen Faktor n^γ ist die Anfragezeit $\mathcal{O}(n^\beta) = \mathcal{O}\left(n^{1-\frac{1}{d(d-1)+1}}\right)$. Damit ist der Algorithmus asymptotisch besser als der sogenannte triviale Scan.

1.2 Halbraum- und Simplex-Bereichsanfragen

Es wird die folgende Problemstellung betrachtet:

Bestimme zu einer Menge A aus n Punkten im \mathbb{R}^d eine Datenstruktur mit linearem Speicher derart, dass zu einem beliebigen offenen Halbraum h^* des \mathbb{R}^d die Anzahl $|A \cap h^*|$ der in diesem Halbraum enthaltenen Punkte berechnet werden kann. In der englisch-sprachigen Literatur wird diese Problemstellung als *half-space counting problem* bezeichnet. Bei der verwandten Problemstellung des *half-space range reporting* wird nicht die Anzahl, sondern die Menge $A \cap h^*$ selbst ausgegeben. Wir behandeln jedoch nur die erste Variante und

bezeichnen daher die Bestimmung von $|A \cap h^*|$ als Halbraum-Bereichsanfrage. Simplex-Bereichsanfragen seien analog die Bestimmung von $|A \cap \sigma^*|$ für ein nicht notwendig voll-dimensionales Simplex σ^* .

Vorgaben und Einschränkungen

Gegeben sei eine Menge $A = \{x_1, x_2, \dots, x_n\}$ aus n Punkten im \mathbb{R}^d . Zur einfacheren Behandlung sei die Punktmenge A in allgemeiner Lage. Das heiÙe, dass durch je d der Punkte genau eine Hyperebene verläuft, die nicht vertikal ist, und dass der Schnitt von $(d + 1)$ solchen Hyperebenen stets leer ist. Diese Annahme wird im vierten Kapitel motiviert, ab welchem Arrangements betrachtet werden. Für die Behandlung nicht-allgemeiner Lage lässt sich eine als *Simulating Simplicity* bezeichnete Technik anwenden. Diese wird bei Edelsbrunner für eine andere Problemstellung mit d -dimensionalen Arrangements durchgeführt. [Ede87, S.185ff.]

Elementare geometrische Algorithmen

Die folgenden Aussagen gehören zur Folklore der algorithmischen Geometrie:

1.2.1 Lemma (Halbraumtest). *Sei ein Punkt $p \in \mathbb{R}^d$ und ein Halbraum h^* gegeben. Dann lässt sich in $\mathcal{O}(1)$ Zeit testen, ob $p \in h^*$ gilt.*

1.2.2 Definition (trivialer Scan für Bereichsanfragen). Man betrachte eine endliche Menge $A = \{x_1, x_2, \dots, x_n\}$. Als trivialen Scan für eine Bereichsanfrage bezeichnet man einen Algorithmus, der für jeden Punkt aus der gegebenen Punktmenge A bestimmt, ob er im angegebenen Bereich enthalten ist. Als Ergebnis gibt der triviale Scan die Anzahl der enthaltenen Punkte zurück.

Es folgt:

1.2.3 Satz. *Der triviale Scan beantwortet Halbraum-Bereichsanfragen sowie Simplex-Bereichsanfragen in $\mathcal{O}(n)$ Laufzeit.*

Das Ziel ist jedoch die Bestimmung einer Datenstruktur mit linearem Speicherbedarf und *sublinearer* Anfragezeit.

1.3 Bekannte Ergebnisse zu Halbraum-Bereichsanfragen

Die erste obere Schranke für Halbebenen-Bereichsanfragen (d.h. Halbraum-Bereichsanfragen im \mathbb{R}^d , $d = 2$) mit linearem Speicher ist $\mathcal{O}(n^\alpha)$ mit $\alpha \approx 0.774$ nach Willard [Wil82], anschließend von Edelsbrunner und Welzl [EW86] zu $\alpha \approx 0.695$ verbessert. Im allgemeinen Fall des \mathbb{R}^d wurde von Yao und Yao [YY85] die erste sublineare obere Schranke gefunden. Diese obere Schranke $\mathcal{O}(n^\alpha)$ mit $\alpha = \lceil \log(2^d - 1)/d \rceil$ ist wegen $\log 2^d = d$ nur marginal kleiner als 1. Die von Haussler und Welzl gefundene Laufzeit stellt eine Verbesserung der

zum damaligen Zeitpunkt bekannten oberen Schranken für alle $d \geq 2$ dar. Es sei bemerkt, dass mittlerweile noch bessere Lösungen der Problemstellung bekannt sind. Insbesondere hat Matoušek in dem Artikel *Efficient Partition Trees* [Mat92] eine Datenstruktur mit $\mathcal{O}(n^{1-1/d}(\log n)^{\mathcal{O}(1)})$ Anfragezeit entworfen, die sich zudem in $\mathcal{O}(n \log n)$ Zeit deterministisch konstruieren lässt. Matoušek bemerkt in dem Artikel, dass die untere Schranke nach Chazelle [Cha89] damit bis auf polylogarithmische Faktoren erreicht wird.

Die von Matoušek entwickelte Datenstruktur baut auf dem Konzept der ε -Netze von Haussler und Welzl auf, geht jedoch weit darüber hinaus und wird im Rahmen dieser Diplomarbeit nicht betrachtet.

1.4 Vorgehensweise zur Lösung der Problemstellung

Die oben zitierten vor Haussler und Welzl entworfenen Lösungen basieren auf einem Partitionsbaum (vgl. [Mat92, 1. Abschnitt]). In diesem Abschnitt wird dessen Grundprinzip erläutert und anhand eines Exkurses gezeigt, dass eine direkte Verwendung eines einfachen Partitionsbaumes nicht ausreichend ist um eine sublineare Laufzeit zu erzielen.

Grundprinzip eines Partitionsbaumes

Gegeben sei eine endliche Menge $A \subset \mathbb{R}^d$. Grundidee ist die Partitionierung (Aufteilung) des \mathbb{R}^d in mehrere offene Regionen, d.h. offene, zusammenhängende Mengen. Diese werden wiederum rekursiv in Teilregionen unterteilt. Jeder Region $r(K)$ wird ein Knoten K zugeordnet. Zudem wird für unsere Anwendung jedem Knoten die Anzahl der in der Region enthaltenen Punkte als Attribut gespeichert. Für Regionen, die nur eine Anzahl von Punkten aus A enthalten, die kleiner oder gleich einer festgelegten Konstanten ist, sind die zugeordneten Knoten Blattknoten. In Blattknoten sind alle aus A enthaltenen Punkte gespeichert und können einzeln abgearbeitet werden.

Damit gilt für einen beliebigen Knoten K und dessen zugehörige Region $r(K)$ die folgende Beobachtung: Die Menge $A_K := A \cap r(K)$ der Punkte aus A , die in $r(K)$ enthalten sind, besteht aus genau jenen Punkten, die im Teilbaum mit der Wurzel K enthalten sind. Desweiteren ist die Menge der Knoten der gleichen Höhe eine Partitionierung der Punktmenge A .

Um eine Bereichsanfrage mit einem Halbraum h^* zu bearbeiten, beginnt man im Wurzelknoten, dem der gesamte \mathbb{R}^d zugeordnet ist.

Bei der Bearbeitung unterscheidet man für die Kindknoten drei Typen von zugeordneten Regionen: Die Regionen außerhalb des Anfragebereichs h^* werden verworfen. Für die komplett in h^* enthaltenen Regionen addiert man unmittelbar die gespeicherte Anzahl der enthaltenen Punkte zur Ausgabe der Bereichsanfrage. Jene Regionen, die von h^* geschnitten werden, sind genauer zu untersuchen. Für deren Knoten wird die Bereichsanfrage mit

h^* rekursiv aufgerufen, so dass ihre Teilregionen nach dem gleichen Schema bearbeitet werden.

Problematisch: Effizienz im Worst-Case

Die Hauptschwierigkeit bei dieser Vorgehensweise ist, das sogenannte Partitionsschema so zu gestalten, dass die sich ergebende Rekursionsgleichung eine gute Laufzeit garantiert.

Im Worst-Case ist dies nicht leicht zu realisieren, vgl. [Mat96, S.549 ff]. Es lassen sich durch einfache Heuristiken gute randomisierte Anfragezeiten realisieren. Die Grundproblematik ist aber, dass eine ungünstig liegende Hyperebene prinzipiell viele Regionen schneiden kann. Eine garantierte (nicht-triviale, d.h. sublineare) Anfragezeit ist dann nicht erreichbar.

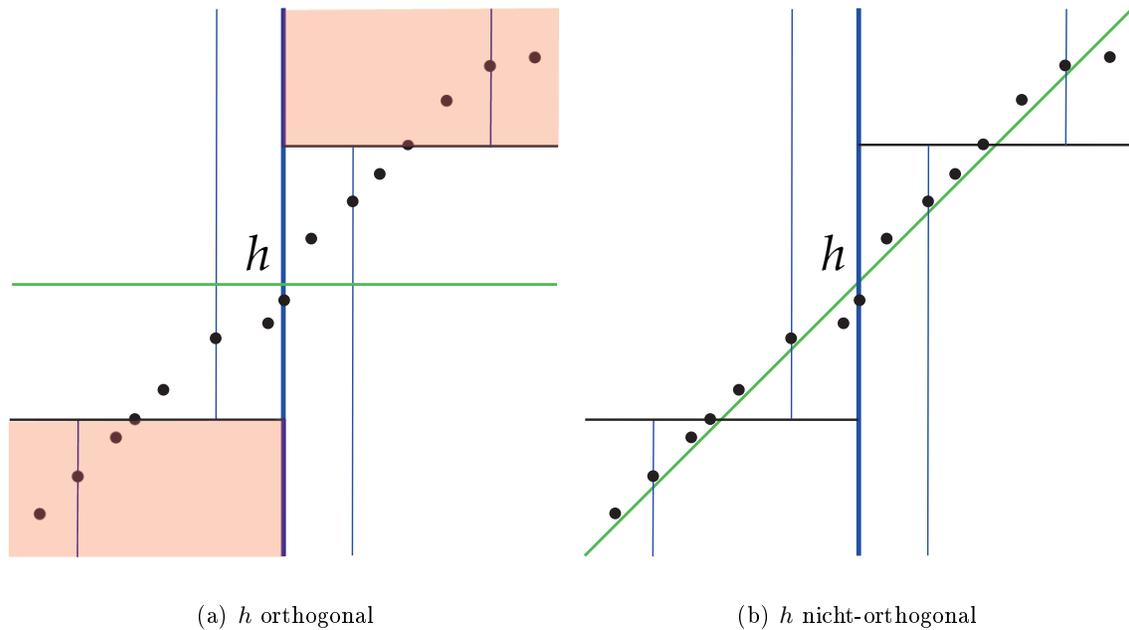
Betrachten wir als ein Beispiel einen kd -Baum.

Exkurs: kd -Baum

Insbesondere im zweidimensionalen erkennt man leicht, dass Partitionsbäume gut für orthogonale Bereichsanfragen geeignet sind. Für Bereichsanfragen in unserem Sinne, bei denen nur die Anzahl der Punkte (*range counting query*) ausgegeben werden soll, ist die Laufzeit $\mathcal{O}(n^{1-1/d})$, vgl. z.B. den Artikel über kd -Bäume von Bentley [Ben75]. Hintergrund ist, dass in jedem d -ten Level des Partitionsbaumes die Hyperebene h komplett in einer der beiden Teilregionen enthalten ist. Abbildung 1.1 zeigt ein Beispiel eines kd -Baumes (mit einer recht speziellen Punktmenge). Jene Regionen, die nicht geschnitten werden, sind in der Abbildung 1.1 (a) rot eingefärbt. Betrachtet man d Levels (in der Abbildung: 2 Levels) des Baumes, so fallen stets 2^{d-1} (in der Abbildung: 2) der 2^d (in der Abbildung: 4) rekursiven Aufrufe weg. Die nicht geschnittenen Regionen können in $\mathcal{O}(1)$ Zeit bearbeitet werden. Dazu werde in jedem Knoten K die Anzahl $|A_K|$ von Punkten der zugeordneten Region A_K gespeichert. Für komplett enthaltene Regionen ist dann nur $|A_K|$ zum Ergebnis der Bereichsanfrage zu addieren. Rekursive Aufrufe sind nur notwendig für geschnittene Regionen. Mit der sich ergebenden Rekursionsgleichung lässt sich die Laufzeit $\mathcal{O}(n^{1-1/d})$ nachweisen. Bei der im Rahmen der Diplomarbeit betrachteten Problemstellungen sind jedoch beliebige Hyperebenen zugelassen. Je nach Punktmenge kann eine ungünstig liegende Hyperebene d dabei deutlich mehr Regionen schneiden. In der Abbildung 1.1 sind die Punkte in der Nähe der Geraden h aus (b) angeordnet. Für diese Gerade h werden alle Regionen geschnitten. Damit ist die Laufzeit $\Omega(n)$. (Das gilt unabhängig davon, ob einige, viele, alle oder gar keine Punkte über der Geraden liegen). Mit dieser kurzen Untersuchung zeigt sich also, dass eine $\mathcal{O}(n)$ -Datenstruktur mit sublinearer Anfragezeit für beliebige Hyperebenen keine triviale Aufgabe ist.

Stichproben-Partitionierung

Ein entscheidender Unterschied bei Haussler und Welzl gegenüber den früheren Untersuchungen ist, dass sie eine Region auf einem Rekursionslevel nicht in eine kleine Anzahl

Abbildung 1.1: kd -Baum

von Regionen aufteilen. Sie teilen die Regionen rekursiv in eine konstante, aber sehr große Anzahl von Regionen auf und garantieren, dass bei jedem rekursiven Aufruf

1. die Anzahl der geschnittenen Regionen und
2. der Bruchteil der dort enthaltenen Punkte angemessen klein ist.

Diese beiden Eigenschaften werden durch Parameter beschrieben und eine parametrisierte Rekursionsgleichung aufgestellt. Es wird gezeigt, dass sich die Parameter so wählen lassen, dass sich eine sublineare Laufzeit ergibt.

Der verwendete Partitionsbaum: (ε, ν) -Baum

Wir bezeichnen den von Haussler und Welzl entworfenen Partitionsbaum, der diese beiden Eigenschaften garantieren soll, als (ε, ν) -Baum (im englischen Original: (ε, ν) -*partition tree*) und definieren diesen formal in Kapitel 6 auf Seite 86. Er basiert darauf, für jede Region mit einer zugehörigen Punktmenge A eine geeignete Stichprobe (Teilmenge) N aus ν Punkten zu wählen mit einer geeignet großen Konstanten ν . Um aus diesen Stichproben N eine Aufteilung der Regionen zu erzeugen, betrachtet man das Arrangement aus $H_d(N)$, der Menge der Hyperebenen durch je d der Punkte. Jede Zelle, die Punkte aus A enthält, beschreibt dann eine Region, welcher ein Kindknoten zugeordnet wird.

Obwohl die Anzahl der Zellen in einem Arrangement, welches man aus einer ν -Stichprobe erhält, mit ungefähr $\nu^{d(d-1)}$ asymptotisch sehr schnell wächst, zeigt sich: Die oben beschriebene erste Eigenschaft wird hinreichend gut für eine sublineare Laufzeit erfüllt. Man beachte, dass d als konstant betrachtet wird und weiter, dass in der Anfragezeit $T(n)$ der

Term $d(d-1)$ auch auftritt.

Rekursive Aufrufe sind dann notwendig für alle von einer Hyperebene h geschnittenen Zellen. Um garantieren zu können, dass der Bruchteil der Punkte in geschnittenen Zellen klein ist (in der späteren Analyse: höchstens ε), wird von Haussler und Welzl das Konzept der ε -Netze eingeführt.

1.5 Motivation der Gliederung der Diplomarbeit

Auf Grundlage der soeben beschriebenen Vorgehensweise wird in diesem Abschnitt die Unterteilung in Kapitel motiviert. Die mathematischen Kapitel 2 bis 5 enthalten die Voraussetzungen für die Einführung von (ε, ν) -Bäumen (Def. 6.2.1, S.86) und deren zielführende Verwendung für Halbraum-Bereichsanfragen in Kapitel 6 sowie Simplex-Bereichsanfragen in Kapitel 7.

Es zeigt sich, dass probabilistische Betrachtungen notwendig sind, um ε -Netze zu finden. Daher beschäftigt sich das zweite Kapitel mit Grundlagen der Wahrscheinlichkeitsrechnung. Im dritten Kapitel werden ε -Netze für abstrakte Bereichsräume eingeführt und Untersuchungen zu ihrer Existenz durchgeführt.

Im vierten Kapitel werden typische geometrische Bereichsräume untersucht.

Im fünften Kapitel wird gezeigt, dass ε -Netze spezieller Bereichsräume (bestehend aus sogenannten $(d+1)$ -Korridoren) in der Lage sind, die weiter oben beschriebene zweite Voraussetzung zu garantieren: Es gibt zu jeder endlichen Menge A eine geeignet kleine Teilmenge N (diese ist ein spezielles ε -Netz), so dass im Arrangement aus $H_d(N)$ gilt: Für eine beliebige Hyperebene h ist der Bruchteil der Punkte aus A , welche in geschnittenen Zellen liegen, höchstens ε . Diese Aussage wird als das zentrale ε -Theorem (Theorem V, S.80) bezeichnet. Das Theorem besagt auch, dass diese Bedingung schon durch eine zufällige Stichprobe N geeigneter Größe realisiert wird.

Im sechsten Kapitel werden auf solchen ε -Netzen basierend (ε, ν) -Bäume eingeführt und ein Algorithmus für Halbraum-Bereichsanfragen mit (ε, ν) -Bäumen (Algorithmus 6.1, S.89) untersucht. Aus den vorliegenden Parametern ergibt sich eine Rekursionsgleichung für die Laufzeit.

Es wird im Theorem VI (S.93) gezeigt, dass geeignete (ε, ν) -Bäume eine Datenstruktur mit linearem Speicherbedarf und der postulierten sublinearen Laufzeit sind. Dabei wird eine technisch sehr aufwändige Abschätzung der Rekursionsgleichung vorausgesetzt. Diese rein analytische Aussage wird als Abschätzungstheorem (Theorem X, S.119) bezeichnet und im Anhang A bewiesen.

Im Theorem VII (S.96) wird gezeigt, dass sich ein solcher (ε, ν) -Baum durch einen randomisierten Algorithmus in erwarteter Laufzeit $\mathcal{O}(n \log n)$ konstruieren lässt.

Im siebten Kapitel wird gezeigt, dass sich mit (ε, ν) -Bäumen auch Simplex-Bereichsanfragen

durchführen lassen und dabei dieselben Größenordnungen auftreten. Damit ist der Hauptteil der Diplomarbeit abgeschlossen. Es folgen im achten Kapitel einige Ergänzungen zu ε -Netzen sowie im neunten Kapitel eine kurze Zusammenfassung des Inhalts.

1.6 Zusammenhang mit kombinatorischer Geometrie und Wahrscheinlichkeitsrechnung

Ein ε -Netz, dessen Größe nicht von n bzw. nicht von der Anzahl der Punkte l in einer Region abhängt, sondern lediglich von der verlangten Genauigkeit ε , kann garantieren, dass in Bereichen ohne Punkte aus dem ε -Netz weniger als der ε -te Teil der Punkte liegt. Die Eigenschaft, ob Punkte aus dem ε -Netz in der Anfrage enthalten sind, lässt sich mit Verwendung einer geeigneten Datenstruktur in hinreichend kleiner Zeit überprüfen. Die Formulierung eines verwendbaren Bereichsraumes gelingt mit Hilfe einer Dualitätsbetrachtung. Um die Existenz geeigneter ε -Netze in diesem Bereichsraum nachzuweisen, muss gezeigt werden, dass dieser eine gewisse kombinatorische Eigenschaft erfüllt, nämlich eine endliche VC-Dimension besitzt, s. Def. 3.3.1, S.40. Dieser Nachweis erfolgt durch eine Dualitätsbetrachtung und kombinatorische Geometrie.

Da ε -Netze durch eine probabilistische Methode nachgewiesen werden, knüpfen diese einen Zusammenhang zwischen der algorithmischen Verwendung von Partitionsbäumen einerseits sowie kombinatorischer Geometrie und Wahrscheinlichkeitsrechnung andererseits. Die probabilistische Methode ist eine Beweistechnik, die im Abschnitt 2.3 (S.23) eingeführt wird.

Algorithmisches Paradigma: Randomisiertes Divide and Conquer

Um die gewünschte Datenstruktur nachzuweisen, verwenden Haussler und Welzl die Existenz geeigneter Teilmengen (Stichproben) für Punktmengen im obigen Sinne. Das ist zunächst völlig nicht-konstruktiv. Der Begriff „Stichprobe“ suggeriert, dass sich das Verfahren auch aus einem etwas anderem Blickwinkel betrachten lässt. Zur Konstruktion der Datenstruktur können *zufällige* Stichproben geeigneter Größe verwendet werden. Dann erfolgt die rekursive Aufteilung der Punktmenge in Teilmengen zufällig, d.h. randomisiert. Die Wahrscheinlichkeit, dass eine dieser zufälligen Stichproben verwendet werden kann, ist bei geeigneter Größe mindestens $1/2$ (s. Kapitel 3).

Unter diesem Blickwinkel ist das Verfahren von Haussler und Welzl ein *randomisiertes Divide and Conquer* (s. Abschnitt 2.2 oder den Artikel *Derandomization in Computational Geometry* von Matoušek [Mat96, Abschnitt 2.1]).

Da Haussler und Welzl jedoch eine deterministische Anfragezeit garantieren wollen, wird zur Konstruktion ein Las-Vegas-Algorithmus (siehe Unterabschnitt 2.2.1, S.22) verwendet. Bei der Konstruktion werden zufällige Stichproben gewählt, aber so lange verworfen, bis schließlich eine geeignete Stichprobe vorliegt. Die erwartete Laufzeit des Konstruktionsalgorithmus ist $\mathcal{O}(n \log n)$, wie in Theorem VII (S.96) gezeigt wird.

Bemerkung zu den verwendeten Bezeichnungen

Wird bei einem Logarithmus keine Basis angegeben, so ist der Logarithmus zur Basis 2 gemeint.

Die verwendeten Bezeichnungen stellen einen Kompromiss zwischen Haussler und Welzl, Matoušek, Motwani [MR95, Anhang C] sowie Edelsbrunner [Ede87] dar. Es wird eine möglichst einheitliche Bezeichnung für gleichartige Objekte verwendet: Koordinaten werden mit griechischen Kleinbuchstaben bezeichnet, elementare geometrische Objekte (Punkte, Hyperebenen und Zellen) mit lateinischen Kleinbuchstaben. Für Tupel von Punkten wird eine vektorartige Schreibweise verwendet. Desweiteren sollen Mengen von elementaren Objekten mit lateinischen Großbuchstaben bezeichnet und Systeme solcher Mengen mit kalligraphischen oder Frakturbuchstaben. Die folgende Tabelle gibt einige Beispiele für Notationen.

Objekttyp	Beispiele
Koordinate	$\chi_1, \chi_d, \pi_1, \eta_1$
Punkt	x, y, p
Hyperebene	g, h
Zelle	f, f_i
Tupel von Punkten	$\bar{x}, \bar{y}, \bar{z}$
Punktmenge	X, A, N
Menge von Hyperebenen	$H_d^+, H_d(N)$
Menge von Zellen	$F(N), Z(h, N)$
System von Mengen	$\mathcal{R}, \mathcal{P}(), \mathcal{Z}, \mathcal{F}_i$

Zu beachten ist, dass die Bezeichnung π_j im dritten Kapitel auch für Permutationen von natürlichen Zahlen verwendet wird. Da im dritten Kapitel keine Koordinaten von Punkten betrachtet werden, kollidieren die Bezeichnungen jedoch nicht.

Mit dieser Bemerkung beenden wir das Einleitungskapitel und kommen zu den benötigten Grundlagen der Wahrscheinlichkeitsrechnung.

Kapitel 2

Zufall

In diesem Kapitel werden Algorithmentypen und mathematische Betrachtungsweisen eingeführt, die in irgendeiner Form das Konzept des Zufalls verwenden. Dabei werden die später eingesetzten *zufälligen Stichproben* von endlichen Mengen stark hervorgehoben.

1 Grundlagen der Wahrscheinlichkeitsrechnung stellt stochastische Begriffe und Werkzeuge bereit.

2 Randomisierte Algorithmen führt die zwei verschiedenen Typen randomisierter Algorithmen ein. Dies sind der Las-Vegas-Algorithmus und der Monte-Carlo-Algorithmus.

3 Die probabilistische Methode: Bei dieser Methode wird „künstlich“ ein Wahrscheinlichkeitsraum konstruiert um stochastische Werkzeuge für fest gegebene Objekte zugänglich zu machen. Über diesen Umweg können Existenzaussagen bewiesen werden.

4 Derandomisierung: Dieser Abschnitt ist ein Exkurs zu derandomisierten Algorithmen. Das sind deterministische Algorithmen, welche aus randomisierten Algorithmen durch „Entfernen des Zufalls“ (Derandomisierung) entstehen.

5 Die Methode der bedingten Erwartungen/Wahrscheinlichkeiten. Diese Methode stellt eine Möglichkeit zur Derandomisierung dar. Sie kann auch in Kombination mit der probabilistischen Methode verwendet werden um einen Existenzbeweis zu führen. Im kommenden Kapitel wird die Methode der bedingten Wahrscheinlichkeiten eingesetzt, um die Existenz von ϵ -Netzen zu beweisen.

2.1 Grundlagen der Wahrscheinlichkeitsrechnung

Als Vorlage zu diesem Abschnitt über Grundlagen der Wahrscheinlichkeitsrechnung dient ein Kapitel im Anhang von Randomized Algorithms von Raghavan und Motwani [MR95, Anhang C]. Im Original werden keine Aussagen bewiesen. Für Beweise und eine allgemeinere oder axiomatischere Behandlung des Themas ergänze man das Kapitel durch ein Buch oder Skript zur Einführung in die Stochastik, z.B. von Schmitz [Sch88]. Die in diesem

Kapitel gegenüber Raghavan und Motwani ergänzten Aussagen werden bewiesen. Grundlegend für die Betrachtungen sind die folgenden zwei Begriffe:

2.1.1 Zufallsexperimente und Ereignisse

Ein Zufallsexperiment ist ein Experiment mit zufälligen Ergebnissen. Es wird auch der Begriff Zufallsprozess verwendet. Dieser Begriff ist angemessen, wenn kein aktiv durchgeführtes Experiment, sondern beispielsweise ein natürlicher Prozess vorliegt. Jede Aussage über ein Zufallsexperiment benötigt eine zugrundeliegende (nicht notwendig endliche) Ergebnismenge Ω . Die möglichen Ergebnisse des Zufallsexperiments sind die Elemente aus Ω und werden als Elementarereignisse ω bezeichnet. Eine Teilmenge $\mathcal{E} \subset \Omega$ heißt Ereignis, ihre Komplementmenge $\bar{\mathcal{E}} := \Omega - \mathcal{E}$ Gegenereignis.

Häufig gibt es verschiedene Möglichkeiten, eine Ergebnismenge einzuführen, von denen manche für eine Situation geeigneter, manche weniger geeignet sind.

Um uns einen intuitiven Zugang zu verschaffen, betrachten wir ein Beispiel für ein Zufallsexperiment mit einem Graphen, welches in den folgenden Abschnitten des Kapitels untersucht wird.

2.1.1 Exkurs („Zufälliger Teilgraph“). Aus einem Graphen $G = (V, E)$ soll ein zufälliger Teilgraph entnommen werden. Was „zufällig“ bedeutet, hängt von der verwendeten Ergebnismenge sowie dem im Anschluss definierten Wahrscheinlichkeitsmaß ab. Eine mögliche Ergebnismenge wäre die Menge $\mathcal{P}(E)$. Jedem Elementarereignis $\omega \subset E$ ließe sich dann der Graph $G_0 = (V_0, \omega)$ mit der Menge ω der in V_0 vorkommenden Knoten zuordnen. Ebenso möglich als Ergebnismenge ist $\mathcal{P}(V)$. Jedem Elementarereignis $\omega \subset V$ lässt sich folgender Graph zuordnen: $G_0 = (\omega, E_0)$ mit der Menge E_0 aller Kanten aus E , deren beide Knoten zu ω gehören. Zur Vereinfachung der Notation lässt sich statt $\mathcal{P}(V)$ (bzw. $\mathcal{P}(E)$) als Ergebnismenge auch $\Omega = \{0, 1\}^n$ mit $n = |V|$ verwenden, wenn die Knoten aus V angeordnet sind, so dass klar ist, welche Knotenmenge mit einem $\omega \in \Omega$ gemeint ist. Diese Schreibweise erfährt später durch den Begriff des Produktraumes zusätzliche Bedeutung.

Um Aussagen über Zufallsexperimente zu treffen, führt man Wahrscheinlichkeitsräume ein.

2.1.2 Wahrscheinlichkeitsräume

Wir beginnen mit der Definition einer σ -Algebra, die für die Einführung eines Wahrscheinlichkeitsmaßes im allgemeinen Fall unverzichtbar ist.

2.1.2 Definition (σ -Algebra). Eine σ -Algebra ist ein Paar (Ω, \mathbb{F}) aus einer Ergebnismenge Ω und \mathbb{F} , einer Menge von Teilmengen von Ω mit den folgenden Eigenschaften:

1. $\emptyset \in \mathbb{F}$.
2. $\mathcal{E} \in \mathbb{F} \Rightarrow \bar{\mathcal{E}} \in \mathbb{F}$.

3. $\mathcal{E}_1, \mathcal{E}_2, \dots \in \mathbb{F} \Rightarrow \mathcal{E}_1 \cup \mathcal{E}_2 \cup \dots \in \mathbb{F}$.

Die Mengen aus \mathbb{F} bezeichnet man als messbar.

Die letzte Eigenschaft ist der Abschluss unter abzählbarer Vereinigung und impliziert zusammen mit der zweiten Eigenschaft den Abschluss unter abzählbarem Schnitt. Aus den ersten beiden Eigenschaften folgt auch, dass stets $\Omega \in \mathbb{F}$ ist. Mit Hilfe der messbaren Mengen definiert man ein Wahrscheinlichkeitsmaß.

2.1.3 Definition (Wahrscheinlichkeitsmaß). Sei (Ω, \mathbb{F}) eine σ -Algebra. Ein Wahrscheinlichkeitsmaß ist eine Abbildung $\mathbf{P} : \mathbb{F} \rightarrow [0, 1] \subset \mathbb{R}_0^+$ mit den folgenden Eigenschaften:

1. $\forall \mathcal{E} \in \mathbb{F}, 0 \leq \mathbf{P}[\mathcal{E}] \leq 1$.
2. $\mathbf{P}[\Omega] = 1$.
3. Für paarweise disjunkte Ereignisse $\mathcal{E}_1, \mathcal{E}_2, \dots$ gilt: $\mathbf{P}[\cup_i \mathcal{E}_i] = \sum_i \mathbf{P}[\mathcal{E}_i]$.

Aus den letzten beiden Eigenschaften folgt insbesondere $\mathbf{P}[\emptyset] = 0$.

Mit diesen Vorüberlegungen lässt sich nun ein Wahrscheinlichkeitsraum definieren:

2.1.4 Definition (Wahrscheinlichkeitsraum). Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathbb{F}, \mathbf{P})$ mit einer σ -Algebra (Ω, \mathbb{F}) und einer darauf definierten Wahrscheinlichkeit \mathbf{P} . Ein Wahrscheinlichkeitsraum heißt endlich, wenn Ω endlich ist.

2.1.5 Bemerkung. Häufig lassen sich allerdings alle Ereignisse „messen“, d.h. allen Ereignissen lässt sich eine Wahrscheinlichkeit zuordnen. Das ist insbesondere bei den für diese Diplomarbeit ausschließlich benötigten endlichen Wahrscheinlichkeitsräumen der Fall. Zur Einführung einer vereinfachten Wahrscheinlichkeitsrechnung hätte man auf die Definition einer σ -Algebra verzichten können und ein Wahrscheinlichkeitsmaß direkt von Ω nach $[0, 1]$ erklären.

Aus Gründen der Konsistenz und Transparenz verwenden wir jedoch die üblichere allgemeinere Definition aus der Wahrscheinlichkeitstheorie. Dies ist keine allzu große Erschwernis. Insbesondere wird die Notation nicht wesentlich aufwendiger, da wir folgende abkürzende Schreibweise verwenden: Ein Paar (Ω, \mathbf{P}) sei ein Wahrscheinlichkeitsraum $(\Omega, \mathbb{F}, \mathbf{P})$ mit $\mathbb{F} = \mathcal{P}(\Omega)$. Dies ist gerade die obige Situation, dass jedes Ereignis $\mathcal{E} \subset \Omega$ messbar und damit $\mathbf{P}[\mathcal{E}]$ für jedes Ereignis zu definieren ist. Die im Rahmen dieser Diplomarbeit verwendeten Wahrscheinlichkeitsräume sind alle endlich. Für diese setzt man stets $\mathbb{F} = \mathcal{P}(\Omega)$.

Notation. Für ein Elementarereignis $\omega \in \Omega$ schreiben wir kurz $\mathbf{P}[\omega]$ statt $\mathbf{P}[\mathcal{E}]$ mit $\mathcal{E} = \{\omega\}$. Gibt man $\mathbf{P}[\omega]$ für alle $\omega \in \Omega$ an, so ist \mathbf{P} wegen der dritten Eigenschaft im Falle eines abzählbaren Wahrscheinlichkeitsraumes hierdurch schon eindeutig festgelegt.

2.1.6 Exkurs. Betrachten wir erneut die zufällige Auswahl eines Teilgraphen. Statt des Prozesses insgesamt lässt sich auch die Auswahl eines bestimmten Knotens v als ein Zufallsexperiment betrachten. Als Ergebnismenge kann man $\{\emptyset, \{v\}\}$ oder $\{0, 1\}$ (1 stehe für Auswahl des Knotens, 0 für Nicht-Auswahl) verwenden. Die einzelnen Auswahlen lassen sich dann mit dem Begriff des Produktraumes zusammenfassen, was häufig technische Vorteile mit sich bringt.

Ab nun sind alle betrachteten Wahrscheinlichkeitsräume endlich, falls nicht explizit anders angegeben.

2.1.7 Definition (Produktraum).

Seien (Ω_i, \mathbf{P}_i) , $i = 1, \dots, n$, Wahrscheinlichkeitsräume.

Gegeben seien die Mengenprodukte $\Omega = \Omega_1 \times \dots \times \Omega_n$ und

$$\mathbb{F} = \mathcal{P}(\Omega_1) \times \dots \times \mathcal{P}(\Omega_n).$$

Dann ist $(\Omega, \mathbb{F}, \mathbf{P})$ mit \mathbf{P} wie folgt ein Wahrscheinlichkeitsraum, der als Produkt der Wahrscheinlichkeitsräume Ω_i bezeichnet wird.

\mathbf{P} sei hierbei durch die Wahrscheinlichkeit eines jeden gemeinsamen Elementarereignisses $(\omega_1, \dots, \omega_n)$ angegeben. Dies sei das Produkt der Wahrscheinlichkeiten der Elementarereignisse im jeweiligen Wahrscheinlichkeitsraum.

Formal ist $\mathbf{P} = \prod_i \mathbf{P}_i$ ein Abbildungsprodukt, d.h. $\mathbf{P} : \mathbb{F} = \prod_i \mathbb{F}_i \rightarrow [0, 1]$ mit

$$\mathbf{P}[(\mathcal{E}_1, \dots, \mathcal{E}_n)] = \prod_i \mathbf{P}_i[\mathcal{E}_i] \in [0, 1].$$

Wie sich leicht zeigen lässt, erfüllt \mathbf{P} in dieser Situation die erforderlichen Eigenschaften und (Ω, \mathbf{P}) ist tatsächlich ein Wahrscheinlichkeitsraum.

2.1.8 Bemerkung. Der Produktraum lässt sich auch für allgemeinere Wahrscheinlichkeitsräume ($\mathbb{F}_i \neq \mathcal{P}_i(\Omega)$) definieren, dann ist jedoch noch mathematischer Aufwand nötig um eine wohldefinierte Menge messbarer Mengen \mathbb{F} zu erhalten. Dieser Aufwand entfällt hier.

Ein zentraler Begriff für Zufallsexperimente ist die bedingte Wahrscheinlichkeit.

2.1.9 Definition (Bedingte Wahrscheinlichkeit). Die bedingte Wahrscheinlichkeit für das Eintreten von \mathcal{E}_1 unter der Annahme von \mathcal{E}_2 mit $\mathbf{P}[\mathcal{E}_2] > 0$ ist gegeben durch :

$$\mathbf{P}[\mathcal{E}_1 | \mathcal{E}_2] := \frac{\mathbf{P}[\mathcal{E}_1 \cap \mathcal{E}_2]}{\mathbf{P}[\mathcal{E}_2]}.$$

Dieser Begriff ermöglicht es eine „totale“ Wahrscheinlichkeit mit Hilfe von bedingten Wahrscheinlichkeiten zu bestimmen:

2.1.10 Satz (Formel der bedingten Wahrscheinlichkeit). Sei $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$ eine Partition der Ergebnismenge Ω . Dann lässt sich die („totale“) Wahrscheinlichkeit eines Ereignisses \mathcal{E} als Summe bedingter Wahrscheinlichkeiten wie folgt berechnen:

$$\mathbf{P}[\mathcal{E}] = \sum_{i=1}^k \mathbf{P}[\mathcal{E}|\mathcal{E}_i]\mathbf{P}[\mathcal{E}_i].$$

Der Satz wird auch als „Satz von der totalen Wahrscheinlichkeit“ bezeichnet. Insbesondere der Fall $k = 2$ ist dabei häufig von Bedeutung:

Sei $p = \mathbf{P}[\mathcal{E}_1]$ und $\mathcal{E}_2 = \Omega \setminus \mathcal{E}_1$. Dann gilt:

$$\mathbf{P}[\mathcal{E}] = p \cdot \mathbf{P}[\mathcal{E}|\mathcal{E}_1] + (1 - p) \cdot \mathbf{P}[\mathcal{E}|\mathcal{E}_2].$$

Durch Umstellung der obigen Formeln erhält man die sogenannte Regel von Bayes.

2.1.11 Lemma (Regel von Bayes). Sei $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$ eine Partition der Ergebnismenge Ω . Dann gilt für jedes Ereignis \mathcal{E} mit $\mathbf{P}[\mathcal{E}] > 0$:

$$\mathbf{P}[\mathcal{E}_i|\mathcal{E}] = \frac{\mathbf{P}[\mathcal{E}_i \cap \mathcal{E}]}{\mathbf{P}[\mathcal{E}]} = \frac{\mathbf{P}[\mathcal{E}|\mathcal{E}_i]\mathbf{P}[\mathcal{E}_i]}{\sum_{j=1}^k \mathbf{P}[\mathcal{E}|\mathcal{E}_j]\mathbf{P}[\mathcal{E}_j]}.$$

Im Folgenden wird der Begriff der Unabhängigkeit von Ereignissen definiert. In der Wahrscheinlichkeitstheorie wird er zwar allgemeiner definiert, für unsere Wahrscheinlichkeitsräume ist die folgende Definition ausreichend:

2.1.12 Definition (Unabhängigkeit). Eine nicht-leere, endliche Familie von Ereignissen $\{\mathcal{E}_i | i \in I\}$ bezeichnen wir als unabhängig, wenn für alle Teilmengen $S \subset I$ gilt:

$$\mathbf{P}\left[\bigcap_{i \in S} \mathcal{E}_i\right] = \prod_{i \in S} \mathbf{P}[\mathcal{E}_i].$$

Als abgeschwächte Form verwendet man auch den Begriff der k -fachen Unabhängigkeit: Die Ereignisse heißen k -fach unabhängig, wenn jede Teilmenge aus k Ereignissen unabhängig ist. Statt „2-fach unabhängig“ sagt man meist „paarweise unabhängig“

2.1.13 Definition (Zufallsvariable). Eine Zufallsvariable X ist eine reellwertige Funktion auf der Ergebnismenge, $X : \Omega \rightarrow \mathbb{R}$, derart dass für alle $x \in \mathbb{R}$ gilt:

$$\{\omega \in \Omega | X(\omega) \leq x\} \in \mathbb{F}.$$

Notation. Für Ereignisse der Form $\mathcal{E} = \{\omega \in \Omega | X(\omega) \leq x\}$ schreiben wir statt $\mathbf{P}[\mathcal{E}]$ auch kurz $\mathbf{P}[X \leq x]$. Entsprechend verwenden wir auch ähnliche Notationen wie $\mathbf{P}[X = x]$ oder $\mathbf{P}[X \geq x]$.

2.1.14 Definition (Verteilung). Zu einer Zufallsvariablen X ist die Verteilung $F_X: \mathbb{R} \rightarrow [0, 1]$ definiert als $F_X(x) = \mathbf{P}[X \leq x]$.

Konvention (zu Zufallsvariablen). Eine Zufallsvariable heißt diskret, wenn sie höchstens abzählbar viele Werte annimmt. Wir betrachten nur diskrete Zufallsvariablen. Die formulierten Aussagen beziehen sich stets auf solche. Häufig werden Zufallsvariablen betrachtet, deren Werte nur natürliche Zahlen sind. Noch spezieller sind die sogenannten Indikatorvariablen, welche die Werte 0 und 1 annehmen.

Eine solche Indikatorvariable X charakterisiert das Eintreten (bzw. Nicht-Eintreten) des Ereignisses \mathcal{E} mit $\mathcal{E} = \{\omega \in \Omega | X(\omega) = 1\}$ und $\bar{\mathcal{E}} = \{\omega \in \Omega | X(\omega) = 0\}$. Es lässt sich leicht sehen, dass sich die Begriffe der bedingten Wahrscheinlichkeit und Unabhängigkeit auf Zufallsvariablen übertragen. Der Grund dafür ist, dass sich Indikatorvariablen auch als andere Möglichkeit betrachten lassen, Ereignisse zu formulieren. Es gilt:

2.1.15 Definition (Unabhängigkeit von Zufallsvariablen). Zwei (diskrete) Zufallsvariablen X und Y heißen unabhängig, wenn für alle $x, y \in \mathbb{R}$ die Ereignisse $\{X = x\}$ und $\{Y = y\}$ unabhängig sind.

Um den Begriff der bedingten Wahrscheinlichkeit auf den sogenannten bedingten Erwartungswert einer Zufallsvariablen zu übertragen, definieren wir zunächst die Wahrscheinlichkeitsdichte.

2.1.16 Definition (Wahrscheinlichkeitsdichte).

Die Wahrscheinlichkeitsdichte $p_X: \mathbb{R} \rightarrow [0, 1]$ einer Zufallsvariablen X ist definiert durch:

$$p_X(x) = \mathbf{P}[X = x].$$

Nun lässt sich der Begriff des Erwartungswertes einer Zufallsvariablen definieren. Er gibt bei einem Zufallsexperiment für eine Zufallsvariable an, welcher Wert zu „erwarten“ ist.

2.1.17 Definition (Erwartungswert). Der Erwartungswert einer Zufallsvariablen X mit Dichte p wird durch die folgende Reihe definiert, falls sie absolut konvergiert:

$$\mathbf{E}[X] = \sum_{x \in \Omega} x \cdot p_X(x).$$

Dabei kann man die Summation auch über dem Bild von X bilden, da die anderen Summanden 0 sind. Wenn der Erwartungswert existiert, schreiben wir $\mathbf{E}[X] < \infty$ und verwenden auch die Sprechweise „der Erwartungswert ist endlich.“

Der Erwartungswert einer beliebigen reell-wertigen Funktion $g(X)$ sei zudem definiert durch $\mathbf{E}[g(X)] = \sum_x g(x) \cdot p_X(x)$.

Für zwei beliebige Zufallsvariablen X und Y gilt $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$. Man beachte, dass diese Eigenschaft immer erfüllt ist und keine Unabhängigkeit oder andere Beziehung von X und Y voraussetzt. Das ist bei der probabilistischen Methode von großem Nutzen und einer der Gründe, warum der Erwartungswert ein so mächtiges Instrument ist. Er motiviert auch die Verwendung von Indikatorvariablen anstelle der korrespondierenden Ereignisse. Als allgemeine Fassung erhält man die

2.1.18 Lemma (Linearität des Erwartungswertes). *Seien X_1, \dots, X_k beliebige Zufallsvariablen. Dann gilt für jede lineare Funktion $h(X_1, \dots, X_k)$:*

$$\mathbf{E}[h(X_1, \dots, X_k)] = h(\mathbf{E}[X_1], \dots, \mathbf{E}[X_k]).$$

2.1.19 Definition (Bedingter Erwartungswert, vergleiche Brönnimann [Brö95]).

Der bedingte Erwartungswert einer Zufallsvariablen unter der Voraussetzung des Ereignisses \mathcal{E} ist definiert durch:

$$\mathbf{E}[X|\mathcal{E}] = \sum_x x \cdot \mathbf{P}\{X = x|\mathcal{E}\}.$$

Diese Definition ist in Analogie zur bedingten Wahrscheinlichkeit. Die Verwendung der Wahrscheinlichkeitsdichte von x wird ersetzt durch die Wahrscheinlichkeit von $\{X = x\}$ unter der Bedingung des Ereignisses \mathcal{E} .

2.1.20 Definition (Varianz). Wir definieren die Varianz einer Zufallsvariablen X formal wie folgt:

$$\mathbf{var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2].$$

Die Varianz existiert nur, wenn die vorkommenden Erwartungswerte definiert sind. Dann verwenden wir auch die Sprechweise „die Varianz ist endlich.“ Die positive Quadratwurzel der Varianz heißt Standardabweichung und wird meist als σ_X bezeichnet. Bei Motwani [MR95] wird die Varianz einer Zufallsvariablen wie allgemein üblich als Spezialfall des k -ten Momentes definiert. Das allgemeine Konzept der k -ten Momente ist für randomisierte Algorithmen und die probabilistische Methode auch hilfreich, wird in unserem Kontext jedoch nicht benötigt.

2.1.21 Beobachtung.

$$\text{Es gilt: } \mathbf{var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

Im Folgenden betrachten wir wichtige Zufallsexperimente sowie wichtige Verteilungen: Insbesondere die Betrachtung von n -stufigen, d.h. n -fach wiederholten (unabhängigen) Zufallsexperimenten sowie Zufallsexperimenten mit nur zwei möglichen Ergebnissen ist wichtig. Durch die Definition sogenannter zufälliger n -Stichproben erhalten wir eine einfache Notation für n -stufige Zufallsexperimente.

Das Bernoulli-Experiment ist ein 1-stufiges Zufallsexperiment mit nur zwei möglichen Ergebnissen (1 bzw. KOPF sowie 0 bzw. ZAHL). Die zugehörige Verteilung heißt Bernoulli-Verteilung. Das n -stufige Bernoulli-Experiment ist eine n -fache Wiederholung des Bernoulli-Experimentes. Die Verteilung, der die Anzahl $0 \leq k \leq n$ von Einsen bzw. KÖPFEN genügt, heißt Binomialverteilung.

2.1.3 Zufällige, unabhängige n -Stichproben

Die n -fache Wiederholung eines Zufallsexperiment auf dem Wahrscheinlichkeitsraum (Ω, \mathbf{P}) ergibt ein sogenanntes n -stufiges Zufallsexperiment auf dem Wahrscheinlichkeitsraum (Ω^n, \mathbf{P}^n) (n -facher Produktraum, vgl. Definition 2.1.7). Ein n -Tupel $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$ bezeichnen wir als n -Stichprobe.

2.1.22 Bemerkung. Man kann zum selben Wahrscheinlichkeitsraum (Ω, \mathbf{P}) nun Stichproben verschiedener Größen wählen oder zwei Stichproben hintereinander auswählen, so dass zu einem (Ω, \mathbf{P}) verschiedene Produkträume betrachtet werden. In Kapitel 2 werden dies beispielsweise (Ω^m, \mathbf{P}^m) und $(\Omega^{2m}, \mathbf{P}^{2m})$ sein. Man beachte, wie praktisch die Notation $\mathbf{P}^m[\omega]$ bzw. $\mathbf{P}^m[\mathcal{E}]$ mit einem Ereignis \mathcal{E} ist: Man sieht dem Wahrscheinlichkeitsmaß \mathbf{P}^m stets an, wie viele Stellen ein betrachtetes n -Tupel oder eine Menge hat.

Diese zufälligen Stichproben bezeichnet man in der Stochastik und Kombinatorik auch als Stichproben mit Wiederholung und Zurücklegen, da ein $\omega_0 \in \Omega$ mehrfach als Element im Vektor ω vorkommen kann und Vektoren verschieden sind, wenn sie sich durch die Positionen der Elemente unterscheiden.

Bemerkung: Nur unabhängige zufällige Stichproben kann man auf diese Weise mit dem n -fachen Produkt Ω^n darstellen, da sich mit der Definition des n -fachen Produkt \mathbf{P}^n direkt (für ein $\omega_0 \in \Omega$) die Unabhängigkeit von Ereignissen $\mathcal{E}_i := \{\omega, \omega_k = \omega_0\}$ und $\mathcal{E}_j := \{\omega, \omega_j = \omega_0\}$ ergibt.

Konvention (zufällige n -Stichprobe). Da wir im Rahmen der Diplomarbeit lediglich zufällige, unabhängige n -Stichproben auf endlichen Mengen mit dem „Wahrscheinlichkeitsmaß der Gleichverteilung“ (Definition im Folgenden) benötigen, meinen wir ab nun mit zufälligen n -Stichproben nun solche.

Zufällige n -Stichproben nach dieser Konvention sind besonders leicht zu behandeln:

2.1.23 Definition (zufällige n -Stichprobe). Sei (Ω, \mathbf{P}) ein endlicher Wahrscheinlichkeitsraum mit Gleichverteilung als Wahrscheinlichkeitsmaß, d.h. $|\Omega| < \infty$ und $\mathbf{P}[\omega_0] = \frac{1}{|\Omega|} \forall \omega_0 \in \Omega$. Dann ist eine (einfache) n -Stichprobe ein Vektor $\omega = (\omega_1, \dots, \omega_n)$ mit $\omega_i \in \Omega$.

Aus der Definition 2.1.7 des Produktraumes folgt unmittelbar, dass für Mengen $\mathcal{E} \subset \Omega^n = \{\omega, \omega = (\omega_1, \dots, \omega_n), \omega_i \in \Omega\}$ stets gilt:

$$\mathbf{P}^m[\mathcal{E}] = \sum_{\omega \in \mathcal{E}} \mathbf{P}^m[\omega] = \sum_{\omega \in \mathcal{E}} \frac{1}{|\Omega|} = \frac{|\mathcal{E}|}{|\Omega|}.$$

2.1.24 Bemerkung. Das Ziehen von zufälligen n -Stichproben, also n -stufigem unabhängigen Ziehen mit Zurücklegen auf einem endlichen Wahrscheinlichkeitsraum mit gleichverteiltem Wahrscheinlichkeitsmaß, bezeichnet man auch als n -stufiges Laplace-Experiment.

2.1.4 Bernoulli-Experiment

Sei (Ω, \mathbf{P}) der endliche Wahrscheinlichkeitsraum mit $\Omega = \{0, 1\}$ und $\mathbf{P}[1] = p \in [0, 1]$. Das (1-stufige) Zufallsexperiment, dem dieser Wahrscheinlichkeitsraum zugrunde liegt, heißt Bernoulli-Experiment: Ein Münzwurf habe mit Wahrscheinlichkeit p das Elementarereignis KOPF. Häufig verwendet man $q := 1 - p$ für die Wahrscheinlichkeit von ZAHL. Sei X eine Zufallsvariable mit Wert 1 für das Ereignis KOPF und 0 sonst. Dann heißt die Verteilung von X Bernoulli-Verteilung mit Parameter p . Die Wahrscheinlichkeitsdichte (nach Definition 2.1.16) einer Bernoulli-verteilten Zufallsvariablen X ist gegeben durch

$$p_X(x) = \begin{cases} 1 - p & \text{falls } x = 0 \\ p & \text{falls } x = 1 \\ 0 & \text{sonst.} \end{cases}$$

Für die Bernoulli-Verteilung gilt: $\mathbf{E}[X] = p$ und $\mathbf{var}[X] = pq = p(1 - p)$. Im allgemeinen Fall $p \neq \frac{1}{2}$ ist die Bernoulli-Verteilung somit keine Gleichverteilung.

2.1.5 Binomialverteilung und n -stufiges Bernoulli-Experiment

Als „ n -stufiges Bernoulli-Experiment mit Parameter p “ bezeichnen wir eine n -fache Wiederholung eines Bernoulli-Experiments. Ihre Ergebnisse (Elementarereignisse) sind voneinander unabhängig. Gegeben seien n unabhängige Bernoulli-verteilte Zufallsvariablen X_1, \dots, X_n auf Ω mit gemeinsamem Parameter p . Die Zufallsvariable X_i sei die Zufallsvariable definiert auf der i -ten Stufe des Bernoulli-Experimentes. (Hierzu stelle man sich $\Omega^n = \Omega_1 \times \dots \times \Omega_n$ vor.) Das n -stufige Bernoulli-Experiment betrachten wir also auf dem n -fachen Produkt des zugehörigen Wahrscheinlichkeitsraumes: Sei (Ω, \mathbf{P}) der Wahrscheinlichkeitsraum eines Bernoulli-Experiments mit (festem) Parameter p . Mit $\Omega^n = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in \Omega\}$ erhalten wir für das Produktmaß nach Definition 2.1.7:

$$\mathbf{P}^n[\omega = (\omega_1, \dots, \omega_n)] = \prod_{i=1}^n \mathbf{P}[\omega_i] = p_{X_1}(\omega_1) \cdot \dots \cdot p_{X_n}(\omega_n).$$

Dann ist (Ω^n, \mathbf{P}^n) ein Wahrscheinlichkeitsraum und liegt diesem n -stufigen Bernoulli-Experiment zugrunde. Für ein Ereignis $\mathcal{E} = \{(\omega_1, \dots, \omega_n), \omega_i \in \Omega, \omega_k = \omega_0\}$ mit einem

festen $\omega_0 \in \Omega$ schreibe man auch kurz $\{\omega_k = \omega_0\}$. Nach Definition von Wahrscheinlichkeit und Produktraum erhält man:

$$\mathbf{P}^n[\omega_k = \omega_0] = \prod_{i \neq k} \mathbf{P}[\Omega] \cdot \mathbf{P}[\omega_0] = 1 \cdot \mathbf{P}[\omega_0] = p_X(\omega_0) = p.$$

Die Verteilung der Zufallsvariablen $X = X_1 + \dots + X_n$ heißt wie zuvor erwähnt Binomialverteilung mit Parametern n und p , manchmal abgekürzt mit $B(n, p)$, und hat die mit $b(k; n, p)$ bezeichnete Wahrscheinlichkeitsdichte. Für $k = 0, 1, \dots, n$ ist:

$$b(k; n, p) = \mathbf{P}[X = k] = \binom{n}{k} \cdot p^k \cdot q^{n-k}.$$

2.1.25 Lemma (Erwartungswert und Varianz der Binomialverteilung). *Für eine binomialverteilte Zufallsvariable X existieren der Erwartungswert und die Varianz und es gilt:*

$$\mathbf{E}[X] = np, \quad \mathbf{var}[X] = npq, \quad q = 1 - p.$$

2.1.26 Bemerkung (Interpretation des Bernoulli-Experiments bei Stichproben).

Betrachte eine Stufe des n -stufigen Bernoulli-Experiments, also ein Bernoulli-Experiment. Da die Bernoulli-Verteilung für $p \neq \frac{1}{2}$ keine Gleichverteilung ist, entspricht ein n -stufiges Bernoulli-Experiment mit $p \neq \frac{1}{2}$ nicht direkt dem n -fachen Ziehen von Stichproben aus einer zweielementigen Menge. In jedem endlichen Raum Ω lässt sich aber leicht ein Ereignis so definieren, dass das Ziehen von Stichproben ein Bernoulli-Experiment und somit eine Binomialverteilung ergibt: Man betrachte eine Teilmenge $R \subset \Omega$ und wähle den Parameter $p = \frac{|R|}{|\Omega|}$. Im Vorgriff auf das folgende Kapitel bezeichnen wir R als Bereich. Ein Ergebnis $\omega_0 \subset R$ bezeichne man als TREFFER und ein Ergebnis $\omega_0 \subset \Omega - R$ als FEHLSCHLAG. Dann gilt $\mathbf{P}[\text{TREFFER}] = p$. Definieren wir uns eine Zufallsvariable X_0 nur mit den Werten 0 und 1 und $X_0(\omega_0) = 1 \Leftrightarrow \omega_0 \in R$, so ist diese eine bernoulli-verteilte Zufallsvariable auf Ω und wir können auf Ω ein Bernoulli-Experiment betrachten (indem wir TREFFER als 1 und FEHLSCHLAG als 0 interpretieren). Damit können wir auf Ω^n wiederum ein n -stufiges Bernoulli-Experiment betrachten. Ist die Zufallsvariable X eine Summe $X = X_1 + \dots + X_n$ aus n bernoulli-verteilten Zufallsvariablen X_i , so ist diese wie im vorigen Abschnitt also binomialverteilt.

Betrachten wir nun zwei Gesetze über den Erwartungswert. Diese haben auch allgemeinere Fassungen, im Kontext der weiteren Arbeit ist aber die eingeschränkte Fassung ausreichend.

2.1.27 Satz (Formel der bedingten Erwartung).

Gegeben sei ein Wahrscheinlichkeitsraum (Ω, \mathbf{P}) und eine Zufallsvariable X mit $\mathbf{E}[X] < \infty$. Sei $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$ eine Partition der Ergebnismenge Ω . Dann lässt sich der („totale“) Erwartungswert einer Zufallsvariablen X als Summe bedingter Erwartungswerte wie folgt berechnen:

$$\mathbf{E}[X] = \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{E}[X|\mathcal{E}_i].$$

Insbesondere der Fall $k = 2$ ist dabei von Bedeutung:

Mit $\mathcal{E} := E_1$, $p := \mathbf{P}[\mathcal{E}]$ und $\mathcal{E}_2 := \bar{\mathcal{E}}$ gilt:

$$\mathbf{E}[X] = p \cdot \mathbf{E}[X|\mathcal{E}] + (1 - p) \cdot \mathbf{E}[X|\bar{\mathcal{E}}].$$

Beweis. Es gilt:

$$\begin{aligned} \mathbf{E}[X] &= \sum_x x \cdot \mathbf{P}\{\{X = x\}\} \\ &\stackrel{\text{def}}{=} \sum_x x \cdot \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{P}\{\{X = x\}|\mathcal{E}_i\} \\ &= \sum_{i=1}^k \sum_x x \cdot \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{P}\{\{X = x\}|\mathcal{E}_i\} \\ &= \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \sum_x x \cdot \mathbf{P}\{\{X = x\}|\mathcal{E}_i\} \\ &= \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{E}[X|\mathcal{E}_i]. \end{aligned}$$

Da nur endliche Wahrscheinlichkeitsräume betrachtet werden, existieren alle Erwartungswerte und man kann die Summationsreihenfolge ändern. In der letzten Gleichung wurde schließlich die Formel der bedingten Wahrscheinlichkeiten mit $\Omega = \mathcal{E} \cup \bar{\mathcal{E}}$ angewendet. \square

2.1.28 Lemma (Einschränkung des Erwartungswertes).

Sei $\Omega = \Omega_1 \times \Omega_2$ Produkt zweier Wahrscheinlichkeitsräume mit Wahrscheinlichkeitsmaß \mathbf{P} . Das Wahrscheinlichkeitsmaß auf Ω_2 sei \mathcal{P}_2 . Sei X eine Zufallsvariable auf Ω , dessen Erwartungswert existiert. Vorgelegt sei ein Ergebnis (Elementarereignis) ω_2 von Ω_2 . Dann entspricht der bedingte Erwartungswert von X unter \mathcal{E} dem Erwartungswert der Zufallsgröße X_{Ω_1} auf Ω_1 , die durch die Projektion von Ω_2 auf ω_2 gegeben ist:

$$\mathbf{E}[X|\{\omega_2\}] = \mathbf{E}[X_{\Omega_1}] \text{ mit } X_{\Omega_1}(\omega_1) := X(\{\omega_1, \omega_2\}).$$

(Die formale Schreibweise als Projektion erscheint hier nicht notwendig.) Dabei ist $\{\omega_2\}$ auf natürliche Weise als Ereignis in Ω zu verstehen, formal korrekt ist das Ereignis als $\Omega_1 \times \{\omega_2\}$ zu schreiben.

Beweis. Der Hauptschritt im Beweis liegt darin, den Erwartungswert korrekt zu erklären: Da X_{Ω_1} auf Ω_1 lebt, ist dies offensichtlich geleistet durch

$$\mathbf{E}[X_{\Omega_1}] = \sum_x x \cdot \mathbf{P}_1[X_{\Omega_1} = x].$$

Es ist also ausreichend, für alle $x \in \mathbb{R}$ zu zeigen:

$$\mathbf{P}_1[X_{\Omega_1} = x] = \mathbf{P}[X = x | \{\omega_2\}].$$

Unmittelbar aus der Definition der bedingten Wahrscheinlichkeit folgt für alle \mathcal{E}_1 :

$$\mathbf{P}_1[\mathcal{E}_1] = \mathbf{P}[\mathcal{E}_1 | \{\omega_2\}].$$

Sei nun $x \in \mathbb{R}$. Durch Summation aller \mathcal{E}_1 mit $X(\mathcal{E}_1) = x$ folgt dann

$$\mathbf{P}_1[X_{\Omega_1} = x] = \mathbf{P}[X = x | \{\omega_2\}].$$

und somit die Behauptung. □

Auch wenn dieses Lemma durch seine Notation abstrakt wirkt, ist die Bedeutung einfach und intuitiv: Sei ein mögliches Teilergebnis eines Zufallsprozesses gegeben. Dann berechnet man auf zwei Weisen die gleiche Größe wie folgt:

1. Den Erwartungswert einer Zufallsvariable des gesamten Zufallsprozesses unter der Bedingung des Teilergebnisses. (Linke Seite.)
2. Den Erwartungswert des übrigen Zufallsprozesses unter gegebenem Teilergebnis. (Rechte Seite.)

2.1.6 Ungleichungen für große Abweichungen (*tail estimates*)

In diesem Unterabschnitt werden zwei Abschätzungen für Wahrscheinlichkeiten formuliert, die für randomisierte Algorithmen und probabilistische Beweise von enormer Bedeutung sind. Sie bieten obere Grenzen für die Wahrscheinlichkeit, dass eine Zufallsvariable erheblich von ihrem Erwartungswert abweicht. Die Sätze und ihre Beweise sind aus dem dritten Kapitel von Raghavan und Motwani [MR95, Kapitel 3] übernommen.

2.1.29 Satz (Markov-Ungleichung). *Sei Y eine Zufallsvariable mit ausschließlich nicht-negativen Werten. Dann gilt für alle positiven Parameter t :*

$$\mathbf{P}[Y \geq t] \leq \frac{\mathbf{E}[Y]}{t}.$$

Beweis. Wir definieren $f(y)$ durch $f(y) = 1$ für $y \geq t$ und $f(y) = 0$ sonst. Nach Definition 2.1.17 gilt für den Erwartungswert der Funktion $\mathbf{P}[Y \geq t] = \mathbf{E}[f(Y)]$. Mit der für alle y gültigen Abschätzung $f(y) \leq \frac{y}{t}$ folgt dann

$$\mathbf{P}[Y \geq t] = \mathbf{E}[f(Y)] \leq \mathbf{E}\left[\frac{Y}{t}\right] = \frac{\mathbf{E}[Y]}{t}. \quad \square$$

Die Markov-Ungleichung ist die stärkste Abschätzung für große Abweichungen im allgemeinen Fall. Mit allgemeinem Fall meinen wir, dass die Zufallsvariable Y nur nicht-negative Werte annimmt und ihr Erwartungswert definiert ist, aber keine weitere Annahme über die

Verteilung gemacht wird. Trotzdem ist diese Ungleichung häufig zu schwach. Auch für den probabilistischen Beweis im zweiten Kapitel wird die schärfere Abschätzung nach Chebyshev (im deutschen Sprachraum als Tschebyschow oder auch Tschebyscheff transkribiert) benötigt. Es gibt noch viele speziellere Ungleichungen wie z.B. die Chernoffungleichung [MR95, Kapitel 4], die je nach Verteilung einer Zufallsvariablen deutlich bessere Abschätzungen liefern können. Für die Verwendung im zweiten Kapitel ist die im Folgenden definierte Chebyshev-Ungleichung jedoch stark genug.

2.1.30 Satz (Chebyshev-Ungleichung). *Sei X eine Zufallsvariable mit endlichem Erwartungswert und endlicher Varianz $\mathbf{var}[X]$. Dann gilt für alle positiven reellen Zahlen k :*

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq k] \leq \frac{\mathbf{var}[X]}{k^2}.$$

D.h. die Wahrscheinlichkeit für eine Abweichung einer Zufallsvariablen um mehr als k vom Erwartungswert lässt sich gegen den Quotient aus der Varianz und k^2 abschätzen. (Für $k = t \cdot \sqrt{\mathbf{var}[X]}$ ergibt sich die im Original von Motwani verwendete Form.)

Beweis. Da für alle $x \in \mathbb{R}$ die Äquivalenz $|x - \mathbf{E}[X]| \geq k \iff (x - \mathbf{E}[X])^2 \geq k^2$ erfüllt ist, gilt:

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq k] = \mathbf{P}[(X - \mathbf{E}[X])^2 \geq k^2].$$

Die Zufallsvariable $Y = (X - \mathbf{E}[X])^2$ hat nach der Definition 2.1.20 der Varianz den Erwartungswert $\mathbf{var}[X]$. Anwendung der Markov-Ungleichung auf die Zufallsvariable Y mit $t = k^2 > 0$ ergibt damit:

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq k] = \mathbf{P}[(X - \mathbf{E}[X])^2 \geq k^2] \leq \frac{\mathbf{var}[X]}{k^2}. \quad \square$$

2.1.7 Anwendung der Chebyshev-Ungleichung für Stichproben

Eine der ganz entscheidenden Abschätzungen für den Beweis von ε -Netzen in Kapitel 2 liefert die Anwendung der Chebyshev-Ungleichung auf n -Stichproben. Daher schreiben wir diese Anwendung hier als Theorem. Man betrachte zufällige n -Stichproben nach Definition 2.1.23, also Vektoren $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$ mit endlichem Ω und dem Wahrscheinlichkeitsmaß der Gleichverteilung \mathbf{P} . Wie in der Interpretation von n -Stichproben als n -stufiges Bernoulli-Experiment in Bemerkung 2.1.26 betrachten wir ein spezielles Ereignis $R \subset \Omega$. Ein $\omega_0 \in R \subset \Omega$ bezeichnen wir erneut als TREFFER und ein $\omega_0 \in \Omega - R$ bezeichnen wir als FEHLSCHLAG und wählen den Parameter $p = \mathbf{P}[\omega] = \frac{|R|}{|\Omega|}$. Die n Zufallsvariablen X_1, \dots, X_n mit $X_i(\omega_i) = 1$ für $\omega_i \in \mathcal{E}$ (TREFFER) und $X_i(\omega_i) = 0$ sonst (FEHLSCHLAG) sind wie im Abschnitt 2.1.26 bernoulli-verteilt. Dann ist die Zufallsvariable $X = X_1 + \dots + X_n$, welche die Anzahl der TREFFER zählt, binomialverteilt. Nach Lemma 2.1.25 über den Erwartungswert und die Varianz der Binomialverteilung folgt dann: Die erwartete Anzahl von TREFFERN ist der Erwartungswert $\mathbf{E}[X] = np$. Die Varianz ist $\mathbf{var}[X] = np(1 - p)$.

Theorem I (Stichprobentheorem) Sei Ω eine endliche Menge und $\omega \in \Omega$ eine zufällige n -Stichprobe nach der Definition 2.1.23 einer (unabhängigen, gleichverteilten) Stichprobe. Sei eine feste Teilmenge $R \subset \Omega$ gegeben. Das Ereignis $\omega_0 \in R \subset \Omega$ heie TREFFER. Man setze $p = \frac{|R|}{|\Omega|}$. Dann gilt fur die Anzahl X der Treffer:

$$\mathbf{P}^n[|X - np| \geq k] \leq \frac{np(1-p)}{k^2}.$$

Beweis. Wie zuvor bemerkt, ist X binomialverteilt mit dem Erwartungswert $\mathbf{E}[X] = np$ und der Varianz $\mathbf{var}[X] = np(1-p)$. Die Abschtzung folgt nun durch Einsetzen des Erwartungswertes und der Varianz in die Chebyshev-Ungleichung (Satz 2.1.30).

2.2 Randomisierte Algorithmen

Diese kurz gehaltene Einfuhrung ist dem ersten Kapitel von Motwani [MR95] entnommen und nur leicht verandert. Zur Ergnzung sei dem Leser das Kapitel 9 („Geometric Algorithms[.]“), insbesondere Abschnitt 9.9 („Random Sampling“), empfohlen. Fur viele Probleme in der algorithmischen Geometrie wie auch der Informatik insgesamt gibt es gegenuber deterministischen Algorithmen deutlich einfachere und effizientere Losungen mit randomisierten, d.h. vom Zufall abhangigen Algorithmen, vgl. [Mat96]. Auch die in dieser Arbeit vorgestellten Algorithmen fur Simplex-Bereichsanfragen verwenden das Konzept der randomisierten Algorithmen.

Daher sollen die beiden Grundtypen Las-Vegas-Algorithmus und Monte-Carlo-Algorithmus hier kurz vorgestellt werden.

2.2.1 Las-Vegas-Algorithmus

Ein Las-Vegas-Algorithmus ist ein Algorithmus, der stets ein korrektes Ergebnis liefert, dessen Laufzeit jedoch vom Zufall abhangt. Meist soll bei Las-Vegas-Algorithmen zumindest eine niedrige erwartete Laufzeit erzielt und bewiesen werden. Ein bekanntes Beispiel ist Quick-Sort, welches stets korrekt sortiert, im erwarteten Fall Laufzeit $\mathcal{O}(n \log n)$ liefert (entspricht der optimalen deterministischen Laufzeit – im ublichen Berechnungsmodell des algebraischen Entscheidungsbaumes), aber auch hohere Laufzeit benotigen kann.

Die Laufzeit kann dabei variieren unter verschiedenen Inputs sowie unter verschiedenen zufalligen Wahlen innerhalb des Algorithmus. „Zufallige“ Wahlen innerhalb des Algorithmus lassen sich interpretieren als feste Wahlen in Abhangigkeit einer zufalligen n -Stichprobe aus den Bits $\{0, 1\}$.

2.2.2 Monte-Carlo-Algorithmus

Als Monte-Carlo-Algorithmus wird ein Algorithmus bezeichnet, der nur mit einer gewissen Wahrscheinlichkeit ein korrektes Ergebnis liefert – dies jedoch mit fester Laufzeit. In den beiden folgenden Abschnitten wird ein Beispiel für einen (trivialen) Monte-Carlo-Algorithmus angegeben, der mit Wahrscheinlichkeit $1/2$ ein korrektes Ergebnis liefert. Durch wiederholte Anwendung kann eine geeignet hohe Erfolgswahrscheinlichkeit garantiert werden. (Voraussetzung dafür ist, dass die Korrektheit des Ergebnisses verifiziert werden kann.) In der Praxis sowie auch als Black-Box innerhalb eines anderen Algorithmus ist dies oft ausreichend, wenn mit schneller Laufzeit verifizierbar ist, ob ein korrektes Ergebnis vorliegt.

Dieser Grundgedanke kommt beim algorithmischen Paradigma des Random Sampling (Ziehung zufälliger Stichproben, auch bezeichnet als randomisiertes Divide & Conquer) zur Anwendung.

2.2.3 Nachteile randomisierter Algorithmen

Es gibt mehrere Gründe, warum ein randomisierter Algorithmus als Lösung für eine Problemstellung gegebenenfalls nicht vollständig zufriedenstellend ist. Seien hier zwei kurz aufgegriffen. Ausführlichere Überlegungen dazu finden sich in den Einleitungstexten der zwei Arbeiten von Matoušek [Mat96] bzw. Brönnimann [Brö95].

1. Zufall kann als eine „Ressource“ verstanden werden, da algorithmisch nur Pseudo-Zufallszahlen erzeugt werden können und auch Pseudo-Zufallszahlen mit gleichmäßiger Verteilung aufwändig zu erzeugen sind.
2. Es überwiegen Situationen, in denen ein exaktes Ergebnis verlangt wird und damit ein Monte-Carlo-Algorithmus inakzeptabel ist. Auch ein Las-Vegas-Algorithmus kann inakzeptabel sein, wenn eine Laufzeit garantiert werden muss. Dies ist beispielsweise bei Echtzeitanwendungen der Fall.

2.3 Die probabilistische Methode

Die probabilistische Methode ist eine Methode aus der Stochastik, mit der die Existenz eines Objektes mit vorgegebenen Eigenschaften bewiesen werden kann. Diese Objekte können beispielsweise Zahlen, Mengen oder Mengenstrukturen, letztendlich auch Datenstrukturen sein. Klassischerweise hat sie Anwendung in der Zahlentheorie und Kombinatorik gefunden, vgl. Erdős und Spencer [ES74]. Später fanden sich jedoch auch Anwendungen in der algorithmischen Geometrie sowie der Informatik insgesamt, wie erkennbar werden wird. Für eine umfangreiche Darstellung der Einsatzmöglichkeiten sei auf Alon, Spencer und Erdős verwiesen [ASP91].

Die (künstliche) Einführung eines Wahrscheinlichkeitsraumes ermöglicht den Einsatz von stochastischen Größen wie des Wahrscheinlichkeitsmaßes und des Erwartungswertes. Dies wird von großem Nutzen sein. Zu beachten ist jedoch, dass durch einen solchen Existenzbeweis das Objekt nicht per se konstruktiv angegeben wird. Ein typisches Beweisschema ist der Nachweis, dass ein zufällig gewähltes Objekt eine bestimmte Eigenschaft mit einer Wahrscheinlichkeit größer als Null erfüllt. Damit gilt insbesondere, dass ein solches Objekt existieren muss.

Ein weiteres Schema ist die Bestimmung des Erwartungswertes einer Zufallsvariable. Hier folgt analog die Existenz eines Objekts, für das die Zufallsvariable mindestens diesen Wert annimmt.

Betrachten wir für die zweite Variante ein Beispiel aus der Graphentheorie [Mer01, MR95], das die Eleganz und Mächtigkeit der probabilistischen Methode demonstriert:

2.3.1 Definition (Cut). Sei $G = (V, E)$ ein endlicher Graph und zur einfachen Darstellung $V = \{1, \dots, n\}$ und $E = \{e_1, \dots, e_m\}$ mit $e_i = \{u_i, v_i\}, V \ni u_i \neq v_i \in V$. Dann ist ein Cut (V_0, V_1) eine disjunkte Zerlegung $V = V_0 \dot{\cup} V_1$ der Knotenmenge. Das Gewicht des Cuts ist definiert als die Anzahl der „Schnittkanten“:

$$w(\text{Cut}) = |\{\{u, v\} \in E : u \in V_0, v \in V_1\}| + |\{\{u, v\} \in E : u \in V_1, v \in V_0\}|.$$

2.3.2 Lemma. Sei $G = (V, E), V = \{1, \dots, n\}, E = \{e_1, \dots, e_m\}, e_i = \{u_i, v_i\}$.

Dann existiert ein Cut mit $w(\text{Cut}) \geq m/2$.

Beweis. Statt einen Cut mit $w(\text{Cut}) \geq m/2$ zu konstruieren, wenden wir die probabilistische Methode durch Bildung eines Wahrscheinlichkeitsraumes an. Es soll jeder Punkt mit Wahrscheinlichkeit $1/2$ unabhängig von den anderen Punkten in V_0 oder V_1 liegen. Sei daher (Ω^n, \mathbf{P}^n) mit $\Omega = \{0, 1\}$ der Wahrscheinlichkeitsraum des Bernoulli-Experiments mit n -facher Wiederholung und Parameter $p = \frac{1}{2}$. Für jedes n -Tupel $\omega \in \Omega$ lässt sich nun der zugehörige Cut_ω definieren:

$$\text{Cut}_\omega := (V_0, V_1) \text{ mit } V_0 = \{i | \omega_i = 0\} \text{ und } V_1 = \{i | \omega_i = 1\}.$$

Schließlich sei X eine Zufallsvariable, welche für jedes ω das Gewicht von Cut_ω angibt:

$$X(\omega) := w(\text{Cut}_\omega).$$

Ziel ist nun zu zeigen, dass $\mathbf{E}[X] \geq \frac{m}{2}$ gilt. Hierzu wird für jede Kante e_i eine Zufallsvariable \hat{e}_i definiert, die charakterisiert, ob e_i eine Schnittkante ist:

$$\hat{e}_i = \begin{cases} 1 & \text{für } u_i \in V_0 \wedge v_i \in V_1 \text{ oder } u_i \in V_1 \wedge v_i \in V_0 \\ 0 & \text{für } u_i \in V_0 \wedge v_i \in V_0 \text{ oder } u_i \in V_1 \wedge v_i \in V_1 \end{cases}$$

Aufgrund der Unabhängigkeit ergibt sich für den Erwartungswert:

$$\begin{aligned} \mathbf{E}[\hat{e}_i] &= 1 \cdot \mathbf{P}[\hat{e}_i = 1] = \mathbf{P}[u_i \in V_0 \wedge v_i \in V_1] && + \mathbf{P}[u_i \in V_1 \wedge v_i \in V_0] \\ &= \mathbf{P}[u_i \in V_0] \cdot \mathbf{P}[v_i \in V_1] && + \mathbf{P}[u_i \in V_1] \cdot \mathbf{P}[v_i \in V_0] \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

Nach Definition ist X die Summe dieser Zufallsvariablen. Wegen der Linearität des Erwartungswertes \hat{e}_i folgt schon:

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^n \hat{e}_i\right] = \sum_{i=1}^n \mathbf{E}[\hat{e}_i] = m \cdot \frac{1}{2} = \frac{m}{2}.$$

An dieser Stelle soll kurz untersucht werden, durch welche Wahl des Parameters p im n -stufigen Bernoulli-Experiment die probabilistische Methode hier Beweiskraft erlangt. Da Kanten aus je zwei verschiedenen Punkten bestehen, ist recht offensichtlich, dass für jeden Punkt die gleiche Wahrscheinlichkeit angesetzt werden sollte, um Erwartungswerte einfach auswerten zu können. Setzen wir nun eine beliebige Wahrscheinlichkeit p an, so erhalten wir für den Erwartungswert einer Kante: $\hat{e}_i = p(1-p)$. Als Funktion von p aufgefasst erreicht diese ihr Maximum bei $p = 1/2$, wie sich leicht zeigen lässt. Am instruktivsten erhält man dies durch die Anwendung der aus der Analysis bekannten Ungleichung vom arithmetischen und geometrischen Mittel: Das Maximum eines Produktes mit konstanter Faktorensomme erhält man genau dann, wenn alle Faktoren gleich sind. Entscheidend für die Anwendung der probabilistischen Methode ist also die „balancierte“ Wahl $p = 1 - p = 1/2$, durch die das stochastische Instrumentarium des Erwartungswertes den maximalen Effekt erzielt. (Diese Wahl wird aufgrund der Symmetrie in der Problemstellung hier als „balanciert“ bezeichnet.) Auch bei einer Derandomisierung von Matoušek für eine Problemstellung im Zusammenhang mit Bereichsanfragen [Mat91, Abschnitt 3, S.507] ist eine für die Problemstellung balancierte Wahl für p ($p = s/n$, n Eingabegröße, s Größe einer zu findenden Teilmenge) zielführend. \square

Im folgenden Abschnitt wird die Derandomisierung in Worten eingeführt. Die gemachten Erklärungen sind angelehnt an die lesenswerten Ausführungen zu Derandomisierung von Matoušek [Mat96].

2.4 Derandomisierung

Die Derandomisierung ist eine algorithmische Technik, welche wir hier einführen. Bei der Derandomisierung wird ein randomisierter Algorithmus durch Ersetzen des zufälligen Inputs derart modifiziert, dass man einen deterministischen Algorithmus erhält. Diesen bezeichnet man als derandomisierten Algorithmus.

Ein solcher Algorithmus ist somit stets deterministisch.

2.4.1 Bemerkung zur Praktikabilität

Berechtigt ist die Frage, weshalb statt der Derandomisierung eines randomisierten Algorithmus nicht direkt ein leistungsfähiger deterministischer Algorithmus verwendet wird. Häufig sind schnellere, oft sehr einfache, randomisierte Algorithmen bekannt, während die besten bekannten deterministischen Algorithmen deutlich langsamer (und komplizierter) sind. Erstaunlicherweise gelingt es häufig, durch Zusatzroutinen aus einem randomisierten Algorithmus einen derandomisierten Algorithmus zu erhalten, der immer noch schneller ist als ein klassischer deterministischer Algorithmus. Ohne den ursprünglichen randomisierten Algorithmus methodisch zu benennen, bliebe dieser entstehende Algorithmus sowie die Analyse seiner Laufzeit jedoch in der Regel unverständlich. Ein äußerst eindrucksvolles Beispiel für dieses Phänomen ist der derandomisierte Algorithmus zur Bestimmung einer konvexen Hülle im \mathbb{R}^d in optimaler Laufzeit nach Brönnimann [Brö95]. Dadurch ist die Charakterisierung eines Algorithmus als derandomisierter (im Gegensatz zu deterministisch) gerechtfertigt.

Bevor wir dies an einem Beispiel untersuchen, sei an dieser Stelle plausibel gemacht, dass diese Zusatzroutinen in gewissen Situationen eine nur so geringfügige Verschlechterung der Laufzeit verursachen, dass derandomisierte Algorithmen mit deterministischen konkurrieren können. Dies ist wichtig, da bei einem trivialen Ersetzen des Zufalls durch Ausprobieren eine exponentielle Laufzeit zu vermuten ist.

2.4.2 Anschauliche Rechtfertigung der Praktikabilität

Oft kann man ohne großen Zusatzaufwand aus dem Wahrscheinlichkeitsraum ein Objekt mit geeigneten Eigenschaften herausgreifen. Als Beispiel hierfür kann die randomisierte Version des Quicksort-Algorithmus dienen [MR95, Theorem 1.1]. Wählt man als Pivot-Element stets den Median, lässt sich die Laufzeit $\mathcal{O}(n \log n)$ garantieren. Diese Wahl ist mit Median-Find in linearer Laufzeit zu bewältigen (Quicksort mit Median-Find nimmt jedoch gewissermaßen eine Sonderstellung unter den derandomisierten Algorithmen ein, nicht zuletzt weil bei seiner Entwicklung die Technik der Derandomisierung als solche noch nicht bekannt gewesen ist.) Die Derandomisierung ist keine einfache kanonische Prozedur, mit der aus jedem randomisierten Algorithmus ein guter derandomisierter entworfen werden kann. Manche Typen von randomisierten Algorithmen (Random Sampling / Divide and Conquer) eignen sich mehr, andere weniger (viele inkrementelle Algorithmen). Für umfassende Untersuchungen sei hier auf die Doktorarbeit von Brönnimann [Brö95] sowie die oben zitierte Arbeit *Derandomization in Computational Geometry* verwiesen [Mat96]. Es gibt jedoch eine systematische (wenn auch nicht kanonische) Methode zur Derandomisierung von Algorithmen, die im Folgenden vorgestellt wird.

2.5 Die Methode der bedingten Erwartungswerte bzw. Wahrscheinlichkeiten

Die Methoden der bedingten Erwartungswerte und Wahrscheinlichkeiten sind sehr verwandte Methoden. Sie beide haben zwei verschiedene Einsatzmöglichkeiten. Einerseits können sie als Teil der probabilistischen Methode in Existenzbeweisen verwendet werden, andererseits sind sie eine systematische Methode zur Derandomisierung [Mat96]. Es wird jeweils ein Wahrscheinlichkeitsraum betrachtet. Man verwendet den Satz II bzw. 2.5.1: „Unter verschiedenen Wahlmöglichkeiten zur Ersetzung eines zufälligen Teilergebnisses durch ein festes Teilergebnis gibt es stets eine solche, bei der der Erwartungswert einer Zufallsgröße beziehungsweise die Wahrscheinlichkeit eines Ereignisses mindestens gleich groß bleibt.“

Zur Derandomisierung kann man schrittweise zufällige Teilergebnisse (in der Regel einzelne Bits) durch feste ersetzen, bis schließlich nur noch feste Werte und damit ein derandomisierter Algorithmus verbleiben.

Will man einen Existenzbeweis mit Hilfe der probabilistischen Methode führen, so ist die Methode der bedingten Erwartung bzw. Wahrscheinlichkeit hilfreich um den Wahrscheinlichkeitsraum einzuschränken. Der verbleibende Wahrscheinlichkeitsraum kann eine einfacher zu untersuchende Gestalt haben. Diese Vorgehensweise wird in dem probabilistischen Existenzbeweis für kleine ε -Netze nach Haussler und Welzl in Satz 3.5.1 angewendet. In diesem Abschnitt wird die gemeinsame Grundidee unter a) bzw. b) erläutert.

2.5.1 Beschreibung der Methode

Wir beschreiben hier die Methode in seinen beiden Varianten. Ein randomisierter Algorithmus hänge von n durch n -fachen Münzwurf gegebenen Bits ab.

Jedem $\omega \in \{0,1\}^n$ lässt sich also ein (Ausgabe-)Objekt des Algorithmus zuordnen. Der Algorithmus erfülle eine der beiden Bedingungen:

- a) Eine Zufallsvariable X , die möglichst groß sein soll, habe einen gewissen Erwartungswert $\mathbf{E}[X]$.
- b) Das Objekt habe mit positiver Wahrscheinlichkeit $p > 0$ eine gewisse Eigenschaft.

Ziel: Es soll auf deterministische Weise ein Objekt konstruiert werden, für welches a) die Zufallsvariable mindestens den obigen Erwartungswert annehme bzw. b) mit Wahrscheinlichkeit 1 die obige Eigenschaft erfülle. Bei der Methode der bedingten Erwartungswerte bzw. Wahrscheinlichkeiten werden die zufälligen Bits sukzessive durch diejenige Wahl (0 oder 1) ersetzt, die für das Ziel geeigneter ist. Mit der Formel für bedingte Erwartung bzw. der Formel für bedingte Wahrscheinlichkeit lässt sich zeigen, dass man bei geeigneter Wahl schließlich ein derartiges Objekt erhält.

Nun folgt ein Beispiel für die Variante b) mit Cuts. Als Vorlage dient der dritte Abschnitt aus einer Vorlesungsmitschrift von Merkle [Mer01]. Das Beispiel findet sich auch bei Raghavan und Motwani [MR95]. In Lemma 2.3.2 wurde die Existenz eines gewissen Cuts bewiesen, an dieser Stelle wird ein solcher Cut algorithmisch bestimmt.

2.5.2 Zum Theorem über die maximale Wahrscheinlichkeit

Grundlage für das Funktionieren der Methode bieten zwei einfache Sätze, die wir im Einleitungstext des Abschnittes wie folgt umgangssprachlich formuliert haben: „Unter verschiedenen Wahlmöglichkeiten zur Ersetzung eines zufälligen Teilergebnisses durch ein festes Teilergebnis gibt es stets eine solche, bei welcher der betrachtete Erwartungswert beziehungsweise die Wahrscheinlichkeit mindestens gleich groß bleibt.“ Diese beiden Sätze basieren auf der Formeln für die bedingte Wahrscheinlichkeit beziehungsweise die bedingte Erwartung. Den Satz über die maximale Wahrscheinlichkeit bezeichnen wir als Theorem, da er im folgenden Kapitel für einen Existenzbeweis eingesetzt wird.

Theorem II (über die maximale Wahrscheinlichkeit) . *Seien ein Wahrscheinlichkeitsraum (Ω, \mathbf{P}) und eine endliche Partitionierung $\Omega = \mathcal{E}_1 \dot{\cup} \mathcal{E}_2 \dot{\cup} \dots \dot{\cup} \mathcal{E}_k$ gegeben. Dann gilt für ein beliebiges Ereignis $\mathcal{E} \subset \Omega$:*

$$\max_{i=1}^k \mathbf{P}[\mathcal{E}|\mathcal{E}_i] \geq \mathbf{P}[\mathcal{E}].$$

Das heißt, die maximale bedingte Wahrscheinlichkeit unter k Ereignissen, von denen genau eines eintritt, ist stets größer oder gleich der (totalen) Wahrscheinlichkeit.

Beweis (indirekt). Anschaulich ist dies völlig klar, da der totale Erwartungswert ein (gewichtetes) arithmetisches Mittel ist. Wir führen den formalen Beweis indirekt durch die Annahme: $\forall i : \mathbf{P}[\mathcal{E}|\mathcal{E}_i] < \mathbf{P}[\mathcal{E}]$ Daraus können wir schnell einen Widerspruch herleiten:

$$\begin{aligned} \mathbf{P}[\mathcal{E}] &= \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{P}[\mathcal{E}|\mathcal{E}_i] \\ &< \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{P}[\mathcal{E}] \\ &= \mathbf{P}[\mathcal{E}] \cdot \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] = \mathbf{P}[\mathcal{E}] \cdot \mathbf{P}[\Omega] \\ &= \mathbf{P}[\mathcal{E}]. \end{aligned} \quad \text{⚡}$$

Das Theorem ergibt sich also unmittelbar aus der Formel der bedingten Wahrscheinlichkeit 2.1.9

$$\text{Also gilt wie gefordert: } \max_{i=1}^k \mathbf{P}[\mathcal{E}|\mathcal{E}_i] \geq \mathbf{P}[\mathcal{E}]. \quad \square$$

2.5.3 Satz über die maximale Erwartung

Der Satz über die maximale Erwartung ist völlig analog. Auch der bedingte Erwartungswert ist ein (gewichtetes) arithmetisches Mittel.

2.5.1 Satz (über die maximale Erwartung). *Sei erneut ein Wahrscheinlichkeitsraum (Ω, \mathbf{P}) mit einer endlichen Partitionierung $\Omega = \{\mathcal{E}_1 \dot{\cup} \mathcal{E}_2 \dot{\cup} \dots \dot{\cup} \mathcal{E}_k\}$ gegeben. Sei X eine Zufallsvariable. Dann existieren die bedingten Erwartungswerte $\mathbf{E}[X|\mathcal{E}_i]$, $i = 1, \dots, k$ und ein maximaler unter diesen ist mindestens so groß wie $\mathbf{E}[X]$:*

$$\max_{i=1}^k \mathbf{E}[X|\mathcal{E}_i] \geq \mathbf{E}[X].$$

Wir führen hier also auch völlig analog einen indirekten Beweis.

Beweis (indirekt). Wir machen die folgenden Annahme:

$$\forall i : \mathbf{E}[X|\mathcal{E}_i] < \mathbf{E}[X].$$

Wir leiten wiederum einen Widerspruch her, denn es folgt:

$$\begin{aligned} \mathbf{E}[X] &= \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{E}[X|\mathcal{E}_i] \\ &< \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] \cdot \mathbf{E}[X] \\ &= \mathbf{E}[X] \cdot \sum_{i=1}^k \mathbf{P}[\mathcal{E}_i] = \mathbf{E}[X] \cdot \mathbf{P}[\Omega] \\ &= \mathbf{E}[X]. \end{aligned} \quad \downarrow$$

Also gilt wie gefordert: $\max_{i=1}^k \mathbf{E}[X|\mathcal{E}_i] \geq \mathbf{E}[X]$. □

Der Satz ergibt sich also unmittelbar aus der Formel des bedingten Erwartungswertes 2.1.27.

2.5.2 Bemerkung. Da die beiden Aussagen im Grunde nur sehr elementare Umstellungen der Formeln der bedingten Wahrscheinlichkeit bzw. des bedingten Erwartungswertes sind, wäre (unter dem Maßstab der Eigenständigkeit ihrer Aussagen) allenfalls eine Formulierung als Korollar angemessen. Die Bezeichnung als Satz bzw. Theorem erfolgt aus einer methodischen Motivation: Die Sätze sind in ihrer Anwendung der zentrale Gegenstand der Methode der bedingten Wahrscheinlichkeit bzw. des bedingten Erwartungswertes. Diese formal durchzuführen ist im Grunde leichtes Handwerk. Umso erstaunlicher ist, dass mit ihrer Hilfe deterministische Algorithmen gewonnen werden können, die ohne probabilistische Betrachtungen unzugänglich sind.

Als ein Beispiel dient der folgende Exkurs in das Gebiet der endlichen Graphen. Anhand diesem wird die Methode der bedingten Erwartungswerte konkret erläutert.

2.5.4 Exkurs zur Erläuterung der Methode: Algorithmus zum Finden von Cuts

Wir erinnern an die obige Definition des Gewichtes eines Cuts:

2.5.3 Definition. Sei wie im obigen Abschnitt $G = (V, E)$ ein Graph mit $V = \{1, \dots, n\}$, $E = \{e_1, \dots, e_m\}$ sowie $w(\text{Cut})$ die Anzahl der Schnittkanten eines $\text{Cuts}(V_0, V_1)$.

Gesucht ist ein deterministischer Algorithmus, der einen Cut mit $w(\text{Cut}) \geq m/2$ berechnet. Dieses Ziel erreicht man durch Derandomisieren des folgenden Monte-Carlo-Algorithmus, welcher einen beliebigen Cut berechnet:

Algorithmus CUT

Eingabe: Graph $G=(V,E)$ mit $V=1,2,\dots,n$

Wähle eine zufällige n -Stichprobe ω aus $\{0, 1\}^n$

$$V_0 = \{i | \omega_i = 0\}$$

$$V_1 = \{i | \omega_i = 1\}$$

Ausgabe: Cut (V_0, V_1)

Es sei X die Zufallsvariable, die jedem $\omega \in \Omega$ das Gewicht des ausgegebenen Cuts zuordnet. Der Erwartungswert von X ist wie zuvor gezeigt $m/2$. Insbesondere existiert also nach dem Satz 2.5.1 über die maximale Erwartung ein Cut mit $w(\text{Cut}) \geq m/2$. Anmerkung: Dieser Monte-Carlo-Algorithmus hat Erfolgswahrscheinlichkeit $\geq 1/2$. (Ohne Beweis.)

Für den gesuchten derandomisierten Algorithmus sollen wie zuvor beschrieben sukzessive die zufälligen Bits durch feste Auswahlen ersetzen. Sei $x = (x_1, x_2, \dots, x_s)$ ein binäres s -Tupel mit $s \leq n$, d.h. eine (feste) s -Stichprobe im Gegensatz zu einer zufälligen s -Stichprobe. Dann sei $\text{CUT}(x)$ der Algorithmus, der durch Festlegung der ersten s Indizes mit $\omega_i = x_i$ aus CUT entsteht (D.h. die weiteren Bits bleiben zufällig.) Zur expliziteren Notation im Folgenden sei das 0-Tupel mit λ bezeichnet, so dass $\text{CUT}(\lambda) = \text{CUT}$ ist.

Algorithmus CUT(x)

Eingabe: Graph $G=(V,E)$ mit $V=\{1,2,\dots,n\}$

Setze $\omega_i = x_i$, $i = 1, \dots, s$.

Wähle paarweise unabhängig jeweils mit $p = 1/2$ zufällige Bits $\omega_{s+1}, \dots, \omega_n$

$$V_0 = \{i | \omega_i = 0\}$$

$$V_1 = \{i | \omega_i = 1\}$$

Ausgabe: Cut (V_0, V_1)

Sei X_x die Zufallsvariable für das Gewicht eines Cuts, den $\text{CUT}(x)$ berechnet. Nun wollen

wir den Satz 2.5.1 über die maximale Erwartung anwenden. Ziel ist es, auch die zufällige Wahl des Bits ω_{s+1} durch eine feste Wahl zu ersetzen, so dass induktiv alle Bits ersetzt werden. Die Ersetzung des Bits ω_{s+1} durch eine feste Wahl entspricht der Betrachtung eines Teilergebnisses des Zufallsexperimentes. Die zwei möglichen Teilergebnisse sind die Ereignisse $\mathcal{E}_0 := \{x_{s+1} = 0\}$ und $\mathcal{E}_1 := \{x_{s+1} = 1\}$. Aus dem Satz 2.5.1 folgt dann:

$$\max_{k=0}^1 \mathbf{E}[X_x | \mathcal{E}_k] \geq \mathbf{E}[X_x].$$

Sei $k \in \{0, 1\}$ so, dass das Maximum angenommen wird. Dann treffen wir die Entscheidung $x_{s+1} := k$. Damit hat der Algorithmus $\text{CUT}(x_1, \dots, x_s, x_{s+1})$ einen mindestens so großen Erwartungswert für die Anzahl der Schnittkanten wie $\text{CUT}(x)$. Iteriert man dies, erhält man ein Schema von Algorithmen mit einer monoton wachsenden Folge für den Erwartungswert der Schnittkanten:

$$\text{CUT} = \text{CUT}(\lambda) \rightarrow \text{CUT}(x_1) \rightarrow \text{CUT}(x_1, x_2) \rightarrow \dots \rightarrow \text{CUT}(x_1, x_2, \dots, x_n).$$

Der letzte Algorithmus verwendet keine zufälligen Bits und ist damit ein deterministischer Algorithmus. Der formal definierte bedingte Erwartungswert ist identisch mit der Anzahl der Schnittkanten: $\mathbf{E}[X_{(x_1, x_2, \dots, x_n)}] = w(\text{Cut}(x_1, \dots, x_n))$. Damit liefert $\text{CUT}(x_1, \dots, x_n)$ einen gewünschten Cut. Wenn man nun noch das Treffen der Entscheidungen algorithmisch spezifiziert, erhält man den derandomisierten Algorithmus:

Algorithmus DERANDOMCUT

Eingabe: Graph $G=(V,E)$ mit $V=\{1,2,\dots,n\}$

FOR($s = 1..n$)

IF $\mathbf{E}[X_{(x_1, \dots, x_{s-1})} | \{\omega_{s+1} = 1\}] \geq \mathbf{E}[X_{(x_1, \dots, x_{s-1})} | \{\omega_{s+1} = 0\}]$

THEN $x_s = 1$

ELSE $x_s = 0$

ENDIF

$V_0 := \{i : x_i = 0\}$

$V_1 := \{i : x_i = 1\}$

Ausgabe: Cut (V_0, V_1)

Zuletzt sei kurz erwähnt, wie analog bei bedingten Wahrscheinlichkeiten zu verfahren ist: Gesucht ist ein Algorithmus, dessen Ausgabe eine gewisse Eigenschaft \mathcal{K} erfüllt. Gegeben sei ein randomisierter Algorithmus ALG mit zufälligen Bits $\omega_1, \dots, \omega_n$. Bewiesen sei (durch eine probabilistische Betrachtung), dass es unter der Menge der möglichen Ausgaben des Algorithmus eine Ausgabe gibt, die diese Eigenschaft erfüllt, d.h. $\mathbf{P}[\mathcal{K}] > 0$. Dann findet man durch die Formel der bedingten Wahrscheinlichkeit eine Folge von Algorithmen

$$\text{ALG} \rightarrow \text{ALG}(x_1) \rightarrow \text{ALG}(x_1, x_2) \rightarrow \dots \rightarrow \text{ALG}(x_1, x_2, \dots, x_n) =: \text{DERANDOMALG},$$

für die $\mathbf{P}[\mathcal{K}]$ monoton wächst. Für die Ausgabe des derandomisierten Algorithmus DERANDOMALG ist dann die Eigenschaft \mathcal{K} nachgewiesen, denn $\mathbf{P}[\mathcal{K}]$ nimmt nur die Werte 1 (falls \mathcal{K} erfüllt wird) und 0 (sonst) an.

Dies ist jedoch nur das allgemeine Schema zur Derandomisierung mit bedingtem Erwartungswert oder bedingter Wahrscheinlichkeit. Findet man keine effiziente sukzessive Berechnung der bedingten Werte, kann die Laufzeit einen hohen polynomiellen Faktor erhalten, womit der derandomisierte Algorithmus wertlos wird. (Wie bei Raghavan und Motwani [MR95] erläutert, gibt es im Falle des Beispiels eine effiziente Berechnung.)

In der Einführung zu derandomisierten Algorithmen wurde skizziert, dass aus einem gegebenen randomisierten Algorithmus durch Anwendung einer Zusatzroutine die zufälligen Wahlen entfernt werden können. Jedoch werden in der Regel für derandomisierte Algorithmen speziell angepasste randomisierte Algorithmen entworfen, die eine geeignete Abschätzung von Wahrscheinlichkeiten oder Erwartungswerten ermöglichen. Der Prozess der Derandomisierung ist keineswegs kanonisch. Häufig ist ein technisch aufwändiger Umgang mit stochastischen Größen erforderlich. Oft ist zunächst eine Anpassung eines bekannten randomisierten Algorithmus notwendig um ihn derandomisieren zu können.

Haussler und Welzl gelingt keine deterministische Konstruktion der Datenstruktur für Halbraum- oder Simplexbereichsanfragen. Bei Haussler und Welzl wird die Methode der bedingten Wahrscheinlichkeit eingesetzt. Dies geschieht als Teil eines probabilistischen Existenzbeweises.

Kapitel 3

Bereichsräume und ε -Netze

In diesem Kapitel wird das abstrakte Konzept eines Bereichsraumes eingeführt. ε -Netze und ε -Approximationen werden in diesem allgemeinen Zusammenhang definiert und untersucht. Es wird auch ein (kombinatorischer) Dimensionsbegriff eingeführt: Die VC-Dimension. Der Begriff der VC-Dimension (oder kurz Dimension) eines Bereichsraumes geht auf Vapnik und Chervonenkis zurück [VC71]. Er wird von Haussler und Welzl verwendet, da viele geometrische Bereichsräume eine endliche VC-Dimension haben und diese (kombinatorische) Eigenschaft schon ausreicht um die Existenz von kleinen ε -Netzen und ε -Approximationen nachzuweisen, wie in diesem Kapitel gezeigt wird. Einen sehr guten aktuellen Überblick, der sich auch als Einführung eignet, bietet Matoušek in *Epsilon-Nets and Computational Geometry* [Mat93]. Dort werden die Begriffe definiert und eine kleine Sammlung von Resultaten zu ε -Netzen bewiesen. Matoušek betrachtet Anwendungen von ε -Netzen für Probleme der algorithmischen Geometrie (das *Cutting lemma* sowie das *Short edge lemma*). Auch die im Rahmen dieser Diplomarbeit behandelte Anwendung von ε -Netzen für Simplex-Bereichsanfragen stellt Matoušek für den zweidimensionalen Fall sprachlich schön und prägnant dar [Mat93, Problem 3.4]. Wie sich im folgenden Kapitel zeigt, ist eine abstraktere Formulierung von Bereichsräumen für die spezielle Problemstellung der Halbraum-Bereichsanfragen hilfreich. Daher führen wir Bereichsräume als abstraktes Konzept ein und untersuchen bestimmte Strukturen und Eigenschaften von ihnen.

3.1 Bereichsräume, ε -Netze und ε -Approximationen

3.1.1 Definition (Bereichsraum). Ein Bereichsraum B ist ein Paar $B = (X, \mathcal{R})$ mit einer Menge X , dessen Elemente x Punkte oder Objekte heißen und einer *Bereichsmenge* $\mathcal{R} \subset \mathcal{P}(X)$. \mathcal{R} ist eine Menge von Teilmengen von X . Die Elemente $R \in \mathcal{R}$ heißen Bereiche. Ein Bereichsraum (X, \mathcal{R}) heißt endlich, wenn X (und damit auch \mathcal{R}) endlich ist.

Haussler und Welzl verwenden für Bereiche die Bezeichnung r sowie für die Bereichsmenge R . Diese Notation wird abgeändert. Einerseits ist hier $r := 1/\varepsilon$ wie bei Matoušek eine Bezeichnung für einen Parameter.

Andererseits sollen Punktmenge mit Großbuchstaben bezeichnet werden, vgl. Einleitungskapitel. Entsprechend soll die Bereichsmenge als Menge von Punktmenge mit einem Fraktur- oder Skriptbuchstaben bezeichnet werden. Um die Bereichsmenge deutlich von Bereichen R unterscheiden zu können, wird die Bezeichnung \mathcal{R} (und nicht etwa \mathcal{R}) verwendet.

3.1.2 Definition (große Bereiche). Sei ein endlicher Bereichsraum $B = (X, \mathcal{R})$ vorgelegt. Sei $\varepsilon > 0$. Dann nennen wir einen Bereich $R \in \mathcal{R}$ groß oder ε -kritisch, wenn $|R| > \varepsilon \cdot |X|$ gilt. Wir definieren die Menge aller großen Bereiche:

$$\mathcal{R}_\varepsilon := \{R \in \mathcal{R} : |R| > \varepsilon \cdot |X|\}.$$

3.1.3 Definition (Unterraum). Gegeben sei ein Bereichsraum (X, \mathcal{R}) und eine Teilmenge $Y \subset X$. Dann ist $(Y, \mathcal{R}|_Y)$ mit $\mathcal{R}|_Y = \mathcal{R} \cap Y := \{R' = R \cap Y \text{ mit } R \in \mathcal{R}\}$ ein durch Y induzierter Bereichsraum, der als Unterraum von X bezeichnet wird. Dabei ist $\mathcal{R} \cap Y$ eine Teilmenge von $\mathcal{P}(Y)$, d.h. verschiedene Bereiche R_1, R_2 induzieren den gleichen Bereich R' in $\mathcal{R} \cap Y$, wenn $R' = R_1 \cap Y = R_2 \cap Y$ gilt. Diese Definition ist sehr fruchtbar, da sich die im Folgenden definierte VC-Dimension (gewissermaßen eine Charakterisierung der kombinatorische Komplexität) eines Bereichsraumes auf seine Unterräume vererbt.

Durch den Begriff des Unterraumes kann man „Bereichsanfragen“ (die wir erst später formal definieren) im Kontext von Bereichsräumen betrachten.

3.1.4 Beispiel (Halbebenen-Bereichsanfrage). Gegeben sei eine endliche Punktmenge A in der Ebene \mathbb{R}^2 . Wird uns eine offene Halbebene $R \subset \mathbb{R}^2$ in der Ebene vorgelegt, so wollen wir berechnen, welche der Punkte aus A in R liegen. Gesucht ist also der Schnitt $A \cap R$. Definieren wir also den Bereichsraum $(X, \mathcal{R}) = (\mathbb{R}^2, \mathcal{R})$, wobei \mathcal{R} die Menge aller offenen Halbräume des \mathbb{R}^2 ist. Wir sind an dem Unterraum $(A, \mathcal{R}|_A)$ interessiert. Dessen Bereiche haben die Form $R' = R \cap A$, sie sind die gesuchten endlichen Punktmenge.

In Abbildung 3.1.4 liegt eine endliche Punktmenge A vor. Desweiteren sind einige Geraden angedeutet, über bzw. unter derer gewisse Teilmenge von A sind. Die Menge aller Teilmenge, die in einer beliebigen Halbebene des \mathbb{R}^2 liegen, ist die Bereichsmenge $\mathcal{R}|_A$.

3.1.5 Bemerkung. Man beachte, dass Bereichsräume abstrakt definiert sind und sich nicht unmittelbar algorithmisch zur Lösung von Bereichsanfragen einsetzen lassen. Dies sieht man leicht ein: Schon die Aufzählung der Bereiche $\mathcal{R}|_A$ entspräche der Aufzählung aller möglichen Ergebnisse eines Algorithmus, wie Matoušek bemerkt [Mat91, Einleitung von Abschnitt 4].

Typischerweise interessiert man sich für endliche Bereichsräume (X, \mathcal{R}) , die als Unterraum aus einem Bereichsraum mit einer größeren Menge, z.B. \mathbb{R}^d entstanden sind. Für die folgenden Betrachtungen sei also nun stets ein $\varepsilon > 0$ und ein endlicher Bereichsraum (X, \mathcal{R})

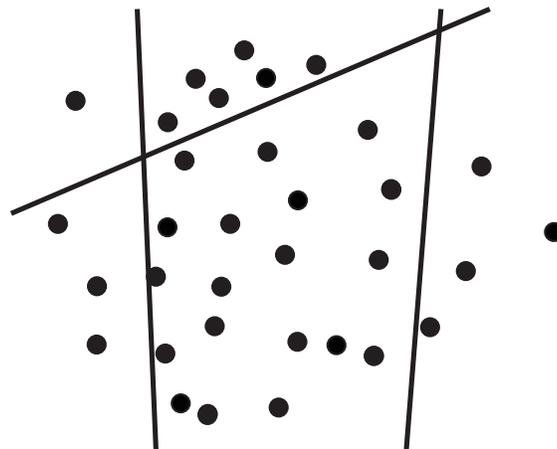


Abbildung 3.1: eine endliche Punktmenge und sechs Halbebenen

vorgelegt. Die beiden folgenden Definitionen führen ε -Netze und ε -Approximationen ein. Haussler und Welzl verwenden leicht abweichende Bezeichnungen:

Während hier ähnlich zu Matoušek ε -Netze und ε -Approximationen für endliche Bereichsräume definiert werden, definieren Haussler und Welzl ε -Netze und ε -Approximationen für endliche Teilmengen von Bereichsräumen. Unter Beachtung der Definition von Unterraum sind diese Vorgehensweisen jedoch äquivalent.

3.1.6 Definition (ε -Netz). Eine Teilmenge $N \subset X$ mit endlichem X heißt ε -Netz von X , wenn jeder große Bereich $R \in \mathcal{R}_\varepsilon$ ein $x \in N$ enthält. Nach Definition 3.1.2 ist \mathcal{R}_ε die Menge der Bereiche R mit $|R| > \varepsilon|X|$. Damit erhält man die formale Definition:

$$N \subset X \text{ ist } \varepsilon\text{-Netz} \iff \forall R \in \mathcal{R}_\varepsilon: |N \cap R| > 0.$$

Äquivalent dazu ist: Eine Teilmenge $N \subset X$ heißt ε -Netz von X , wenn jeder Bereich R , der kein Element aus N enthält, höchstens $\varepsilon|X|$ Elemente aus X enthält, also $R \notin \mathcal{R}_\varepsilon$ ist:

$$N \subset X \text{ ist } \varepsilon\text{-Netz} \iff (|N \cap R| = 0 \Rightarrow |R| \leq \varepsilon|X|).$$

Wir deuten nun ein Beispiel für ein ε -Netz an:

3.1.7 Beispiel (ε -Netz in Bezug auf Halbebenen). Betrachten wir die Abbildung 3.2 und bezeichnen die Punktmenge mit A . Bei Wahl eines geeigneten ε (geschätzt $\frac{1}{5}$) ist in jeder Halbebene, die mehr $\varepsilon \cdot |A|$ Punkte enthält, ein markierter Punkt. Man beachte, dass nur einige dieser Halbebenen dargestellt werden. Mit einem geeignet gewählten ε ist die Menge $N \subset A$ der markierten Punkte ein ε -Netz in Bezug auf Halbebenen, d.h. ein ε -Netz von A im weiter oben betrachteten Bereichsraum $(A, \mathcal{R}_{|A|})$.

Es folgt die Definition eines den ε -Approximationen verwandten Mengensystems:

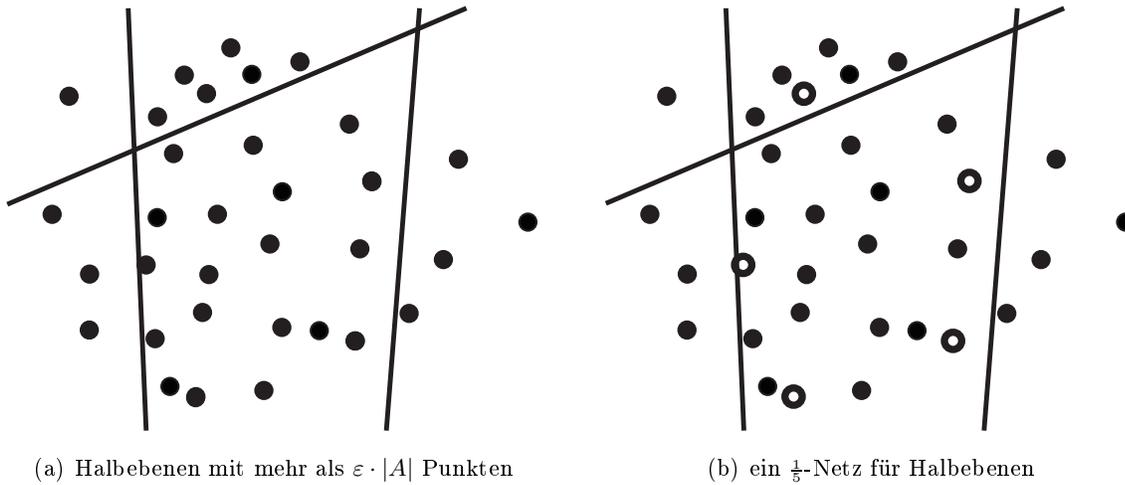


Abbildung 3.2: Beispiel eines ε -Netzes

3.1.8 Definition (ε -Approximation). Eine Teilmenge $V \subset X$ heißt ε -Approximation, wenn für alle Bereiche $R \in \mathcal{R}$ gilt:

$$\left| \frac{|V \cap R|}{|V|} - \frac{|R|}{|X|} \right| \leq \varepsilon.$$

Eine Teilmenge $V \subset X$ heißt also ε -Approximation, wenn V die Menge X in folgendem Sinne für alle Bereiche R approximiert: Für beliebige $R \in \mathcal{R}$ weicht der Bruchteil der in der „Stichprobenmenge“ V enthaltenen Elemente um höchstens ε vom Bruchteil der in der Grundmenge X ab.

ε -Approximationen wurden historisch vor ε -Netzen eingeführt. Diese beiden Mengen in Bereichsräumen stehen in enger Beziehung zueinander. Der einfachste, triviale Zusammenhang findet sich in der folgenden Beobachtung:

3.1.9 Beobachtung. Sei $V \subset X$ eine ε -Approximation. Dann ist V auch ein ε -Netz.

Beweis. Sei $R \in \mathcal{R}_\varepsilon$, also $|R|/|X| > \varepsilon$. Nach der Definition für ε -Approximationen folgt:

$$\left| \frac{|V \cap R|}{|V|} - \frac{|R|}{|X|} \right| \leq \varepsilon \implies \left| \frac{|V \cap R|}{|V|} \right| > 0 \implies |V \cap R| > 0. \quad \square$$

Umgekehrt gilt dies nicht: Hinzufügen beliebiger Elemente zu einem ε -Netz ergibt wieder ein ε -Netz, für ε -Approximationen ist dies offensichtlich nicht der Fall.

Mit ε -Approximationen lässt sich ein einfacher Algorithmus für approximative Bereichsanfragen konstruieren. Sei zum Beispiel (\mathbb{R}^2, H^*) der Bereichsraum der Halbebenen im \mathbb{R}^2 und (A, \mathcal{R}) der von einer endlichen Menge A erzeugte Unterraum. Nehmen wir an, dass sich zu einem $\varepsilon > 0$ eine ε -Approximation $V \subset A$ bestimmen lässt. Dann lässt sich mit Hilfe einer solchen ε -Approximation zu einer Anfrage-Halbebene h^* mit beliebig kleinem relativen Fehler der gesuchte Wert $|A \cap h^*|$ bestimmen.

Eingabe: Halbebene h^*

- 1: wähle ein $\varepsilon > 0$
- 2: bestimme eine ε -Approximation $V \subset A$ mit möglichst kleinem $|V|$
- 3: bestimme $|V \cap h^*|$
- 4: setze $i \leftarrow \frac{1}{|V|} \cdot |V \cap h^*|$

Ausgabe: $i \cdot |A|$

Algorithmus 3.1: Approximative Bereichsanfragen für eine endliche Menge A

$$\text{Es gilt: } \left| i - \frac{z}{|A|} \right| \leq \varepsilon \implies |i \cdot |A| - z| \leq \varepsilon \cdot |A|,$$

das heißt die Ausgabe $i \cdot |A|$ von Algorithmus 3.1 ist ungefähr der gesuchte Wert $|A \cap h^*|$.

Die Abweichung hängt dabei von dem gewählten $\varepsilon > 0$ ab.

Falls $|V|$ deutlich kleiner als $|A|$ ist, so ist die Berechnung von $i \cdot |A|$, d.h. die Bearbeitung der approximativen Bereichsanfrage, deutlich schneller als die Berechnung von $|A \cap h^*|$.

3.2 Existenz von kleinen ε -Netzen

In diesem und den folgenden Abschnitten befassen wir uns mit der Existenz von kleinen ε -Netzen und ε -Approximationen sowie der Wahrscheinlichkeit, dass eine „zufällige Teilmenge“ ein ε -Netz bzw. eine ε -Approximation ist. Für Existenzbeweise wendet man in der Regel die probabilistische Methode an, d.h. man wählt aus der Menge X zufällige m -Stichproben \bar{x} und weist nach, dass diese (als Menge) bei einer hinreichenden Stichproben-Größe m ein ε -Netz bzw. eine ε -Approximation bilden.

Obwohl zufällige m -Stichproben keine Mengen sind, ist dies wesentlich fruchtbarer als die „zufällige“ Auswahl von Teilmengen, da der zugrundeliegende Wahrscheinlichkeitsraum viel leichter zu handhaben ist.

Wir erinnern kurz an die Definition 2.1.23 (S.16) einer zufälligen m -Stichprobe

$\bar{x} = (x_1, \dots, x_m) \in X^m$ auf der endlichen Menge X . Es ist zu beachten, dass wir in diesem Kapitel für ein m -Tupel die Schreibweise \bar{x} verwenden, um $\bar{x} \in X^m$ von $x \in X$ zu unterscheiden. Man verwendet die beiden Wahrscheinlichkeitsmaße \mathbf{P} auf X und \mathbf{P}^m auf X^m mit:

$$\begin{aligned} \mathbf{P}[x] &= \frac{1}{|X|} \text{ für alle } x \in X, \\ \mathbf{P}[x \in R] &= \frac{|R|}{|X|} \text{ für alle } R \subset X, \\ \mathbf{P}^m[\bar{x}] &= \frac{1}{|X|^m} \text{ für alle } \bar{x} \in X^m \text{ und} \\ \mathbf{P}^m[\bar{x} \in \mathcal{E}] &= \frac{|\mathcal{E}|}{|X|^m} \text{ für alle } \mathcal{E} \subset X^m. \end{aligned}$$

Die sehr angenehme Eigenschaft des Wahrscheinlichkeitsraumes (X^m, \mathbf{P}^m) ist, dass die Ereignisse $[x_j \in R]$ und $[x_k \in R']$ für $j \neq k$ und beliebige Mengen $R, R' \subset X$ stets unabhängig sind. Dies ist inhaltlich völlig klar, da die Ziehung eines Elementes unabhängig von den vorigen und nachfolgenden Elementen ist. Die formale Darstellung ist jedoch etwas mühsam. Man verwendet Gleichung 3.2 und die Definition des Wahrscheinlichkeitsmaßes \mathbf{P}^m auf dem Produktraum (X^m, \mathbf{P}^m) . Zunächst gilt:

$$\mathbf{P}^m[x_j \in R] = \left(\prod_{i \neq j} \mathbf{P}[X] \right) \cdot \mathbf{P}[x_j \in R] = \left(\prod_{i \neq j} 1 \right) \cdot \mathbf{P}[x_j \in R] = \mathbf{P}[x_j \in R].$$

Wir weisen erneut (vgl. Bemerkung 2.1.22, S.16) darauf hin, dass die Schreibweise \mathbf{P}^m impliziert, dass $[x_j \in R]$ im Term $\mathbf{P}^m[x_j \in R]$ ein Ereignis \mathcal{E} aus $X^m = X \times \dots \times X$ darstellt und schreiben dieses zur Klarstellung ausnahmsweise explizit hin:

$$\mathcal{E} = \{\bar{x} \in X^m : x_j \in R\} = \left(\prod_{i=1}^{j-1} X \right) \times \{x_j\} \times \left(\prod_{i=j+1}^m X \right), \mathcal{E} \subset X^m.$$

Wir zeigen nun, dass auch $\mathbf{P}^m[x_j \in R | x_k \in R'] = \mathbf{P}[x_j \in R]$ gilt:

$$\begin{aligned} \mathbf{P}^m[x_j \in R | x_k \in R'] &= \frac{\mathbf{P}^m[x_j \in R \wedge x_k \in R']}{\mathbf{P}^m[x_k \in R']} \\ &= \frac{\left(\prod_{i \notin \{j, k\}} \mathbf{P}[X] \right) \cdot \mathbf{P}[x_j \in R] \cdot \mathbf{P}[x_k \in R']}{\left(\prod_{i \neq k} \mathbf{P}[X] \right) \cdot \mathbf{P}[x_k \in R']} \\ &= \frac{1 \cdot \mathbf{P}[x_j \in R] \cdot \mathbf{P}[x_k \in R']}{1 \cdot \mathbf{P}[x_k \in R']} \\ &= \mathbf{P}[x_j \in R] = \mathbf{P}^m[x_j \in R]. \end{aligned}$$

Die erste Umformung ergibt sich aus der Definition 2.1.9 (S.12) der bedingten Wahrscheinlichkeit. Die weiteren Umformungen ergeben sich sofort aus der Definition 2.1.7 des Produktraumes und der Definition 2.1.23 (S.16) von m -Stichproben.

Zwecks einer weniger formalen Darstellung verwenden wir einige Sprechweisen mit offensichtlichen Bedeutungen. Einige von diesen listen wir hier auf.

3.2.1 Sprechweisen (für Stichproben aus Bereichsräumen).

\bar{x} ist ε -Netz	Die Menge $\{x_1, \dots, x_m\}$ (ohne Duplikate) ist ein ε -Netz von X .
\bar{x} ist kein ε -Netz	Die Menge $\{x_1, \dots, x_m\}$ (ohne Duplikate) ist kein ε -Netz von X .
x trifft R	$x \in R$ (für einen Bereich $R \subset X$).
\bar{x} trifft R	$\exists i = 1, \dots, m: x_i \in R$ $\hat{=}$ Es gibt ein x_i in \bar{x} , welches R trifft.
x verfehlt R	$x \notin R$ (für einen Bereich $R \subset X$).
\bar{x} verfehlt R	$x_1 \notin R \wedge x_2 \notin R \wedge \dots \wedge x_m \notin R$ ($R \subset X$) $\hat{=}$ Es gibt kein x_i in \bar{x} , welches R trifft.

3.2.2 Satz (Existenz nach Matoušek [Mat93]).

Sei $1 > \varepsilon > 0$ und ein endlicher Bereichsraum (X, \mathcal{R}) gegeben. Man setze $s = |\mathcal{R}|$.

Dann gilt für diesen Bereichsraum:

$$\text{Es existiert ein } \varepsilon\text{-Netz } N \text{ von } X \text{ mit höchstens } \frac{1}{\varepsilon} \log s \text{ Elementen.}$$

Konvention. Wenn ein Term t mit reellen Größen wie ε eine Anzahl beschreibt, so ist stets die nach oben gerundete natürliche Zahl $\lceil t \rceil$ gemeint. Die Rundungsoperation wird dann nicht mitnotiert um dem Term ein einfacheres Aussehen zu geben. In diesem Sinne lässt sich beispielsweise sagen: Es gibt eine Teilmenge $N \subset X$ mit $\nu = \frac{1}{\varepsilon} \log s$ Elementen, die ein ε -Netz von X ist. (Man beachte: Wenn $N' \subset N$ mit $\nu' < \nu$ ein ε -Netz ist, so ist auch nach der Definition auch N ein ε -Netz.)

Beweis (Nach Matoušek [Mat93]). Betrachte eine m -Stichprobe $\bar{x} = (x_1, \dots, x_m)$ von X geschrieben als m -Tupel $\bar{x} \in X^m$. Nach Definition 2.1.23 ist \mathbf{P} die diskrete Gleichverteilung auf X und \mathbf{P}^m die diskrete Gleichverteilung auf X^m .

Wir betrachten zunächst einen großer Bereich $R \in \mathcal{R}_\varepsilon$. Für $x \in \bar{x}$ gilt:

$$\mathbf{P}[x \notin R] = \frac{|X| - |R|}{|X|} < 1 - \varepsilon.$$

Daraus folgt:

$$\mathbf{P}^m[\forall x \in \bar{x}: x \notin R] = \prod_{i=1}^m \mathbf{P}[x \notin R] < (1 - \varepsilon)^m.$$

Mit der leichten technischen Abschätzung $1 - \varepsilon \leq e^{-\varepsilon m}$ folgt:

$$\mathbf{P}^m[\bar{x} \text{ verfehlt } R] = \mathbf{P}^m[\forall x \in \bar{x}: x \notin R] < e^{-\varepsilon m}.$$

Man gelangt nun zu einer Abschätzung für $\mathbf{P}^m[\bar{x} \text{ ist } \varepsilon\text{-Netz}]$:

$$\begin{aligned} \mathbf{P}^m[\exists R \in \mathcal{R}_\varepsilon: \bar{x} \text{ verfehlt } R] &< |\mathcal{R}_\varepsilon| \cdot e^{-\varepsilon m} \leq s \cdot e^{-\varepsilon m} \\ \implies \mathbf{P}^m[\bar{x} \text{ ist } \varepsilon\text{-Netz}] &= 1 - \mathbf{P}^m[\exists R \in \mathcal{R}_\varepsilon: \bar{x} \text{ verfehlt } R] > 1 - s \cdot e^{-\varepsilon m}. \end{aligned}$$

Für $m \geq \frac{1}{\varepsilon} \log s$ folgt:

$$\begin{aligned} \mathbf{P}^m[\bar{x} \text{ ist } \varepsilon\text{-Netz}] &> 1 - s \cdot e^{-\varepsilon \cdot \frac{1}{\varepsilon} \log s} = 1 - s \cdot e^{-\log s} = 1 - \frac{s}{s} = 0 \\ \implies \exists \bar{x} \in X^m: \bar{x} &\text{ ist ein } \varepsilon\text{-Netz.} \end{aligned}$$

Im letzten Beweisschritt wurde nur die Definition des endlichen Wahrscheinlichkeitsmaßes \mathbf{P}^m verwendet: Für $p = \mathbf{P}^m[\mathcal{E}] > 0$ ist $|\mathcal{E}| = p \cdot |X|^m > 0$. Diese Folgerung ist eine Anwendung der probabilistischen Methode (vgl. Abschnitt 2.3, S.23). \square

Für die von Haussler und Welzl verwendete Datenstruktur ist es allerdings nötig ein ε -Netz zu finden, dessen Größe nicht von der Größe der Objektmenge ist. Ein solches ε -Netz hat in diesem Sinne eine konstante Größe. Solche ε -Netze existieren nur für gewisse Bereichsräume. Dies ist eine kombinatorische Eigenschaft, die nach Vapnik und Chervonenkis als VC-Dimension bezeichnet wird [VC71].

3.3 Die VC-Dimension von Bereichsräumen

Wir beginnen diesen Abschnitt mit der Definition der Vapnik-Chervonenkis-Dimension (kurz: Dimension oder VC-Dimension) eines Bereichsraumes und geben anschließend ein Beispiel für einen Bereichsraum mit einer endlichen sowie ein Beispiel für einen Bereichsraum mit einer unendlichen VC-Dimension an.

Anschließend wird in Lemma 3.3.4 gezeigt, dass für einen Bereichsraum mit endlicher VC-Dimension d auch die Unterräume eine endliche Dimension kleiner oder gleich d haben.

Im Satz 3.3.7 (S.43) wird eine obere Schranke für die Anzahl der Bereiche in einem endlichen Unterraum $(A, \mathcal{R}|_A)$ (endlich: $\Leftrightarrow |A| < \infty$) eines Bereichsraumes mit endlicher VC-Dimension angegeben. Diese Abschätzung bietet die Grundlage für die beiden folgenden Abschnitte, in denen die Existenz von ε -Approximationen und ε -Netzen konstanter Größe nachgewiesen wird.

3.3.1 Definition (VC-Dimension). Gegeben sei ein Bereichsraum (X, \mathcal{R}) . Dann ist seine VC-Dimension (kurz: Dimension) $\dim(X, \mathcal{R})$ definiert durch:

$$\dim(X, \mathcal{R}) = \begin{cases} \max\{|A|: A \subset X \text{ endlich und } \mathcal{R}|_A = \mathcal{P}(A)\}, & \text{bzw.} \\ \infty, & \text{falls das obige Maximum nicht existiert.} \end{cases}$$

Dies ist die größte Anzahl von Elementen, die eine endliche Menge $A \subset X$ haben kann, so dass folgende Eigenschaft erfüllt ist: Zu einer beliebigen Teilmenge $A' \subset A$ gibt es einen Bereich R aus \mathcal{R} mit $R \cap A = A'$. Falls eine endliche Teilmenge A diese Eigenschaft erfüllt, sagen wir auch, dass A durch \mathcal{R} bzw. (X, \mathcal{R}) überdeckt (*shattered*) wird.

Falls beliebig große Mengen A diese Eigenschaft erfüllen, existiert keine größte Anzahl von Elementen und ist die Dimension unendlich. Es lässt sich auch $\dim(\mathcal{R})$ statt $\dim(X, \mathcal{R})$ schreiben, wenn klar ist, auf welchen Bereichsraum sich \mathcal{R} bezieht.

3.3.2 Beispiel (elementarer Bereichsraum mit Dimension 2). Man betrachte den Bereichsraum der abgeschlossenen Intervalle auf den reellen Zahlen. Dieser ist $(\mathbb{R}, \mathcal{R})$, $\mathcal{R} = \{R: R = [a, b] \text{ mit } a, b \in \mathbb{R}\}$. Man erkennt leicht, dass $\dim(\mathbb{R}, \mathcal{R}) = 2$ gilt:

Sei $A = \{x_1, x_2\}$ mit $x_1 + 1 < x_2$. Dann gilt für die Bereiche $R_1 = [x_1, x_2]$, $R_2 = [x_1, x_1 + 1]$,

$R_3 = [x_1 + 1, x_2]$ und $R_4 = [x_1 - 2, x_1 - 1]$: $R_1 \cap A = \{x_1, x_2\}$, $R_2 \cap A = \{x_1\}$, $R_3 \cap A = \{x_2\}$ und $R_4 \cap A = \emptyset$. Also gilt:

$$\mathcal{R}|_A = \{\{x_1, x_2\}, \{x_1\}, \{x_2\}, \emptyset\} = \mathcal{P}(A).$$

Sei nun $A = \{x_1, x_2, x_3\}$ eine dreielementige Menge und o.B.d.A. $x_1 < x_2 < x_3$. Diese lässt sich nie überdecken, da es keinen Bereich $R = [a, b]$ gibt mit $R \cap A = \{x_1, x_3\}$. Also gilt für $|A| \geq 3$ stets $\mathcal{R}|_A \neq \mathcal{P}(A)$.

Im nächsten Kapitel (Lemma 4.3.2, S.69) wird gezeigt, dass der Bereichsraum der Halbebenen im \mathbb{R}^2 bzw. der Bereichsraum der Halbräume im \mathbb{R}^d (sowie deren Unterräume) höchstens die Dimension $2 + 1$ bzw. $d + 1$ haben.

Dies ist nicht selbstverständlich, da es auch elementare Bereichsräume gibt, die keine endliche Dimension haben:

3.3.3 Beispiel (elementarer Bereichsraum mit Dimension ∞). Der Bereichsraum $(\mathbb{R}^2, \mathcal{R})$ mit $\mathcal{R} = \{R \subset \mathbb{R}^2: R \text{ konvex}\}$ hat keine endliche Dimension: Sei $n \in \mathbb{N}$ und $A = \{x_1, x_2, \dots, x_{n+1}\}$ ein konvexes Polygon mit $n + 1$ Eckpunkten. Dann wird A überdeckt, wie man leicht erkennt: Sei $A' \subset A$. Sei $R = CH(A') \in \mathcal{R}$ die konvexe Hülle von A' . Da A ein konvexes Polygon ist, folgt $R \cap A = A'$. Also gibt es endliche überdeckte Mengen A mit beliebig vielen Elementen und es gilt:

$$\dim(\mathbb{R}^2, \{R \subset \mathbb{R}^2: R \text{ konvex}\}) = \infty.$$

Direkt aus der Definition der VC-Dimension ergibt sich, dass sich die VC-Dimension auf Unterräume vererbt:

3.3.4 Lemma (Vererbung der VC-Dimension). *Sei (X, \mathcal{R}) ein Bereichsraum mit $\dim(X, \mathcal{R}) = d$. Dann hat auch der von Y induzierte Unterraum eine endliche VC-Dimension und diese ist kleiner oder gleich d .*

Beweis. Annahme: Der von $Y \subset X$ induzierte Bereichsraum $(Y, \mathcal{R}|_Y)$ hat eine VC-Dimension $\geq d + 1$ (oder unendlich). Dann gibt es eine von $\mathcal{R}|_Y = \mathcal{R} \cap Y$ überdeckte Teilmenge $A \subset Y$ mit $|A| \geq d + 1$. Für jeden Bereich $R \in \mathcal{R}$ gilt wegen $A \subset Y$ aber $R \cap A = (R \cap Y) \cap A$ und es ist $R \cap Y \in \mathcal{R}|_Y$. Damit wird $A \subset X$ auch von (X, \mathcal{R}) überdeckt und es folgt: Die VC-Dimension von (X, \mathcal{R}) ist größer oder gleich $d + 1$. Dies ist im Widerspruch zur Annahme. \square

Die entscheidende Konsequenz endlicher VC-Dimension d für einen Bereichsraum (X, \mathcal{R}) ist, dass die Anzahl der Bereiche durch ein festes Polynom in $|X|$ vom Grad d beschränkt ist. Diese Funktion auf den natürlichen Zahlen definieren wir zunächst:

3.3.5 Definition. $\Phi_d: \mathbb{N} \rightarrow \mathbb{N}$ sei für $d \in \mathbb{Z}$ definiert durch

$$\Phi_d(n) = \sum_{k=0}^d \binom{n}{k}.$$

Konvention: $\binom{n}{k} = 0$ für $k > n$, so dass $\Phi_d(n) := \sum_{k=0}^n \binom{n}{k} = 2^n$ für $d > n$.

Im Falle von $d = 2$ ergibt sich für $n \geq 2$: $\Phi_2(n) = 1 + n + n(n-1)/2$. Dies ist tatsächlich ein Polynom 2-ten Grades in n . Ein mit Arrangements vertrauter Leser kennt dieses Polynom als die größte Anzahl von Flächen, in welche die Ebene durch eine Menge aus n Geraden partitioniert wird. Für 3 Geraden erhält man zum Beispiel $\Phi_2(3) = 1 + 3 + 3 = 7$. Dieser sehr entscheidende Zusammenhang gilt auch für d -dimensionale Arrangements und wird im nächsten Kapitel zu geometrischen Bereichsräumen diskutiert.

Da $\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!}$ ein Polynom vom Grad k in n ist, ergibt Φ_d tatsächlich ein Polynom in n vom Grad d . Haussler und Welzl definieren $\Phi_d(n)$ durch eine Rekursion und beweisen die Formel in Definition 3.3.5. Wir gehen den umgekehrten Weg und beweisen aus der obigen expliziten Definition die Rekursionsformel.

3.3.6 Lemma (Rekursion für $\Phi_d(n)$). Für $n, d \geq 1$ gilt:

$$\Phi_d(n-1) + \Phi_{d-1}(n-1) = \Phi_d(n).$$

Beweis. Wir beweisen zunächst, dass die Rekursionsformel für den Binomialkoeffizienten $\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}$ auch für $k > n-1$ (und $n \geq 1$) mit der Konvention $\binom{n}{i} = 0$ für $i > n$ erfüllt wird. Wir unterscheiden zwei Fälle:

1. $k = n$: Es gilt

$$\binom{n-1}{n-1} + \binom{n-1}{n} = 1 + 0 = 1 = \binom{n}{n}.$$

2. $k > n$: Es gilt

$$\binom{n-1}{k-1} + \binom{n-1}{k} = 0 + 0 = 0 = \binom{n}{k}.$$

Den Beweis erhält man nun durch einfaches Umformen der Definition von $\Phi_d(n)$:

$$\begin{aligned}
 \Phi_d(n-1) + \Phi_{d-1}(n-1) &= \sum_{k=0}^d \binom{n-1}{k} + \sum_{k=0}^{d-1} \binom{n-1}{k} \\
 &= \sum_{k=0}^d \binom{n-1}{k} + \sum_{k=1}^d \binom{n-1}{k-1} \\
 &= \binom{n-1}{0} + \sum_{k=1}^d \left[\binom{n-1}{k} + \binom{n-1}{k-1} \right] \\
 &= 1 + \sum_{k=1}^d \binom{n}{k} \\
 &= \binom{n}{0} + \sum_{k=1}^d \binom{n}{k} = \sum_{k=0}^d \binom{n}{k} \\
 &= \Phi_d(n).
 \end{aligned}$$

Damit ist die Aussage für alle $n, k \geq 1$ bewiesen. \square

3.3.7 Lemma (Bereichsanzahl). *Gegeben sei ein endlicher Bereichsraum (X, \mathcal{R}) mit VC-Dimension d . Dann gilt für die Anzahl $n = |\mathcal{R}|$ seiner Bereiche:*

$$|\mathcal{R}| \leq \Phi_d(n).$$

Beweis. [HW87, Lemma 3.1]

Der Beweis wird mit vollständiger Induktion nach d und n geführt.

Im Falle von $d = 0$ wird nach der Definition der VC-Dimension höchstens die leere Menge überdeckt, man erhält also $\mathcal{R} \subset \{\emptyset\}$ und damit $|\mathcal{R}| \leq 1 = \Phi_0(n)$.

Für $n = 0$ erhält man wegen $X = \emptyset$ erneut $\mathcal{R} \subset \{\emptyset\}$.

Es gilt $\Phi_d(0) = \sum_{k=0}^d \binom{0}{k} = \binom{0}{0} + 0 = 1 \geq |\mathcal{R}|$.

Äußere Induktionsannahme: Die Behauptung gilt für alle endlichen Bereichsräume mit VC-Dimension $\leq d-1$ mit einem $d \geq 1$.

Innere Induktionsannahme: Die Behauptung gilt für alle endlichen Bereichsräume mit VC-Dimension d und höchstens $n-1$ Elementen mit einem $n \geq 1$.

Induktionsbehauptung: Die Behauptung gilt für einen Bereichsraum $B = (X, \mathcal{R})$ mit VC-Dimension d und $|X| = n$.

Induktionsschritt:

Wir betrachten zwei Bereichsräume auf der $(n-1)$ -elementigen Menge $X - \{x\}$:

$$B^{-x} := (X - \{x\}, \mathcal{R}^{-x})$$

ist einer der beiden Bereichsräume auf $X - \{x\}$. Seine Bereiche $R \in \mathcal{R}^{-x}$ entstehen durch Entfernung von x aus einem jeden Bereich von \mathcal{R} , d.h. es wird definiert:

$$\mathcal{R}^{-x} := \{R - \{x\} : R \in \mathcal{R}\}.$$

B^{-x} ist damit der von der Menge $X - \{x\}$ induzierte Unterraum von (X, \mathcal{R}) .

$$B^{(x)} := (X - \{x\}, \mathcal{R}^{(x)})$$

ist der andere Bereichsraum auf der Menge $X - \{x\}$. Die Bereichsmenge ist definiert als

$$\mathcal{R}^{(x)} := \{R \in \mathcal{R} : x \notin R, R \cup \{x\} \in \mathcal{R}\}.$$

Die Bereiche $R \in \mathcal{R}^{(x)}$ sind jene Teilmengen von $X - \{x\}$, die auch in der Bereichsmenge \mathcal{R} enthalten sind und für die gleichzeitig $R \cup \{x\}$ in \mathcal{R} ist. Man beobachtet:

$$|\mathcal{R}| = |\mathcal{R}^{-x}| + |\mathcal{R}^{(x)}|.$$

Dies ergibt sich direkt aus der Konstruktion der Bereichsmengen \mathcal{R}^{-x} und $\mathcal{R}^{(x)}$, wir geben für diese Beobachtung hier dennoch den Beweis an. Jeder Bereich aus \mathcal{R} wird auf einen Bereich aus \mathcal{R}^{-x} abgebildet. Es werden offensichtlich nie mehr als zwei Bereiche aus \mathcal{R} auf denselben Bereich aus \mathcal{R}^{-x} abgebildet, da nur ein Element x entfernt wird. Zwei verschiedene Bereiche R_1, R_2 aus \mathcal{R} werden nach der Definition von \mathcal{R}^{-x} genau dann auf denselben Bereich $R \in \mathcal{R}^{-x}$ abgebildet, wenn $R = R_1$ und $R \cup \{x\} = R_2$ (oder mit vertauschten R_1, R_2). Es gilt stets $x \notin R$. Also ist diese Aussage gleichbedeutend mit $x \notin R, R \in \mathcal{R}$ und $R \cup \{x\} \in \mathcal{R}$. Damit folgt:

$$|\mathcal{R}| - |\mathcal{R}^{-x}| = |\{R \in \mathcal{R} : x \notin R, R \cup \{x\} \in \mathcal{R}\}| = |\mathcal{R}^{(x)}|.$$

Wir beweisen nun:

$$|\mathcal{R}| \leq \Phi_d(n-1) + \Phi_{d-1}(n-1) = \Phi_d(n).$$

Zunächst ist B^{-x} nach Konstruktion ein Bereichsraum mit höchstens $n-1$ Elementen und hat offensichtlich höchstens die endliche VC-Dimension d . (Gilt nach Lemma 3.3.4.) Nach der inneren Induktionsannahme gilt also $|\mathcal{R}^{-x}| \leq \Phi_d(n-1)$.

Es bleibt zu zeigen: $|\mathcal{R}^{(x)}| \leq \Phi_{d-1}(n-1)$.

Wir beweisen also, dass der Bereichsraum $B^{(x)}$ mit $n-1$ Elementen höchstens die VC-Dimension $d-1$ hat. Dies geschieht mit der Definition der VC-Dimension:

Nehmen wir an, dass eine endliche Menge A von $B^{(x)}$ überdeckt wird. Dann ist $x \notin A$.

Wir zeigen nun, dass die Menge $A \cup \{x\}$ von (X, \mathcal{R}) überdeckt wird:

Sei A' eine beliebige Teilmenge von A . Dann gibt es einen Bereich R in $\mathcal{R}^{(x)}$ mit $R \cap A = A'$. Nach der Definition von $\mathcal{R}^{(x)}$ ist sowohl $R \in \mathcal{R}^{(x)}$ als auch $R \cup \{x\} \in \mathcal{R}^{(x)}$. Es ist $R \cap A \cup \{x\} = A'$ und $(R \cup \{x\}) \cap (A \cup \{x\}) = A' \cup \{x\}$.

Also werden beliebige Teilmengen von $A \cup \{x\}$ durch einen Bereich in \mathcal{R} geschnitten, d.h. $A \cup \{x\}$ wird von \mathcal{R} überdeckt. Da die VC-Dimension von (X, \mathcal{R}) höchstens d ist, hat somit der Bereichsraum $B^{(x)} = (X - \{x\}, \mathcal{R}^{(x)})$ höchstens VC-Dimension $d-1$. Es folgt schließlich:

$$|R| \leq |\Phi_d(n-1)| + |\Phi_{d-1}(n-1)| = |\Phi_d(n)|. \quad \square$$

3.3.8 Beispiel. Betrachte den Bereichsraum (X, \mathcal{R}) mit $X = \{1, 2, 3\}$, $n := |X| = 3$ und $\mathcal{R} = \{\emptyset, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Es ist $|\mathcal{R}| = 7$. Da $\{2, 3\}$ überdeckt wird, ist seine VC-Dimension $d \geq 2$. Offensichtlich wird keine 3-elementige Menge überdeckt, da $\{1\} \notin \mathcal{R}$. (Tatsächlich ist $|R| = 2^3 - 1 = \binom{3}{0} + \binom{3}{1} + \binom{3}{2} = \Phi_2(3)$.) Betrachten wir die Bereichsräume $B^{-3} = (X - \{3\}, \mathcal{R}^{-3})$ und $B^{(3)} = (X - \{3\}, \mathcal{R}^{(3)})$. Es sind:

$$\begin{aligned} \mathcal{R} &= \{\{\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\} \\ \mathcal{R}^{-3} &= \{\{\}, \{2\}, \{\}, \{1, 2\}, \{1\}, \{2\}, \{1, 2\}\} \\ &= \{\{\}, \{1\}, \{2\}, \{1, 2\}\} \\ \mathcal{R}^{(3)} &= \{\{\}, \{2\}, \{1, 2\}, \{\}\} \\ &= \{\{\}, \{2\}, \{1, 2\}\} \end{aligned}$$

Die Mengen sind dabei bündig gesetzt, damit man erkennt, wie sich \mathcal{R}^{-3} und $\mathcal{R}^{(3)}$ ergänzen: Es ist augenscheinlich $7 = |\mathcal{R}| = |\mathcal{R}^{-3}| + |\mathcal{R}^{(3)}| = 4 + 3$. Verfolgen wir weiter die Beweisschritte: Nach der Definition der VC-Dimension hat B^{-3} VC-Dimension ≤ 2 (hier: $= 2$) und nach der Induktionsannahme folgt:

$$|R^{-3}| \leq \Phi_2(2) = \binom{2}{0} + \binom{2}{1} + \binom{2}{2} = 2^2 - 1 = 3.$$

Wir vollziehen auch nach, dass die VC-Dimension von $\mathcal{R}^{(3)}$ kleiner oder gleich 1 ist: Angenommen, $B^{(3)}$ würde eine zweielementige Menge $\{y, z\}$ überdecken (es kommt natürlich hier nur $\{1, 2\}$ in Frage), so würde (X, \mathcal{R}) die Menge $\{y, z, 3\}$ mit drei Elementen überdecken, aber (X, \mathcal{R}) hat VC-Dimension 2. Also gilt:

$$\begin{aligned} |\mathcal{R}^{(3)}| &\leq \Phi_1(2) = \binom{2}{0} + \binom{2}{1} = 2^2 - 1 = 3 \\ \Rightarrow |\mathcal{R}| &\leq \Phi_2(2) + \Phi_1(2) = \Phi_2(3). \end{aligned}$$

Mit diesem Dimensionsbegriff für Bereichsräume lassen sich ε -Netze und ε -Approximationen für endliche Mengen A finden, deren Größe nicht von $|A|$, sondern nur ε und der Dimension des Bereichsraumes abhängt.

3.4 ε -Approximationen der konstanten Größe $\mathcal{O}\left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$

Haussler und Welzl zitieren einen Satz nach Vapnik und Chervonenkis, der eine Aussage über ε -Approximationen konstanter Größe beinhaltet [VC71].

3.4.1 Satz (über ε -Approximationen nach Vapnik und Chervonenkis). *Es gibt eine Konstante $c > 0$, so dass gilt:*

Sei (X, \mathcal{R}) ein Bereichsraum der Dimension d . Seien $1 > \varepsilon > 0$ und $\delta \leq 1$. Dann gilt für

jeden endlichen Teilraum $(A, \mathcal{R}|_A)$ und eine zufällige m -Stichprobe \bar{x} von A die folgende Implikation:

$$m \geq \frac{c}{\varepsilon^2} \left(d \log \frac{d}{\varepsilon} + \log \frac{1}{\delta} \right) \implies \mathbf{P}^m[\bar{x} \text{ ist } \varepsilon\text{-Netz}] > 1 - \delta.$$

Die Aussage wird nicht bewiesen, da sie für die weiteren Ausführungen nicht relevant ist.

3.4.2 Beispiel. Der vorangegangene Satz ist nun hilfreich um das Beispiel zu approximativen Bereichsanfragen durchzuführen. Genauer gesagt bietet der Satz einen Monte-Carlo-Algorithmus für die Bestimmung einer ε -Approximation, welche Teil des Preprocessings ist. Dieser Monte-Carlo-Algorithmus hat eine Fehlerwahrscheinlichkeit von δ mit

Eingabe: Punktmenge A und Parameter m

- 1: Erzeuge eine ε -Approximation von A mit m Elementen (m wie oben).
- 2: Weise $V[i]$, $i = 1, \dots, size$ ein zufälliges Element aus A zu.
- 3: Entferne die Duplikate aus dem Vektor $V[i]$.

Ausgabe: V ▷ V stellt nun eine Menge dar und kann zurückgegeben werden.

Algorithmus 3.2: Monte-Carlo-Algorithmus zur Erzeugung einer ε -Approximation

$$\log \delta = d \cdot \log \frac{d}{\varepsilon} - \frac{\varepsilon^2 \nu}{c}.$$

3.4.3 Bemerkung. Mit ε -Approximationen lassen sich Bereichsanfragen mit beliebiger Genauigkeit beantworten, wie Haussler und Welzl bemerken [HW87, Example 5]. Für die exakte Beantwortung benötigen Haussler und Welzl jedoch ε -Netze. Das mag insofern überraschen, als jede ε -Approximation auch ein ε -Netz ist. Es zeigt sich jedoch bei tieferer Analyse, dass die von Vapnik und Chervonenkis nachgewiesenen ε -Approximationen noch zu groß sind. Betrachten wir die Größe von m in Abhängigkeit von ε , so erhalten wir:

$$m(\varepsilon) \in \mathcal{O} \left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon} \right) = \mathcal{O}(r^2 \log r).$$

Dabei ist $r := \frac{1}{\varepsilon}$. Häufig ist r eine angenehmere Variable als ε um Beziehungen auszudrücken. Daher wird für $0 < \varepsilon < 1$ häufig auch $r = \frac{1}{\varepsilon} > 0$ benutzt. Diese Schreibweise geht auf Matoušek zurück, der sie häufig verwendet. Insofern kann man auch von einem $\frac{1}{r}$ -Netz bzw. Ähnlichem sprechen und die Größenordnung wie oben angeben. Hier wird dies auch häufig getan, wenn es die Darstellung vereinfacht oder einen Zusammenhang offensichtlicher macht.

3.5 ε -Netze der konstanten Größe $\mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$

In diesem Abschnitt nach Haussler und Welzl wird eine Existenzaussage für ε -Netze bewiesen. Auch bei diesem hängt die Größe des ε -Netzes nicht von der Größe der Objektmenge ab. Sie hat in Abhängigkeit von ε die Größenordnung

$$m(\varepsilon) \in \mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right) = \mathcal{O}(r \log r).$$

Damit sind diese ε -Netze deutlich kleiner als die ε -Approximationen aus dem vorigen Abschnitt. Diese Existenzaussage von Haussler und Welzl bietet die Grundlage für ihren Datenstruktur zur Beantwortung von Halbraum- und Simplex-Bereichsanfragen. Die Größe des ε -Netzes ist abhängig von der als endlich vorauszusetzenden VC-Dimension.

Vorgriff von Theorem III: Existenz von ε -Netzen der Größe $\mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$:

Es gibt eine Konstante c_d , so dass für jeden endlichen Bereichsraum (X, \mathcal{R}) mit VC-Dimension d und alle $0 < \varepsilon < 1/2$ gilt:

$$\text{Es existiert ein } \varepsilon\text{-Netz } N \text{ von } X \text{ mit } |N| \leq c_d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}.$$

Haussler und Welzl beweisen diese Aussage mit der probabilistischen Methode durch die Betrachtung von zufälligen Stichproben. Versuchen wir zunächst plausibel zu machen, wie ein solcher Beweis funktionieren kann. In irgendeiner Form fließt die endliche VC-Dimension ein. Diese sagt aus: Für eine feste endliche Teilmenge E der Größe m' ist die Anzahl der Bereiche im induzierten Unterraum kleiner oder gleich $\Phi_d(m')$. Dies ist ein Polynom vom Grad d in m' . Damit eine zufällige Stichprobe irgendeiner Größe ein ε -Netz ist, müsste man also die Wahrscheinlichkeit mithilfe einer festen Teilmenge E der Größe m' abschätzen und zeigen, dass die Wahrscheinlichkeit für die Verfehlung eines jeden Bereiches des induzierten Unterraum ein so kleines $h(m')$ ist, dass der Term $\Phi_d(m') \cdot h(m')$ kleiner als 1 wird. Ein Anzeichen dafür wäre, dass der Grenzwert gegen 0 geht. Hierfür hinreichend wäre eine Exponentialfunktion $h(m') = 2^{-\alpha m'}$ mit einer Konstanten $\alpha > 0$. Wir wollen allerdings nicht nur die Existenz eines ε -Netzes einer beliebigen Größe nachweisen (dieses wäre trivial), sondern ein m finden, das nur von der VC-Dimension abhängt und in Abhängigkeit von $r = 1/\varepsilon$ nur die Größenordnung $\mathcal{O}(r \log r)$ hat. Daher wäre eine Abschätzung gegen $\Phi_d(m') \cdot 2^{-\alpha m'}$ mit einem möglichst großen α nützlich. Wir zeigen im folgenden Satz eine Aussage dieser Form mit $\alpha = \varepsilon m/2$. Für den Beweis des Theorems zeigen wir dann, dass für $m \geq c_d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$ mit einer Konstanten c_d der Term kleiner als 1 wird.

Eine schöne anschauliche Beschreibung des Beweisweges bietet Matoušek in *Epsilon-Nets and Computational Geometry* [Mat93]. Dort wird die gleiche Beweisidee (für eine schärfere Schranke nach Komlos, Woeginger und Pach) [KPW92] durchgeführt.

3.5.1 Satz (Wahrscheinlichkeit von ε -Netzen).

Sei (X, \mathcal{R}) ein Bereichsraum mit endlicher VC-Dimension d und $(A, \mathcal{R} \cap A)$ ein endlicher Unterraum. Sei $\varepsilon > 0$ und $m \geq 8/\varepsilon$. Betrachte eine m -Stichprobe $\bar{x} \in A^m$. Dann gilt:

$$\mathbf{P}^m[\bar{x} \text{ ist kein } \varepsilon\text{-Netz}] \leq 2 \cdot \Phi_d(2m)2^{-\varepsilon m/2}.$$

Wir beweisen die Aussage schrittweise durch eine Ungleichungskette mit den Lemmata A, B und C. Sei also für die Lemmata A, B und C stets (X, \mathcal{R}) ein Bereichsraum mit endlicher VC-Dimension d . Desweiteren wird das Hilfslemma D für Lemma B benötigt und das Hilfslemma E für Lemma C. Für $2m$ -Stichproben \bar{z} schreiben wir im Folgenden $\bar{z} = \overline{xy}$ und meinen damit

$$\bar{z} = (z_1, \dots, z_m, z_{m+1}, \dots, z_{2m}), \text{ wobei } \bar{x} = (x_1, \dots, x_m), \bar{y} = (y_1, \dots, y_m),$$

$$\text{und } z_i = \begin{cases} x_i & \text{für } 1 \leq i \leq m \\ y_{i-m} & \text{für } m < i \leq 2m \end{cases}$$

3.5.2 Sprechweisen (Ergänzung zu den Sprechweisen 3.2.1, S.38).

\bar{x} verfehlt R $x_1 \notin R, \dots, x_m \notin R$ (vgl. 3.2.1).

\bar{x} trifft R häufig $x_i \in R$ für mindestens $\varepsilon m/2$ Indizes i .

(Wird verwendet für große Bereiche R , d.h. $R \in \mathcal{R}_\varepsilon \subset X$.)

$\bar{z} = \overline{xy}$ separiert R \bar{x} trifft R häufig und \bar{y} verfehlt R .

$\bar{z} = \overline{xy}$ separiert $\exists R \in \mathcal{R}_\varepsilon$: \bar{z} separiert R .

Es wird die folgende Schreibweise für Ereignisse in X^m verwendet: Mit $\mathbf{m}[\mathcal{K}]$ wird die Menge aller m -Tupel $\bar{x} \in X^m$ bezeichnet, welche die Eigenschaft \mathcal{K} erfüllen.

$$K_\varepsilon^m := \mathbf{m}[\bar{x} \text{ ist kein } \varepsilon\text{-Netz}]$$

$$F_R^m := \mathbf{m}[\bar{x} \text{ verfehlt } R]$$

$$H_R^m := \mathbf{m}[\bar{x} \text{ trifft } R \text{ häufig}]$$

$$T_\varepsilon^{2m} := \mathbf{2m}[\bar{z} = \overline{xy} \text{ separiert}]$$

Weiter bezeichnen wir mit $L_R(\bar{x})$ die Anzahl der Treffer von \bar{x} , d.h. die Anzahl der Indizes mit $x_i \in R$. Wenn m , R oder ε im Kontext klar sind, lassen wir diese Parameter gegebenenfalls weg. Die Bezeichnungen sind an „kein ε -Netz“, **F**EHLSCHLAG, „häufig“ sowie „trennt“ angelehnt.

Dass T von ε abhängt, erklärt sich dadurch, dass nur große Bereiche $R \in \mathcal{R}_\varepsilon$ bei der Eigenschaft *separiert* betrachtet werden. Bei Vergleich mit dem Original von Haussler und Welzl ist zu beachten, dass diese andere Bezeichnung verwenden: Q_ε^m , Z_R^m sowie J_ε^{2m} .

3.5.3 Lemma (A: Abschätzung der Fehlbereiche gegen die Separationsmenge).

Sei $\varepsilon > 0$ und $m \geq \frac{8}{\varepsilon}$. Dann gilt:

$$\mathbf{P}^m[K_\varepsilon^m] < 2\mathbf{P}^{2m}[T_\varepsilon^{2m}].$$

Beweis. [HW87, Lemma 3.4]

Sei $R \in \mathcal{R}_\varepsilon$ fest. Dann betrachtet man das mehrstufige Bernoulli-Experiment aus Bemerkung 2.1.26 (S.18) mit der Teilmenge $R \subset X$ und $p = \frac{|R|}{|X|} \geq \varepsilon$. Nehmen wir zunächst $p = \varepsilon$ an. TREFFER sind Ziehungen aus dem Bereich R . Dann ist die Anzahl der TREFFER $L = L_R$ eine binomialverteilte Zufallsvariable mit Erwartungswert $\mathbf{E}[L] = mp$ und Varianz $\mathbf{var}[L] = mp(1-p)$. Nach dem Stichprobentheorem (Theorem I, S.22), welches den Satz von Chebyshev auf die Binomialverteilung anwendet, gilt dann:

$$\begin{aligned} \mathbf{P}^m[X^m - H_R^m] &\leq \mathbf{P}^m[|L - mp| > m\varepsilon/2] \\ &= \mathbf{P}^m[|L - \varepsilon m| > \varepsilon m/2] \leq \frac{\varepsilon(1-\varepsilon)m}{(\varepsilon m/2)^2} \\ &= \frac{4(1-\varepsilon)}{\varepsilon m} < \frac{4}{\varepsilon m} \leq \frac{1}{2}. \end{aligned}$$

Für $p > \varepsilon$ ist die Abweichung vom Erwartungswert noch größer.

Für jeden großen Bereich $R \in \mathcal{R}_\varepsilon$ ist also die Wahrscheinlichkeit, dass eine m -Stichprobe diesen Bereich häufig trifft, größer als $\frac{1}{2}$:

$$\mathbf{P}^m[H_R^m] > \frac{1}{2} \iff |H_R^m| > \frac{1}{2} \cdot |X|^m.$$

Sei eine $2m$ -Stichprobe gegeben, die separiert: $\bar{z} = \overline{xy} \in T_\varepsilon^{2m}$. Dann ist $\bar{x} \in K_\varepsilon^m$ und es gibt nach Definition einen Bereich $R \in \mathcal{R}_\varepsilon$ mit $\bar{y} \in H_R^m$. Also gilt:

$$|T_\varepsilon^{2m}| \geq |K_\varepsilon^m| \cdot |H_R^m| > |K_\varepsilon^m| \cdot \frac{1}{2} \cdot |X|^m.$$

Es folgt also aus der Definition der Wahrscheinlichkeitsmaße:

$$\mathbf{P}^{2m}[T_\varepsilon^{2m}] > \mathbf{P}^m[K_\varepsilon^m] \cdot \frac{1}{2}. \quad \square$$

Die Lemmata B und C beschäftigen sich mit der weiteren Abschätzung.

Notation. Wir bezeichnen die Permutationsgruppe der Indizes $1, \dots, 2m$ wie üblich mit $S_{2m} = \{\pi_1, \dots, \pi_{(2m)}!\}$ die Permutationsgruppe der Indizes $1, \dots, 2m$. Für $j = 1, \dots, (2m)!$ ist also π_j eine (verschiedene) Permutation der Indizes. $\pi_j(\bar{z}) \in X^{2m}$ ist eine Permutation von \bar{z} . Bemerkung: Bei einem Term der Form $|j: \pi_j(\bar{z}) \text{ erfüllt } \dots|$ durchlaufe j stets die Werte $1, \dots, n!$ so dass π_j die Elemente der Permutationsgruppe S_n durchläuft. Damit wird in einem solchen Term die Anzahl der Permutationen π_j aufgezählt, die eine gewisse Eigenschaft erfüllen.

3.5.4 Definition. Zu $\bar{z} \in X^{2m}$ definiere $f(\bar{z})$ als die Anzahl der Permutationen $\pi_j(\bar{z})$, die separieren, d.h.:

$$f(\bar{z}) := |j: \pi_j(\bar{z}) \in T|.$$

3.5.5 Lemma (B: Kombinationen).

$$\mathbf{P}^{2m}[T_\varepsilon^{2m}] \leq \max_{\bar{z} \in X^{2m}} \frac{f(\bar{z})}{(2m)!}.$$

Das bedeutet: Wählt man \bar{z} so, dass die Anzahl der Permutationen, die separieren, maximal ist, so ist der Bruchteil dieser Permutationen von \bar{z} größer oder gleich $\mathbf{P}^{2m}[T]$. Dies ist die Wahrscheinlichkeit, dass eine zufällige $2m$ -Stichprobe separiert. Die Aussage wird im Original mit der Bezeichnung $\Theta(\bar{x})$ für $f(\bar{z})$ formuliert [HW87, S.136, Z. 2-3].

Beweis. Wir beweisen die Aussage mit der Methode der bedingten Wahrscheinlichkeiten. Zugrundeliegend ist die Beobachtung, dass eine zufällige (gleichverteilte) Permutation einer zufälligen gleichverteilten Stichprobe ($\in X^{2m}$) wieder eine zufällige gleichverteilte Stichprobe ($\in X^{2m}$) ergibt. Diese Beobachtung werden wir weiter unten (\dagger) verwenden. Zur exakten Formulierung definieren wir uns also als Wahrscheinlichkeitsraum den Produktraum $X^{2m} \times S_{2m}$ mit der Wahrscheinlichkeit $\tilde{\mathbf{P}}^{2m} = \mathbf{P}^{2m} \times \hat{\mathbf{P}}^{2m}$. Elemente sind also Paare (\bar{x}, π) aus einer $2m$ -Stichprobe und einer Permutation $\pi \in S_{2m}$. Die Wahrscheinlichkeit eines Elementarereignisses (\bar{x}, π) ist:

$$\tilde{\mathbf{P}}^{2m}[(\bar{x}, \pi)] = \mathbf{P}^{2m}[\bar{x}] \cdot \hat{\mathbf{P}}^{2m}[\pi] = \frac{1}{X^{2m}} \cdot \frac{1}{(2m)!}.$$

Wir sind interessiert daran, ob eine zufällige Permutation einer zufälligen Stichprobe eine vorgegebene Stichprobe ergibt, also an Ereignissen der Form

$$\mathcal{E}(\bar{y}) := \{(\bar{x}, \pi) \in X^{2m} \times S_{2m} : \pi(\bar{x}) = \bar{y}\}, \quad \mathcal{E}(\bar{y}) \subset X^{2m} \times S_{2m}.$$

Nach der obigen Beobachtung (\dagger) ist

$$\tilde{\mathbf{P}}^{2m}[\mathcal{E}(\bar{y})] = \mathbf{P}^{2m}[\bar{y}]. \quad (3.1)$$

für alle $\bar{y} \in X^{2m}$. Wir beweisen diese Behauptung später im Hilfslemma 3.5.10 (S.54) jedoch auch formal. Die Aussage dieses Hilfslemmas ist die intuitiv klare Formel $|\mathcal{E}(\bar{y})| = |j : \pi_j(\bar{x}) = \bar{y}| = (2m)!$ für alle \bar{y} . Damit gilt:

$$\tilde{\mathbf{P}}^{2m}[\mathcal{E}(\bar{y})] = (2m)! \cdot \tilde{\mathbf{P}}^{2m}[(\bar{x}, \pi)] = \frac{1}{|X|^{2m}} = \mathbf{P}^{2m}[\bar{y}].$$

und es folgt unmittelbar:

$$\tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)] = \mathbf{P}^{2m}[T], \quad \text{wobei } \mathcal{E}(T) := \bigcup_{\bar{y} \in T} \mathcal{E}(\bar{y}).$$

Es gilt nach dem Satz über die totale Wahrscheinlichkeit (Lemma 2.1.10, S.13):

$$\tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)] = \sum_{\bar{z} \in X^{2m}} \tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)|\bar{x} = \bar{z}] \cdot \tilde{\mathbf{P}}^{2m}[\bar{x} = \bar{z}] = \sum_{\bar{z} \in X^{2m}} \tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)|\bar{x} = \bar{z}] \cdot \frac{1}{|X|^{2m}}.$$

Wir summieren also für jedes \bar{z} die Wahrscheinlichkeit auf, dass eine zufällige Permutation $\pi(\bar{z})$ in T liegt und multiplizieren mit der Wahrscheinlichkeit für die Bedingung, dass dieses \bar{z} gewählt wird. Nach der Methode der bedingten Wahrscheinlichkeiten (siehe Theorem II über die maximale Wahrscheinlichkeit) gibt es dann ein \bar{z} , für das diese bedingte Wahrscheinlichkeit mindestens so groß ist wie die totale Wahrscheinlichkeit:

$$\max_{\bar{z} \in X^{2m}} \tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)|\bar{x} = \bar{z}] \geq \tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)]. \quad (3.2)$$

Zu einem gegebenen \bar{z} betrachte man nun $\tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)|\bar{x} = \bar{z}]$. Inhaltlich ist recht klar, dass dies zu einem festen \bar{z} die Wahrscheinlichkeit ist, dass eine zufällige Permutation $\pi_j(\bar{z})$ in T ist, d.h. separiert. Da im Wahrscheinlichkeitsraum $(S_{2m}, \hat{\mathbf{P}}^{2m})$ zu einem $\pi_j \in S_{2m}$ jedoch das Ereignis „ $\pi_j(\bar{z}) \in T$ “ nicht formal erklärt ist, wird der folgende technische Weg notwendig: Wir bestimmen beide Wahrscheinlichkeiten durch die recht mühsame Aufzählung der Elementarereignisse im Produktraum $\tilde{\mathbf{P}}^{2m} = \mathbf{P}^{2m} \times \hat{\mathbf{P}}^{2m}$.

$$\begin{aligned} \tilde{\mathbf{P}}^{2m}[\mathcal{E}(T)|\bar{x} = \bar{z}] &\stackrel{\text{def}}{=} \frac{\tilde{\mathbf{P}}^{2m}[\mathcal{E}(T) \wedge \bar{x} = \bar{z}]}{\tilde{\mathbf{P}}^{2m}[\bar{x} = \bar{z}]} \\ &= \frac{\tilde{\mathbf{P}}^{2m}[(\bar{x}, \pi_j) \in (X^{2m} \times S_{2m}) : (\bar{x} = \bar{z}) \wedge (\pi_j(\bar{x}) \in T)]}{\tilde{\mathbf{P}}^{2m}[(\bar{x}, \pi_j) \in (X^{2m} \times S_{2m}) : (\bar{x} = \bar{z})]} \\ &= \frac{|(\bar{x}, \pi_j) \in (X^{2m} \times S_{2m}) : (\bar{x} = \bar{z}) \wedge (\pi_j(\bar{x}) \in T)|}{|X|^{2m} \cdot |S_{2m}|} \\ &= \frac{|(\bar{x}, \pi_j) \in (X^{2m} \times S_{2m}) : (\bar{x} = \bar{z})|}{|X|^{2m} \cdot |S_{2m}|} \\ &= \frac{|(\bar{x}, \pi_j) \in (X^{2m} \times S_{2m}) : (\bar{x} = \bar{z}) \wedge (\pi_j(\bar{z}) \in T)|}{|(\bar{x}, \pi_j) \in (X^{2m} \times S_{2m}) : (\bar{x} = \bar{z})|} \\ &= \frac{|j : \pi_j(\bar{z}) \in T|}{1 \cdot |S_{2m}|} \quad (\text{da } \bar{z} \text{ fest}) \\ &= \frac{f(\bar{z})}{(2m)!}. \end{aligned}$$

Wendet man nun Gleichung 3.2 und Gleichung 3.1 an, ergibt sich:

$$\mathbf{P}^{2m}[T] \leq \frac{f(\bar{z})}{(2m)!}. \quad \square$$

Die jetzt noch fehlende Aussage ist das folgende Lemma, das für ein beliebiges \bar{z} die Anzahl der Kombinationen abschätzt.

3.5.6 Lemma (Lemma C: Abschätzung der Kombinationen).

Sei \bar{z} eine $2m$ -Stichprobe aus X . Dann gilt:

$$\frac{f(\bar{z})}{(2m)!} \leq \Phi_d(2m) \cdot 2^{-\frac{\varepsilon \cdot m}{2}}.$$

Beweis. [HW87, Lemma 3.5] Sei $\bar{z} \in X^{2m}$ und $E_{\bar{z}} = E = \{z_i, i = 1, \dots, 2m\} \subset X$ die von \bar{z} gebildete Teilmenge (Treffermenge) von X . Wenn z_i einen Bereich $R \in \mathcal{R}$ trifft, so trifft z_i auch $R \cap E$.

Wir betrachten nun jedoch den vom Bereichsraum (X, \mathcal{R}) durch E induzierten Unterraum $(E, \mathcal{R}|_E)$. Damit gilt:

$$\bar{z} \text{ separiert } R \Leftrightarrow \bar{z} \text{ separiert } R \cap E.$$

Dabei ist die rechte Aussage in Bezug auf den Unterraum $(E, \mathcal{R}|_E)$ zu verstehen. Dessen Betrachtung ist nützlich, da offensichtlich jede Permutation $\pi_j(\bar{z})$ die gleiche Teilmenge $E_{\bar{z}}$ bildet. Damit gilt also für alle Permutationen $\pi_j(\bar{z})$ im Bereichsraum $(E, \mathcal{R}|_E)$:

$$\pi_j(\bar{z}) \text{ separiert } R \Leftrightarrow \pi_j(\bar{z}) \text{ separiert } R \cap E.$$

Da (X, \mathcal{R}) ein Bereichsraum mit der endlichen VC-Dimension d ist, hat nach dem Vererbungslemma 3.3.4 (S.41) auch der Unterraum $(E, \mathcal{R}|_E)$ höchstens die endliche VC-Dimension d . Weil \bar{z} eine $2m$ -Stichprobe ist, gilt $|E| \leq 2m$. Damit lässt sich nun die entscheidende Abschätzung für die Anzahl der Bereiche $R \cap E \in \mathcal{R}|_E$ durchführen. Nach dem Lemma 3.3.7 (S.43) über die Abschätzung der Bereiche eines Bereichsraumes mit der VC-Dimension d gilt:

$$|\tilde{\mathcal{R}}| \leq \Phi_d(|E|) \leq \Phi_d(2m).$$

Mithilfe der obigen Betrachtungen lässt sich $f(\bar{z})$ nun abschätzen:

$$\begin{aligned} f(\bar{z}) &= |j: \pi_j(\bar{z}) \text{ separiert}| \\ &= |j: \exists R \in \mathcal{R}_\varepsilon: \pi_j(\bar{z}) \text{ separiert } R| \\ &= |j: \exists R \in \mathcal{R}_\varepsilon: \pi_j(\bar{z}) \text{ separiert } R \cap E| \\ &\leq \sum_{\tilde{R} \in (\mathcal{R}_\varepsilon \cap E)} |j: \pi_j(\bar{z}) \text{ separiert } \tilde{R}| \\ &\leq \sum_{\tilde{R} \in (\mathcal{R}_\varepsilon \cap E)} (2m)! \cdot 2^{-\varepsilon m/2} \\ &\leq \sum_{\tilde{R} \in (\tilde{\mathcal{R}})} (2m)! \cdot 2^{-\varepsilon m/2} \\ &\leq \Phi_d(2m) \cdot (2m)! \cdot 2^{-\varepsilon m/2}. \end{aligned} \tag{3.3}$$

Dabei wurde für die Ungleichung 3.3 das im Anschluss bewiesene Hilfslemma 3.5.7 verwendet und beachtet, dass für $R \in \mathcal{R}_\varepsilon$ die Trefferanzahl l von \bar{z} für $R \in \mathcal{R}_\varepsilon$ größer oder gleich $\frac{\varepsilon m}{2}$ ist, wenn es eine separierte Permutation $\pi_j(z)$ gibt. Damit ist auch für $\tilde{R} \in (\mathcal{R}_\varepsilon \cap E)$ die Trefferanzahl $l \geq \frac{\varepsilon m}{2}$.

Es folgt also die Behauptung:

$$\frac{f(\bar{z})}{(2m)!} \leq \Phi_d(2m) \cdot 2^{-\varepsilon m/2}. \quad \square$$

3.5.7 Lemma (D: Kombinatorisches Hilfslemma zu B).

Sei zu einem Bereich $\tilde{R} \in \mathcal{R}|_E = (\mathcal{R} \cap E)$ die Anzahl der Treffer von \bar{z} gegeben durch $l \geq \frac{\varepsilon m}{2}$. Sei $g(\bar{z})$ die Anzahl der Permutationen von \bar{z} , so dass \bar{z} durch \tilde{R} separiert wird. Dann gilt:

$$\frac{g(\bar{z})}{(2m)!} = \frac{\binom{m}{l}}{\binom{2m}{l}} \leq 2^{-l}.$$

3.5.8 Bemerkung. $\binom{m}{l}$ ist die Anzahl der Möglichkeiten, l Elemente (hier: Treffer) aus (den zweiten) m Positionen auszuwählen. $\binom{2m}{l}$ ist die Anzahl der Möglichkeiten, l Elemente (Treffer) aus (allen) $2m$ Positionen auszuwählen. Haussler und Welzl stellen schlicht fest, dass der Bruchteil der separierten Permutationen $g(\bar{z})$ gleich dem angegebenen Bruch ist. Einen alternativen Beweis mit einer Interpretation des Bruches als Wahrscheinlichkeit führen wir anschließend. Beim folgenden Beweis bestimmen wir zunächst $g(\bar{z})$. Er mag je nach persönlicher Vorliebe geradliniger erscheinen.

Beweis. Wir definieren für $\bar{x} \in X^{2m}$ die Treffermenge

$$I(\bar{x}) := \{i: [x_i \text{ trifft } R]\} \subset \{1, \dots, 2m\}.$$

Es gilt: Eine Permutation $\pi_j(\bar{z})$ separiert einen Bereich \tilde{R} aus $\mathcal{R} \cap E_{\bar{z}}$ nur, wenn $I(\pi_j(\bar{z}))$ ausschließlich Indizes größer oder gleich $(m+1)$ enthält. Dann separiert $\pi_j(\bar{z})$ den Bereich \tilde{R} tatsächlich, da $|I(\pi_j(\bar{z}))| = l \geq \frac{\varepsilon m}{2}$. Die Anzahl $g(\bar{z})$ zu bestimmen, ist damit ein kombinatorisches Auswahlproblem, das nun gelöst wird. Damit eine Permutation $\pi_j(\bar{z})$ durch R separiert wird, müssen sämtliche Treffer in der zweiten Hälfte von $\pi_j(\bar{z})$ liegen. Es sind l Treffer aus m Positionen auszuwählen. Dies geschieht auf $\binom{m}{l}$ Weisen. D.h. es gibt $\binom{m}{l}$ Wahlmöglichkeiten, die Trefferpositionen zulässig festzusetzen. Für jede dieser Wahlmöglichkeiten ist die Anzahl der Permutationen zu bestimmen. Betrachte eine feste Wahl der Treffer-Positionen. Dann gibt es $l!$ Möglichkeiten, die l Treffer beliebig auf den l Trefferpositionen anzuordnen. Unabhängig davon gibt es $(2m-l)!$ Möglichkeiten, die übrigen $(2m-l)$ Elemente beliebig auf den $(2m-l)$ Fehlschlagpositionen anzuordnen. Damit ist die Anzahl der Permutationen, die diese feste Wahl der Treffer-Positionen erfüllt, stets durch $l!(2m-l)!$ gegeben. Damit ist die Anzahl der Permutationen $g(\bar{z})$ gegeben durch das Produkt:

$$g(\bar{z}) = \binom{m}{l} \cdot l!(2m-l)! .$$

Die gesuchte Formel erhält man nun schnell. Man wendet lediglich die Definition des Binomialkoeffizienten $\binom{2m}{l}$ an:

$$\frac{g(\bar{z})}{(2m)!} = \binom{m}{l} \cdot \frac{l!(2m-l)!}{(2m)!} = \frac{\binom{m}{l}}{\binom{2m}{l}}. \quad \square$$

Wir führen nun einen alternativen Beweis durch probabilistische Interpretation, da dieser die Bedeutung des Bruches $\frac{\binom{m}{l}}{\binom{2m}{l}}$ schöner verdeutlicht.

Beweis. Zu einer Permutation \bar{x} sei $I(\bar{x})$ die Menge der Trefferpositionen (Indizes). Man betrachte alle solchen $A = \{i_1, \dots, i_l\}$ und bezeichne sie als Trefferwahl. Um $\frac{\binom{m}{l}}{\binom{2m}{l}}$ als Wahrscheinlichkeit zu interpretieren, soll die Größe des Wahrscheinlichkeitsraumes $\binom{2m}{l}$ sein. Als Ergebnismenge Ω_\star verwenden wir daher die Menge der Trefferwahlen und als Wahrscheinlichkeitsmaß \mathbf{P}_\star die Gleichverteilung. Damit ist $|\Omega_\star| = \binom{2m}{l}$ die Anzahl der Auswahlen von l Trefferpositionen aus $2m$ Indizes. Wir bezeichnen eine Trefferwahl A als zulässig, wenn eine Permutation \bar{x} mit $A(\bar{x}) = A$ den Bereich R separiert. Die Anzahl der zulässigen Trefferwahlen beträgt dann $\binom{m}{l}$. Damit ergibt sich:

$$\mathbf{P}_\star[A \text{ zulässig}] = \frac{\binom{m}{l}}{\binom{2m}{l}}.$$

Jede Auswahl A hat gleich viele Elemente \bar{x} , die diese Trefferwahl realisieren, d.h. $A(\bar{x}) = A$. Diese Anzahl ist $l! \cdot (2m-l)!$, vgl. voriger Beweis. Damit entspricht die Gleichverteilung auf den Permutationen der Gleichverteilung auf den Trefferwahlen:

$$\frac{g(\bar{z})}{(2m)!} = \tilde{\mathbf{P}}[\bar{x} \text{ separiert}] = \tilde{\mathbf{P}}[A(\bar{z}) \text{ zulässig}] = \mathbf{P}_\star[A \text{ zulässig}] = \frac{\binom{m}{l}}{\binom{2m}{l}}.$$

Um die Darstellung im folgenden nicht mit zu vielen Wahrscheinlichkeitsmaßen zu überfrachten, wurde das Wahrscheinlichkeitsmaß \mathbf{P}_\star nur in diesem Beweis verwendet. Es ist im Folgenden nicht mehr zu beachten. \square

3.5.9 Bemerkung. Die Abschätzung $\frac{\binom{m}{l}}{\binom{2m}{l}} \leq 2^{-l}$ besagt, dass die Wahrscheinlichkeit bei l Treffern nur zulässige Treffer zu erhalten, kleiner gleich der Wahrscheinlichkeit ist, bei einem l -fachen Münzwurf keinen Treffer zu erhalten. Da die Anzahl $l \geq \frac{\varepsilon m}{2}$ der Treffer in einem separierten Bereich relativ groß ist, wird die Wahrscheinlichkeit der Separierung also so klein, dass auch bei Summierung über $\leq \Phi_d(2m)$ Bereiche ein Wert kleiner 1 erhalten werden kann.

3.5.10 Lemma (E: Kombinatorisches Hilfslemma zu C). Für jedes n -Tupel $\bar{z} \in X^n$ ist die Gesamtzahl der Möglichkeiten, \bar{z} als eine Permutation aus irgendeinem n -Tupel $\bar{y} \in X^n$ zu erhalten, gleich der Fakultät von n . Formal schreibt sich das als:

$$\sum_{\bar{y} \in X^n} |j : \pi_j(\bar{y}) = \bar{z}| = n!$$

Mit dieser Formel folgt unmittelbar:

$$\begin{aligned}\tilde{\mathbf{P}}^n[\mathcal{E}(\bar{y})] &= \tilde{\mathbf{P}}^n[\{(\bar{x}, \pi) \in X^n \times S_n : \pi(\bar{x}) = \bar{y}\}] \\ &= (n)! \cdot \tilde{\mathbf{P}}^n[(\bar{x}, \pi)] = \frac{1}{|X|^n} \\ &= \mathbf{P}^n[\bar{y}].\end{aligned}$$

3.5.11 Bemerkung. Für \bar{z} , die nur aus verschiedenen Elementen bestehen, ist dies offensichtlich, da zu jedem $\bar{y} \in X^n$ dann genau ein π_j existiert mit $\pi_j(\bar{y}) = \bar{z}$. Wie man leicht übersieht, ist die Aussage für \bar{z} mit mehrfach vorkommenden Elementen jedoch strenggenommen zu zeigen. Für diese \bar{z} gibt es nicht zu jedem \bar{y} entsprechende Permutationen. Falls es jedoch eine Permutation gibt, so gibt es stets derart viele, dass sich insgesamt wiederum $n!$ ergibt.

Beweis. Betrachte ein festes $\bar{z} \in X^n$ und setze $r = |j : \pi_j(\bar{z}) = \bar{z}|$.

(Die Variable r wird sich im Beweis wegekürzen. Es ist auch möglich, r konkret anzugeben:

Wenn \bar{z} Elemente aus X mit Vielfachheiten n_1, \dots, n_k , $\sum n_i = n$ enthält, ergibt sich: $r = \prod_{i=1}^k (n_i)!$

Dies ist leicht zu zeigen oder in einem Standardwerk zur Kombinatorik wie z. B. Schmitz nachzuschlagen [Sch88].)

Sei $A(\bar{z})$ die Menge aller \bar{y} so dass gilt: $\exists j : \pi_j(\bar{y}) = \bar{z}$. („Äquivalenzklasse von \bar{z} “). Es gilt also:

$$\bar{y} \notin A(\bar{z}) \Rightarrow |j : \pi_j(\bar{y}) = \bar{z}| = 0. \quad (3.4)$$

Wir betrachten nun ein $\bar{y} \in A(\bar{z})$ und ein festes $\varphi \in S_n$ mit $\phi(\bar{y}) = \bar{z}$. Es gilt die folgende Äquivalenzbeziehung:

$$(\pi_j \circ \varphi)(\bar{y}) = \bar{z} \Leftrightarrow \pi_j(\varphi(\bar{y})) = \bar{z} \Leftrightarrow \pi_j(\bar{z}) = \bar{z}.$$

Aufgrund der Eindeutigkeit der Inversen in der Gruppe S_n ist die Abbildung $t : S_n \rightarrow S_n$ definiert durch

$$t(\pi_j) = \pi_j \circ \varphi =: \varphi_j$$

eine Bijektion. Also gilt $|j : \varphi_j(\bar{y}) = \bar{z}| = |j : \pi_j(\bar{z}) = \bar{z}| = r$. Mit Umbenennung von φ_j in π_j ergibt sich:

$$\bar{y} \in A(\bar{z}) \Rightarrow |j : \pi_j(\bar{y}) = \bar{z}| = r. \quad (3.5)$$

Da jedes Element π_j ein eindeutiges Inverses π_j^{-1} hat, folgt umgekehrt auch:

$$\bar{y} \in A(\bar{z}) \Rightarrow |j : \pi_j(\bar{z}) = \bar{y}| = r.$$

Da jedes $\bar{y} \in A(\bar{z})$ genau r Mal durch $\pi_j(\bar{z})$ aufgezählt wird, gilt also:

$$|A(\bar{z})| = \frac{|j : \pi_j \in S_n|}{r} = \frac{n!}{r}. \quad (3.6)$$

Mit den Gleichungen 3.4, 3.5 und 3.6 können wir nun das Lemma beweisen:

$$\begin{aligned}
 \sum_{\bar{y} \in X^n} |j: \pi_j(\bar{y}) = \bar{z}| &= \sum_{\bar{y} \notin A(\bar{z})} |j: \pi_j(\bar{y}) = \bar{z}| + \sum_{\bar{y} \in A(\bar{z})} |j: \pi_j(\bar{y}) = \bar{z}| \\
 &= 0 + \sum_{\bar{y} \in A(\bar{z})} |j: \pi_j(\bar{y}) = \bar{z}| \\
 &= \sum_{\bar{y} \in A(\bar{z})} r = r \cdot \sum_{\bar{y} \in A(\bar{z})} 1 = r \cdot |A(\bar{z})| \\
 &= r \cdot \frac{n!}{r} = n! .
 \end{aligned}$$

Dabei haben wir zunächst X^n in $A(\bar{z})$ und dessen Komplement zerlegt und anschließend der Reihe nach die Gleichungen 3.4, 3.5 und 3.6 eingesetzt. Es gilt also wie gefordert:

$$\forall \bar{z} \in X^n: \sum_{\bar{y} \in X^n} |j: \pi_j(\bar{y}) = \bar{z}| = n! . \quad \square$$

3.5.12 Bemerkung. Man kann dieses Lemma auch als Symmetrieeigenschaft bezeichnen: Alle \bar{z} einer Äquivalenzklasse sind symmetrisch bezüglich S_n . Es gibt eine probabilistische Interpretation des Lemmas: Es ließe sich auf X^{2m} das Wahrscheinlichkeitsmaß \mathbf{P}_{\star}^{2m} mithilfe des Wahrscheinlichkeitsmaßes $\tilde{\mathbf{P}}^{2m}$ auf $X^{2m} \times S_{2m}$ definieren durch

$$\mathbf{P}_{\star}^{2m}[\bar{x}] := \tilde{\mathbf{P}}^{2m}[\mathcal{E}(\bar{x})].$$

Dann erhält man für alle $\bar{x} \in X^{2m}$ mit der Definition $\tilde{\mathbf{P}}^{2m} = \mathbf{P}^{2m} \times \hat{\mathbf{P}}^{2m}$:

$$\begin{aligned}
 X^{2m}: \quad \mathbf{P}_{\star}^{2m}[\bar{x}] &= \sum_{\bar{x} \in X^{2m}} \mathbf{P}^{2m}[\bar{x}] \cdot \frac{|j: \pi_j(\bar{y}) = \bar{z}|}{2m!} \\
 &= \mathbf{P}_{\star}^{2m}[\bar{x}] = \frac{2m!}{2m!} \cdot \mathbf{P}^{2m}[\bar{x}] \\
 &= \mathbf{P}^{2m}[\bar{x}] \\
 \implies \mathbf{P}_{\star}^{2m} &= \mathbf{P}^{2m}.
 \end{aligned}$$

Dieses Lemma beinhaltet insofern die intuitiv plausible Aussage, dass eine zufällige Permutation einer zufälligen Stichprobe die Gleichverteilung erhält (oder mathematischer ausgedrückt: respektiert). Um die Darstellung nicht mit der Notation eines weiteren Wahrscheinlichkeitsmaßes zu überfrachten, wird jedoch auf die Definition von \mathbf{P}_{\star}^{2m} verzichtet.

Beweis (von Satz 3.5.1, S.47). Sei $m \geq \frac{8}{\varepsilon}$. Dann erhält man mit den Lemmata A, B und C eine Ungleichungskette, aus der sich der Satz ergibt. Nach Lemma A (3.5.3, S.48) gilt:

$$\mathbf{P}^m[K_\varepsilon^m] \leq 2\mathbf{P}^{2m}[T_\varepsilon^{2m}].$$

Wir wählen \bar{z} so, dass $f(\bar{z})$ maximal wird. Nach Lemma B (3.5.5, S.50) gilt dann:

$$\mathbf{P}^{2m}[T] \leq \frac{f(\bar{z})}{(2m)!}.$$

Nach Lemma C (3.5.6, S.51) gilt:

$$\frac{f(\bar{z})}{(2m)!} \leq \Phi_d(2m) \cdot 2^{-\varepsilon m/2}.$$

Es ergibt sich also:

$$\mathbf{P}^m[\bar{x} \text{ ist kein } \varepsilon\text{-Netz}] = \mathbf{P}^m[K_\varepsilon^m] \leq 2\Phi_d(2m) \cdot 2^{-\varepsilon m/2}. \quad \square$$

Wir haben soeben den Satz 3.5.1 bewiesen:

In einem Bereichsraum (X, \mathcal{R}) mit der VC-Dimension d gilt für zufällige m -Stichproben x mit $m \geq 8/\varepsilon$:

$$\mathbf{P}^m[\bar{x} \text{ ist kein } \varepsilon\text{-Netz}] \leq 2\Phi_d(2m) \cdot 2^{-\varepsilon m/2}.$$

Mit diesem lässt sich durch geeignete Wahl von m die Existenz von kleinen ε -Netzen und die Wahrscheinlichkeit, dass eine kleine Stichprobe ein ε -Netz ist, abschätzen.

Wir können nun das angekündigte Theorem formulieren, das uns die Grundlage für den Beweis einer Datenstruktur mit Simplex-Bereichsanfragen liefert.

Es stellt sich allerdings auch die Frage nach der Bestimmung eines Epsilon-Netzes. In Bezug auf eine deterministische Konstruktion sieht die Antwort zunächst ernüchternd aus: Haussler und Welzl gelingt keine deterministische Konstruktion und auch Matoušek gelingt nur eine Konstruktion mit angemessener Laufzeit für den zweidimensionalen Fall. [Mat89] Angemessen sei dabei einer Laufzeit, die eine Laufzeit $\mathcal{O}(n \log n)$ für die Bestimmung der Datenstruktur ermöglicht.

Der Satz 3.5.1 ermöglicht es auch das Folgende zu zeigen: Die verwendeten Stichproben sind derart geeignet, dass für ein hinreichend großes m auch eine beliebige kleine vorgelegte Fehlerwahrscheinlichkeit für das Finden eines ε -Netzes erreicht werden kann.

Wir formulieren zunächst einen Satz, der eine Aussage über die Existenz und auch die „Wahrscheinlichkeit“ von ε -Netzen beinhaltet:

3.5.13 Satz (über die Existenz von kleinen ε -Netzen). *Sei (X, \mathcal{R}) ein endlicher Bereichsraum der Dimension d . Sei $1 > \delta > 0$. Dann gilt für eine zufällige m -Stichprobe \bar{x} von X die folgende Implikation:*

$$m \geq \max\left(\frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon}, \frac{4}{\varepsilon} \log \frac{2}{\delta}\right) \implies \mathbf{P}^m[\bar{x} \text{ ist } \varepsilon\text{-Netz}] \geq 1 - \delta.$$

Beweis. [HW87, Korollar 3.7] Da bei Vapnik und Chervonenkis eine entsprechende Aussage über ε -Approximationen bewiesen wird, zeigen auch Haussler und Welzl zunächst diese technische Abschätzung nach Vapnik und Chervonenkis. Sei

$$m = \lceil \max\left(\frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon}, \frac{4}{\varepsilon} \log \frac{2}{\delta}\right) \rceil.$$

Nach dem Satz 3.5.1 (S.47) gilt wegen $m \geq \frac{8}{\varepsilon}$:

$$\mathbf{P}^m[\bar{x} \text{ ist kein } \varepsilon\text{-Netz}] \leq 2 \cdot \Phi_d(2m)2^{-\varepsilon m/2}.$$

Wir wenden diesen Satz also an und formen elementar um:

$$\begin{aligned} 2\Phi_d(2m) &\leq \delta 2^{\varepsilon m/2} \\ \iff \frac{\varepsilon m}{2} &\geq \log \Phi_d(2m) + \log \frac{2}{\delta}. \end{aligned}$$

Aus dem rechten Term im obigen Maximum folgt: $\varepsilon m/4 \geq \log(2/\delta)$.

Zu zeigen bleibt die folgende Abschätzung für den anderen Summanden:

$$\frac{\varepsilon m}{4} \geq \log \Phi_d(2m). \quad (3.7)$$

Hinreichend ist es, die Ungleichung 3.7 für $m = \lceil \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon} \rceil$ nachzuweisen, denn dann gilt die Aussage auch für größere m .

1. Fall: $d > 1$.

Dann ist $\Phi_d(2m) \leq 2m^d$ nach der Definition von Φ_d . Mit dieser Abschätzung verbleibt zu zeigen:

$$\begin{aligned} 2d \log \frac{8d}{\varepsilon} &\geq \log 2m^d = d \log \left(\frac{16d}{\varepsilon} \log \frac{8d}{\varepsilon} \right) \\ \iff 2 \log \frac{8d}{\varepsilon} &\geq \log \left(\frac{16d}{\varepsilon} \log \frac{8d}{\varepsilon} \right) \\ \iff \left(\frac{8d}{\varepsilon} \right)^2 &\geq \frac{16d}{\varepsilon} \log \frac{8d}{\varepsilon} \\ \iff \frac{8d}{\varepsilon} &\geq 2 \log \frac{8d}{\varepsilon} \\ \iff \frac{4d}{\varepsilon} &\geq \log \frac{8d}{\varepsilon}. \end{aligned}$$

Die letzte Aussage folgt aus $4d/\varepsilon \geq 2$, $\log(2 \cdot 2) = 2$ und daraus, dass der Logarithmus im Bereich $[2, \infty]$ Steigung kleiner 1 hat.

2. Fall: $d = 1$.

Dann ist $\Phi_d(2m) = 1 + 2m$. Es bleibt zu zeigen:

$$\begin{aligned} 2 \log \frac{8}{\varepsilon} &\geq \log 1 + 2m \\ \iff 2 \log \frac{8}{\varepsilon} &\geq \log 1 + \frac{16}{\varepsilon} \log \frac{8}{\varepsilon} \quad (\text{wegen } m = \lceil \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon} \rceil) \\ \iff \frac{8^2}{\varepsilon^2} &\geq 1 + \frac{16}{\varepsilon} \log \frac{8}{\varepsilon} \\ \iff 8^2 &\geq \varepsilon^2 + 16\varepsilon \cdot \log \frac{8}{\varepsilon} \\ \iff 64 &\geq \varepsilon^2 + 16 \cdot 3 \cdot \varepsilon + 16 \cdot \varepsilon \cdot \log \frac{1}{\varepsilon}. \end{aligned}$$

Die letzte Aussage lässt sich für $0 < \varepsilon < \frac{1}{2}$ zeigen, indem man z. B. durch Kurvendiskussion $\varepsilon \log \frac{1}{\varepsilon} < \frac{7}{8}$ nachweist. Damit wurde der Satz bewiesen:

$$\mathbf{P}^m[\bar{x} \text{ ist } \varepsilon\text{-Netz}] = 1 - \mathbf{P}^m[\bar{x} \text{ ist kein } \varepsilon\text{-Netz}] \geq 1 - \delta. \quad \square$$

Zwei unmittelbare Konsequenzen des Satzes formulieren wir neu als Theorem, da diese Konsequenzen in genau dieser Form später verwendet werden.

Theorem III (ε -Netze konstanter Größe.) *Zu einem Bereichsraum (X, \mathcal{R}) der Dimension d gibt es stets eine Konstante c_d , so dass bei gegebenem $1/2 > \varepsilon > 0$ für jede endliche Menge A mit $\nu := c_d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$ stets gilt:*

1. Existenz: *Es existiert ein ε -Netz N mit $|N| \leq \nu$.*

2. Wahrscheinlichkeit: *Eine zufällige ν -Stichprobe ist mindestens mit Wahrscheinlichkeit $1/2$ ein ε -Netz.*

Das beachtliche dieser Aussage ist, dass ν nicht von $|A|$ abhängt und in Abhängigkeit des gewählten ε nur die Größenordnung $\mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$ hat.

Die Einschränkung $1/2 > \varepsilon$ hat technische Gründe. Sie stört nicht, da nur kleine ε interessant sind.

Beweis. Haussler und Welzl formulieren diese Aussage nicht als Theorem, sondern arbeiten mit der Aussage von Satz 3.5.13 über ε -Netze und bemerken im Beweis von [HW87, Lemma 4.8], dass daher eine solche Konstante existiere. Wir führen hier den Beweis.

Man wähle $\delta = 1/2$. Dann gilt für die Terme in Satz 3.5.13:

$$\frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon} \geq \frac{4}{\varepsilon} \log \frac{2}{\delta}.$$

Es ist nach Satz 3.5.13 also zu zeigen:

$$\begin{aligned} c_d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} &\geq \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon} \\ \iff c_d \cdot \log \frac{1}{\varepsilon} &\geq 8d \cdot (\log 8d + \log \frac{1}{\varepsilon}) \\ \iff \frac{c_d}{8d} \cdot \log \frac{1}{\varepsilon} &\geq \log 8d + \log \frac{1}{\varepsilon} \\ \iff \log \frac{1}{\varepsilon} \left(\frac{c_d}{8d} - 1 \right) &\geq \log 8d \\ \iff \log \frac{1}{\varepsilon} &\geq \frac{\log 8d}{\frac{c_d}{8d} - 1}. \end{aligned}$$

Da $\varepsilon < 1/2$ gilt, ist $\log 1/\varepsilon > \log 2 = 1$. Für die Konstante $c_d := 8d \log 8d + 8d$ erhalten wir:

$$\frac{\log 8d}{\frac{c_d}{8d} - 1} = \frac{\log 8d}{1 + \log 8d - 1} = 1.$$

und damit also:

$$c_d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \geq \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon} \text{ für alle } 0 < \varepsilon < \frac{1}{2}.$$

Also ist eine zufällige ν -Stichprobe \bar{x} für $\nu = c_d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$ mindestens mit Wahrscheinlichkeit $\frac{1}{2}$ ein ε -Netz. Dies ist die zweite Aussage.

Wir haben mit einer probabilistischen Methode auch gezeigt, dass eine ν -Stichprobe mit dieser Eigenschaft existiert. Dies lässt sich nochmals schematisch hinschreiben:

$$\begin{aligned} \mathbf{P}^\nu[\bar{x} \text{ ist } \varepsilon\text{-Netz}] &\geq \frac{1}{2} \\ \implies |\{\bar{x} \text{ ist } \varepsilon\text{-Netz}\}| &\geq \frac{1}{2}|A|^\nu > 0 \\ \implies \exists \bar{x} \in A^\nu : \bar{x} \text{ ist } \varepsilon\text{-Netz}. \end{aligned}$$

□

Für eine solche Stichprobe \bar{x} sei $N \subset A$ die Menge der Elemente von \bar{x} . Dann ist N ein ε -Netz von A und es gilt $|N| \leq \nu = c_d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$.

3.5.14 Bemerkung. Bei der Einschränkung $1/2 > \varepsilon$ lässt sich $1/2$ durch jeden positiven Wert kleiner 1 ersetzen. Mit dem Wert 1 wie in [HW87, Lemma 4.8] lässt sich die Aussage leider nicht aus dem Satz (bzw. [HW87, Theorem 3.8]) herleiten: Für ε beliebig nahe 1 betrachte man die Gleichung

$$\log \frac{1}{\varepsilon} \geq \frac{\log 8d}{\frac{c_d}{8d} - 1} =: b.$$

Dabei ist b klein, aber stets positiv. Wählt man $\varepsilon = \frac{1}{2}^{b+1} < 1$, so folgt $\log \frac{1}{\varepsilon} < b$ im Widerspruch zur Forderung.

3.5.15 Bemerkung. Wie man Satz 3.5.13 sofort ansieht, gibt es zu jedem $0 < \delta < 1$ eine entsprechende Konstante, aber für die später verwendete randomisierte Konstruktion der Datenstruktur ist $\delta = \frac{1}{2}$ hinreichend.

3.5.16 Bemerkung. Wir sind also am Ziel dieses Kapitels und können zu jedem Bereichsraum mit endlicher VC-Dimension die Existenz eines ε -Netz der Größenordnung $\mathcal{O}(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$ voraussetzen. Desweiteren gibt es stets ein hinreichend großes $\nu \in \mathcal{O}(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$, so dass eine zufällige ν -Stichprobe mindestens mit Wahrscheinlichkeit $\frac{1}{2}$ ein ε -Netz ist.

In den nächsten beiden Kapiteln befassen wir uns mit geometrischen Bereichsräumen um aus dem Theorem III eine Aussage über geeignete Stichproben N von endlichen Mengen $A \subset \mathbb{R}^d$ zu machen, mit derer Hilfe sich eine Datenstruktur für Halbraum-Bereichsanfragen gewinnen lässt.

Kapitel 4

Geometrische Bereichsräume

In diesem Kapitel beschäftigen wir uns mit typischen Bereichsräumen im \mathbb{R}^d . Zunächst werden Notationen für bekannte Begriffe wie Punkte und Hyperebenen festgelegt.

Notation. Um eine Verwechslung eines Punktes x mit einem n -Tupel $\bar{x} = (x_1, \dots, x_n)$ aus Punkten zu vermeiden, werden Koordinaten mit kleinen griechischen Buchstaben notiert. Für die Punkte $x \in \mathbb{R}^d$ bzw. $p \in \mathbb{R}^d$ schreiben wir also beispielsweise $x = (\chi_1, \dots, \chi_d)$ bzw. $p = (\pi_1, \dots, \pi_d)$. Enthält ein Punkt $x_i \in \mathbb{R}^d$ bereits einen Index, wird der Punkt ggf. notiert mit $x_i = (\chi_{i,1}, \dots, \chi_{i,d})$.

Für die Untersuchung der Lage eines Punktes in Bezug auf eine Hyperebene vereinbaren wir wie folgt:

Konvention. Im \mathbb{R}^d sei die d -te Achse als vertikale Achse ausgezeichnet.

Wir definieren nun wie angekündigt den Begriff Hyperebene:

4.0.1 Definition (Hyperebene). Eine Hyperebene h im \mathbb{R}^d ist ein $(d-1)$ -dimensionaler affiner Teilraum des \mathbb{R}^d . h wird als vertikal bezeichnet, wenn sie eine Parallele der vertikalen Achse enthält. Für jede nicht-vertikale Hyperebene h gibt es eindeutige reelle Zahlen η_1, \dots, η_d , so dass sich h als die folgende Punktmenge angeben lässt:

$$h = \left\{ x \in \mathbb{R}^d : \eta_1 \chi_1 + \eta_2 \chi_2 + \dots + \eta_{d-1} \chi_{d-1} + \eta_d \right\}.$$

Die Lage eines Punktes p in Bezug auf eine Hyperebene h wird wie folgt definiert:

4.0.2 Definition (Lage eines Punktes p in Bezug auf h).

$$\begin{aligned} p \text{ liegt über } h &: \iff \pi_d > \eta_1 \pi_1 + \eta_2 \pi_2 + \dots + \eta_{d-1} \pi_{d-1} + \eta_d \\ p \text{ liegt auf } h &: \iff \pi_d = \eta_1 \pi_1 + \eta_2 \pi_2 + \dots + \eta_{d-1} \pi_{d-1} + \eta_d \\ p \text{ liegt unter } h &: \iff \pi_d < \eta_1 \pi_1 + \eta_2 \pi_2 + \dots + \eta_{d-1} \pi_{d-1} + \eta_d. \end{aligned}$$

Die Lage einer Hyperebene h in Bezug auf einen Punkt p wird entsprechend definiert. Eine Hyperebene h unterteilt den \mathbb{R}^d in drei zusammenhängende Teile: $\mathbb{R}^d = h^* \dot{\cup} h \dot{\cup} h^\dagger$. Die beiden *offenen* Teile h^* und h^\dagger heißen *Halbräume*. Wir machen also die folgende

4.0.3 Definition (Halbraum). Ein d -dimensionaler Halbraum $h^* \subset \mathbb{R}^d$ ist die offene Punktmenge auf einer Seite der d -dimensionalen Hyperebene h . Ist h^* ein Halbraum, so sei h^\dagger stets der Halbraum auf der anderen Seite von h .

Wir setzen direkt fort mit der Definition eines positiven Halbraumes:

4.0.4 Definition (positiver Halbraum). Ein Halbraum h^* heißt *positiver* Halbraum, wenn die Hyperebene h vertikal oder wenn h^* die Punktmenge über h ist.

Einen positiven Halbraum bezeichnen wir mit h^+ , falls h nicht-vertikal ist.

Für vertikale Hyperebenen h sind nach dieser Definition beide Halbräume positiv. Zu einem positiven Halbraum h^+ wird der gegenüberliegende Halbraum mit h^- bezeichnet. Der Abschluss eines Halbraumes h^* wird als \bar{h}^* notiert und auch abgeschlossener Halbraum genannt. Die Menge aller (offenen) Halbräume sei mit H_d^* bezeichnet, die aller positiven (offenen) mit H_d^+ . Die Menge der nicht-vertikalen Hyperebenen sei $H_d^\#$. Wir kommen nun zur Definition eines Simplex im \mathbb{R}^d .

4.0.5 Definition (Simplex). Ein Simplex σ^* ist definiert als Schnitt von $(d+1)$ (nicht notwendig verschiedenen) Halbräumen h_i^* :

$$\sigma^* := \bigcup_{i=1}^{d+1} h_i^*.$$

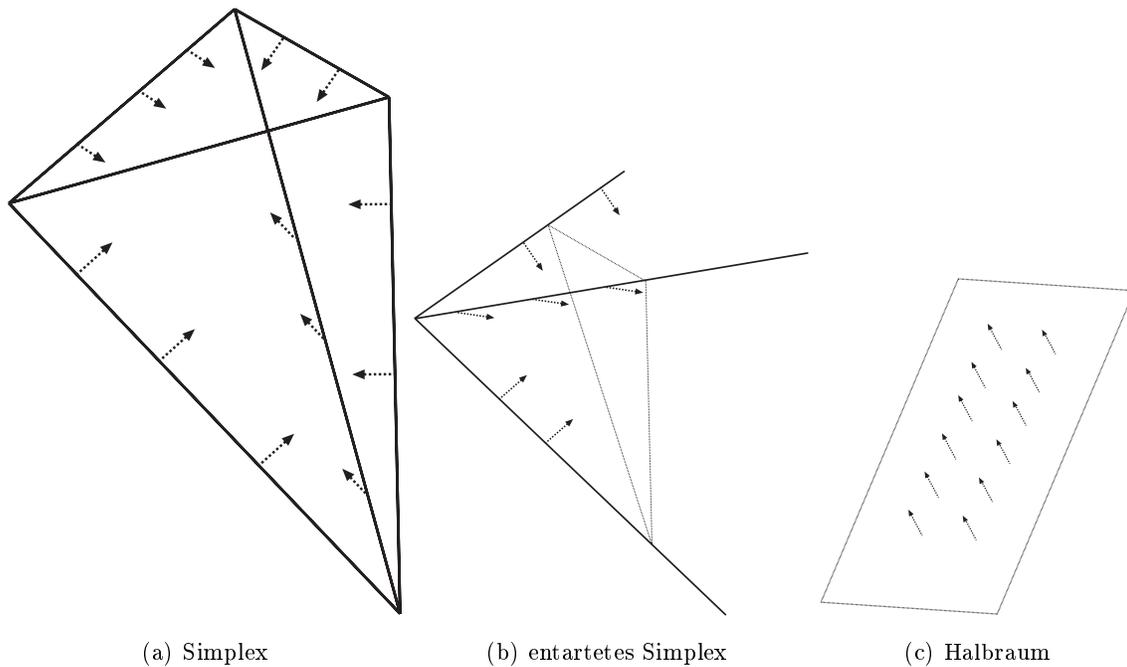
Der Rand σ eines Simplex σ^* ist definiert als der (topologische) Rand von σ^* und damit der Schnitt der zugehörigen Hyperebenen h_i .

Abbildung 4.1 zeigt drei Simplex im \mathbb{R}^3 . Wie man sieht, kann ein Simplex nach dieser Definition auch „entarten“. Insbesondere ist für einen Halbraum h^* auch der Schnitt der Halbräume $h_1^* = h_2^* = h_3^* = h^*$ ein Spezialfall eines Simplex σ^* .

Die Definition des Simplex werden wir später verwenden. Zunächst untersuchen wir jedoch Hyperebenen und Halbräume genauer.

Für die Konstruktion eines geeigneten Partitionsbaumes aus Stichproben N führen wir die folgende Menge von Hyperebenen ein:

4.0.6 Definition ($H_d(N)$). Sei N eine endliche Punktmenge aus $\nu \geq d$ Punkten. Dann ist $H_d(N)$ definiert als die Menge aller Hyperebenen (nicht Halbräume), die jeweils d dieser Punkte enthalten.



(a) Simplex

(b) entartetet Simplex

(c) Halbraum

Abbildung 4.1: Simplicies

Man sagt, dass N eine Punktmenge in allgemeiner Lage ist, wenn jede Hyperebene $h \in H_d(N)$ genau d der Punkte enthält und keine $(d+1)$ Hyperebenen einen gemeinsamen Punkt enthalten. Wir definieren B_d^+ , den Bereichsraum der positiven Halbräume im \mathbb{R}^d , d.h.

$$B_d^+ := (\mathbb{R}^d, H_d^+ \cup \mathbb{R}^d).$$

Dieser Bereichsraum ist ein wichtiger geometrischer Bereichsraum. Ein weiterer Bereichsraum im \mathbb{R}^d ist B_d^* , der Bereichsraum der offenen Halbräume im \mathbb{R}^d , d.h.

$$B_d^* := (\mathbb{R}^d, H_d^* \cup \mathbb{R}^d).$$

Relevant für Betrachtungen mit ε -Netzen sind jeweils deren endliche Unterräume. Die Hinzunahme des Bereiches \mathbb{R}^d vereinfacht später die formale Darstellung der sogenannten Korridore. Wie man leicht erkennt, verändert dies nicht die VC-Dimension, da für jede endliche Menge A stets ein (positiver) Halbraum existiert, der ganz A enthält. Wir erinnern an die Definition von ε -Netzen, nach der in endlichen Bereichsräumen (A, \mathcal{R}) für ein ε -Netz N gilt: Bereiche R , die keine Punkte aus N enthalten, enthalten höchstens den Bruchteil ε der Punkte aus A :

$$(R \cap N = \emptyset) \implies |R| = |R \cap A| \leq \varepsilon \cdot |A|.$$

Man betrachte zu $B_d^+ = (\mathbb{R}^d, \mathcal{R})$ mit $\mathcal{R} := H_d^+ \cup \mathbb{R}^d$ einen endlichen Unterraum $(A, \mathcal{R}|_A)$. Ein Bereich aus dem Unterraum hat dann die Form $h^+ \cap A = \{x \in A : x \text{ über } h\}$. Es

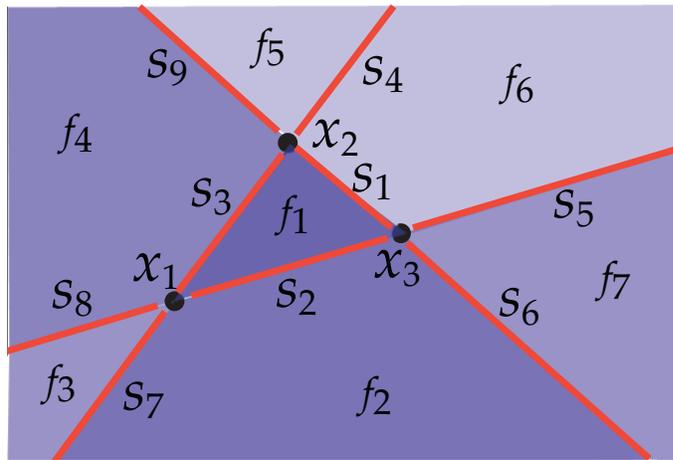


Abbildung 4.2: Arrangement aus 3 Hyperebenen

sind Untersuchungen zu diesen Bereichen anzustellen. Wichtig ist vor allem, die Anzahl dieser Bereiche zu bestimmen um zu zeigen, dass es eine größte Anzahl von Elementen gibt, für die eine Teilmenge A überdeckt werden kann. Da ein solcher Bereich durch Lagebeziehungen der Form x über/auf/unter h charakterisiert wird, bietet die kombinatorische Geometrie Grundlagen für solche Untersuchungen. Zwei dieser Grundlagen sind erstens sogenannte *Arrangements* sowie zweitens die *geometrische Dualität*, welche die Lagebeziehung bei einem solchen Bereich in einen dualen Raum überträgt. Es zeigt sich, dass die Anzahl der Bereiche durch die Anzahl der Zellen in einem Arrangement des dualen Raums ausgedrückt werden kann. In den folgenden zwei Abschnitten werden daher diese beiden relevanten Grundlagen der kombinatorischen Geometrie eingeführt. Als Quelle hierzu dient Edelsbrunner [Ede87], in dem sich weiterführende Aussagen über Arrangements finden. Auch die in den Kapiteln 6 und 7 zu Halbraum-Bereichsanfragen und Simplex-Bereichsanfragen verwendeten algorithmischen Aussagen finden sich in diesem Lehrbuch. Dies sind Aussagen über den Speicherbedarf und Laufzeit von Operationen in geeigneten Repräsentationen von Arrangements. Die dritte und letzte im Rahmen dieser Diplomarbeit behandelte Grundlage der kombinatorischen Geometrie betrifft den Begriff der Zone in einem Arrangement. Sie wird im nächsten Kapitel eingeführt.

4.1 Grundlage: Arrangements von Hyperebenen

4.1.1 Definition (Arrangement). Für eine endliche Menge H aus Hyperebenen im \mathbb{R}^d ist das Arrangement $\mathfrak{A}(H)$ eine Unterteilung (Partitionierung) des \mathbb{R}^d in zusammenhängende Teile (kurz Teile) verschiedener Dimension, welche durch die Hyperebenen $h \in H$ realisiert wird.

Zwei Punkte p und q gehören per Definition genau dann zum gleichen zusammenhängenden Teil, wenn sie in Bezug auf jede Hyperebene $h \in H$ dieselbe Lage haben. Damit ist gemeint, dass sie entweder beide auf h liegen oder beide in demselben Halbraum h^* liegen. Da hier nicht alle Teile von Bedeutung sind, beschränken wir uns auf die Betrachtung der zusammenhängenden Teile der Dimension 0, 1 und d .

(Für weitergehende Untersuchungen sei auf Edelsbrunner verwiesen [Ede87, Kapitel 1]. Dort wird ein zusammenhängendes Teil eines Arrangements formal angegeben durch den Positionsvektor eines Punktes in Bezug auf die Hyperebenen aus h .)

Die Teile der Dimension d heißen Zellen f . Mit $F(\mathfrak{A}(H))$ wird die Menge der Zellen aus $\mathfrak{A}(H)$ bezeichnet. Zellen f sind stets offen, wie man leicht einsieht: Ein Punkt eines zusammenhängenden Teils z der Dimension d kann in Bezug auf keine Hyperebene $h \in H$ die Lage „liegt auf“ beschreiben, da sonst die Dimension $\leq d-1$ wäre. Damit hat ein jeder Punkt p aus dieser Teile einen positiven Abstand zu jeder Hyperebene. Also ist z offen.

Die nulldimensionalen zusammenhängenden Teile sind Punkte. Dies sind die Schnittpunkte der Hyperebenen. Die eindimensionalen Teile heißen Kanten oder Segmente.

Die Anzahl $|F(\mathfrak{A}(H))|$ der Zellen sowie die Anzahl anderer Teile in einem Arrangement sind ein wichtiges Maß für dessen Komplexität, welches in der algorithmischen Geometrie in vielen Zusammenhängen auftritt.

Ein weiterer wichtiger Begriff für Arrangements ist die Inzidenz, welche beschreibt, wann zwei Teile aneinander grenzen. Die Punkte, die an eine Zelle grenzen, heißen Eckpunkte der Zelle.

Relevant für die später betrachteten Partitionsbäume sind Arrangements aus $H_d(N)$, der Menge der Hyperebenen durch je d der Punkte einer endlichen Menge N . (Bei diesen ist N Stichprobe einer größeren Menge A). Daher werden drei kurze Bezeichnungen verwendet, wenn im Kontext klar ist, dass N eine Punktmenge ist:

$\mathfrak{A}(N)$ bezeichnet das Arrangement $\mathfrak{A}(H_d(N))$.

$F(N)$ bezeichnet die Menge der Zellen in $\mathfrak{A}(N)$.

Die im nächsten Kapitel definierte Zone von h im Arrangement $\mathfrak{A}(N)$ wird mit $Z(h, N)$ bezeichnet.

4.1.2 Lemma (Anzahl der Zellen). *Für die Anzahl $m := |F(N)|$ der Zellen in $\mathfrak{A}(N)$ mit $|N| = \nu$ gilt*

$$m = |F(N)| \leq \Phi_d \left(\binom{\nu}{d} \right),$$

wobei Gleichheit bei allgemeiner Lage auftritt. Damit ist gemeint, dass jede Hyperebene aus $H_d(N)$ genau d Punkte aus N enthält und sich stets nur d Hyperebenen in einem Punkt aus \mathbb{R}^d schneiden. Wir erinnern an die Definition 3.3.5 (S.42): $\Phi_d(n) = \sum_{k=0}^d \binom{n}{k}$.

Es wird ein Beweis skizziert, der das Extremalprinzip und eine Bijektion zwischen Zellen und Punkten verwendet.

Beweis. Die Begriffe „nach unten beschränkt“ sowie minimal verwenden wir hier wie üblich in der Analysis für die reellen Zahlen \mathbb{R} angewandt auf die d -te Koordinate. Eine Zelle f heißt also „nach unten beschränkt“, wenn für $x \in \mathbb{R} : x \in f$ die d -te Koordinate χ_d einen gewissen Wert nicht unterschreitet. Es gibt einen Eckpunkt einer solchen Zelle mit minimaler d -Koordinate oder kurz einen minimalen Eckpunkt. Ohne Einschränkung verlangen wir, dass es jeweils genau einen solchen gibt. (Sonst lässt sich stets eine geeignet kleine Drehung des Koordinatensystems vornehmen, welche diese Situationen aufhebt.) Wir ordnen nun jeder nach unten beschränkten Zelle seinen minimalen Eckpunkt zu. Man beobachtet nun, dass dies eine Bijektion zwischen der Menge der Schnittpunkte im Arrangement und der Menge der nach unten beschränkten Zellen darstellt. Einerseits wird jeder Schnittpunkt zugeordnet, da mindestens eine nach unten beschränkte Zelle angrenzt, andererseits teilen sich aufgrund der allgemeinen Lage offensichtlich keine zwei Zellen den gleichen minimalen Schnittpunkt. Damit gibt es $\binom{\nu}{d}$ nach unten unbeschränkte Zellen. Man betrachte nun eine Hyperebene $g := \{(\chi_1, \dots, \chi_{d-1}, -s), s \in \mathbb{R}^+\}$, die unter allen Schnittpunkten liegt. (Die Menge aller Punkte mit gleicher d -Koordinate.) Ohne Einschränkung schneidet diese jede der vorgelegten Hyperebenen. (Falls sie das nicht tut, verwende man eine durch eine kleine Drehung erzeugte Hyperebene g' , welche unter allen Schnittpunkten liegt.) Man betrachte nun die Situation in der $(d-1)$ -dimensionalen Hyperebene g . Jede Hyperebene eingeschränkt ergibt eine $(d-2)$ -dimensionale Hyperebene in dieser. Also erhält man ein $(d-1)$ -dimensionales Arrangement aus n Hyperebenen. (An dieser Stelle wird auf eine formale Darstellung dieses Arrangements verzichtet.) Somit ergibt sich die Formel für allgemeine Lage per vollständiger Induktion nach d , da für den Induktionsanfang $d=1$ gilt: ν Punkte im \mathbb{R}^1 unterteilen die reelle Gerade in $\nu+1 = \binom{\nu}{0} + \binom{\nu}{1}$ Intervalle.

Es gilt also $|F(N)| = \Phi_d\left(\binom{\nu}{d}\right)$ für Punktmengen in allgemeiner Lage.

Für Punktmengen N in nicht-allgemeiner Lage lässt sich stets eine leicht veränderte Punktmenge N' finden mit $|F(N')| \geq |F(N)|$. Mit dieser (als *perturbation* bezeichneten) Technik zeigt man, dass $\Phi_d\left(\binom{\nu}{d}\right)$ stets eine Abschätzung nach oben darstellt. Da bei den später verwendeten Punktmengen allgemeine Lage vorausgesetzt wird, führen wir diese Technik hier nicht explizit durch und verweisen auf [Ede87, Theorem 1.3]. \square

4.1.3 Lemma (Anzahl der Schnittzellen). *Für eine beliebige Hyperebene h ist die Anzahl m der Zellen in $\mathfrak{A}(N)$, welche h schneiden, begrenzt durch*

$$m \leq \Phi_{d-1}\left(\binom{\nu}{d}\right).$$

Beweis. O.B.d.A. lässt sich voraussetzen, dass h jede Hyperebene aus $H_d(N)$ (echt) schneidet, da sonst die Anzahl der geschnittenen Zellen nicht größer wird. Man betrachte die Situation in der $(d-1)$ -dimensionalen Hyperebene h . Jede Hyperebene aus $H_d(N)$ eingeschränkt auf h ergibt eine $(d-2)$ -dimensionale Hyperebene in dieser. Damit erhält man

ein $(d-1)$ -dimensionales Arrangement aus $|H_d(N)|$ Hyperebenen. Also folgt die Aussage aus dem vorigen Lemma. \square

4.1.4 Lemma. Die Anzahl $m = |F(N)|$ der Zellen in $\mathfrak{A}(N)$ mit $|N| = \nu$ lässt sich für $d < \nu$ abschätzen durch

$$m \leq \nu^{d(d-1)}.$$

Diese technische Abschätzung wird hier explizit formuliert und bewiesen. Haussler und Welzl verwenden sie in der ersten Ungleichung des Beweises von Theorem 4.9 [HW87, S.144 unten].

Beweis. Nach Definition des Binomialkoeffizienten lässt sich abschätzen:

$$\binom{\nu}{d} = \frac{\nu \cdot (\nu-1) \cdot \dots \cdot (\nu-d+1)}{d!} \leq \frac{\nu \cdot \nu \cdot \dots \cdot \nu}{d!} < \frac{\nu^d}{d!}.$$

Setzt man dies in $\Phi_{d-1} \left(\binom{\nu}{d} \right)$ ein, erhält man nach Lemma 4.1.2

$$m < \Phi_{d-1} \left(\frac{\nu^d}{d!} \right) = \binom{\nu^d/d!}{0} + \binom{\nu^d/d!}{1} + \dots + \binom{\nu^d/d!}{d-1}.$$

Nun schätzt man jeden dieser Binomialkoeffizienten mit $\binom{n}{k} < n^k$ ab. (Diese Abschätzung ist wie oben, jedoch abgeschwächt, da $d! \geq 1$.)

$$\begin{aligned} m &< \left(\frac{\nu^d}{d!} \right)^0 + \left(\frac{\nu^d}{d!} \right)^1 + \dots + \left(\frac{\nu^d}{d!} \right)^{d-1} \leq 1 + \frac{\nu^{d \cdot 1}}{(d!)^1} + \dots + \frac{\nu^{d \cdot (d-1)}}{(d!)^{d-1}} \\ &\leq 1 + \frac{\nu^{d \cdot 1}}{d!} + \dots + \frac{\nu^{d \cdot (d-1)}}{d!} \leq 1 + (d-1) \frac{\nu^{d \cdot (d-1)}}{d!} \leq 1 + \nu^{d(d-1)}. \end{aligned}$$

Aus $m \in \mathbb{N}$ ergibt sich die Behauptung

$$m \leq \nu^{d(d-1)}. \quad \square$$

4.2 Grundlage: Geometrische Dualität

Die geometrische Dualität im \mathbb{R}^d ist eine Abbildung \mathcal{D} , welche die Menge der Punkte im \mathbb{R}^d bijektiv auf die Menge der (nicht-vertikalen) Hyperebenen $H_d^\#$ abbildet.

Die geometrische Dualität \mathcal{D} im \mathbb{R}^d ist unter Verwendung der Koordinatendarstellung einer nicht-vertikalen Hyperebene die wie folgt definierte bijektive Abbildung $\mathcal{D}: \mathbb{R}^d \rightarrow H_d^\#$, deren Umkehrabbildung $H_d^\# \rightarrow \mathbb{R}^d$ ebenfalls mit \mathcal{D} bezeichnet wird.

4.2.1 Definition (geometrische Dualität im \mathbb{R}^d).

$$\mathcal{D}(p) := h \text{ mit } h = \left\{ x \in \mathbb{R}^d : \chi_d = 2\eta_1\chi_1 + 2\eta_2\chi_2 + \dots + 2\eta_{d-1}\chi_{d-1} - \eta_d \right\} \text{ und}$$

$$\mathcal{D}(h) := p \text{ mit } \mathcal{D}(p) = \mathcal{D}(\mathcal{D}(h)) = h.$$

Die Dualität wird auf natürliche Weise erweitert auf Punktmenge und Mengen von Hyperebenen, so dass gilt:

$$\mathcal{D}(P) = \{\mathcal{D}(p) : p \in P\} \text{ und } \mathcal{D}(\mathcal{D}(P)) = \{\mathcal{D}(h) : h \in \mathcal{D}(P)\} = P.$$

Einsetzen lässt sich die geometrische Dualität für zahlreiche Aussagen, die mit der Lage von Punkten in Arrangements zusammenhängt. Grund dafür ist die Beobachtung, dass die \mathcal{D} die Lage „umkehrt“. Diese wird durch das folgende Lemma ausgedrückt:

4.2.2 Lemma (duale Lage). *Sei $p \in \mathbb{R}^d$ und $h \in H_d^\#$. Dann gilt:*

$$\begin{aligned} p \text{ liegt auf } h &\iff \mathcal{D}(h) \text{ liegt auf } \mathcal{D}(p) \\ p \text{ liegt über } h &\iff \mathcal{D}(h) \text{ liegt über } \mathcal{D}(p) \\ p \text{ liegt unter } h &\iff \mathcal{D}(h) \text{ liegt unter } \mathcal{D}(p). \end{aligned}$$

Beweis. Der Beweis benötigt lediglich einfache Algebra und wird daher nicht geführt. \square

Man beachte, dass bei Edelsbrunner die erste Eigenschaft als *incidence preservation* bezeichnet wird [Ede87, Abschnitt 1.4]. Für zahlreiche weitere Eigenschaften von \mathcal{D} bietet sich ein Korollar von Edelsbrunner an [Ede87, Korollar 1.6].

Mit Hilfe der eingeführten Begriffe und dem Konzept der Dualität lässt sich nun der Bereichsraum der Halbräume des \mathbb{R}^d eingehend untersuchen.

4.3 Der Bereichsraum der Halbräume des \mathbb{R}^d

4.3.1 Lemma (Dimension des Bereichsraumes der positiven Halbräume).

Der Bereichsraum $B_d^+ = (\mathbb{R}^d, H_d^+ \cup \mathbb{R}^d)$ der positiven Halbräume im \mathbb{R}^d hat die VC-Dimension d , d.h. es gilt:

$$\dim B_d^+ = d.$$

Beweis. [HW87, S.132 oben]

Wir zeigen zunächst, dass eine n -elementige Teilmenge $A = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ überdeckt werden kann, wenn $n < d$ gilt. Dazu betrachten wir das duale Bild $\mathcal{D}(A) = \{\mathcal{D}(x_1), \dots, \mathcal{D}(x_n)\}$. Dies sind n Hyperebenen, die den \mathbb{R}^d nach Lemma 4.1.2 in ein Arrangement mit $m \leq \Phi_d(n)$ Zellen aufteilen. Jede Zelle ist die Menge aller Punkt mit derselben Lage in Bezug auf sämtliche Hyperebenen $\mathcal{D}(x_1), \dots, \mathcal{D}(x_n)$: Sei h^+ ein positiver Halbraum im \mathbb{R}^d und h die zugehörige Hyperebene. Setze $A' := h^+ \cap A$. Dann gilt: Ein Punkt x_i liegt über der Hyperebene h genau dann, wenn $x_i \in A'$. Im Dualen gilt äquivalent: Der Punkt $\mathcal{D}(h)$ liegt über der Hyperebene $\mathcal{D}(x_i)$ genau dann, wenn $x_i \in A'$ gilt.

1. Fall: $n > d$.

Aus der Definition 3.3.5 (S.42) von $\Phi_d(n)$ ergibt sich $m \leq \Phi_d(n) < 2^n$. Da es weniger als 2^n Zellen gibt, gibt es weniger als 2^n verschiedene Teilmengen A' , die von einem positiven Halbraum als Schnittmenge von A gebildet werden. Also kann A nicht überdeckt werden, und es folgt:

$$\dim B_d^+ \leq d.$$

2. Fall: $n = d$.

Wir wählen die Punktmenge A so, dass allgemeine Lage nach Lemma 4.1.2 (S.65) vorliegt. Dann gilt für die Anzahl m der Zellen im Arrangement $m = \Phi_d(n) = 2^n$ nach der Definition von ϕ_d (3.3.5 unten). Wählen wir 2^n Punkte aus verschiedenen Zellen, so sind deren duale Bilder Hyperebenen h_i , $1 \leq i \leq 2^n$ mit verschiedenen Schnittmengen $A' = A \cap h_i$. Da es genau 2^n Teilmengen von A gibt, kommt jede Menge auch vor und es gilt: A wird überdeckt.

Es folgt $\dim B_d^+ \geq d$. Zusammen folgt also die Behauptung:

$$\dim B_d^+ = d. \quad \square$$

Haussler und Welzl geben an, dass sich durch eine leichte Erweiterung des obigen Arguments $\dim B_d^* = d + 1$ zeigen lässt. Hier wird nur die obere Schranke $\dim B_d^* \leq d + 1$ nachgewiesen, da diese für den Einsatzzweck völlig ausreicht. Natürlich ist $\dim B_d^* \geq d$, da $H_d^* \supset H_d^+$. Außerdem ist es nicht schwer, eine $d + 1$ -elementige Teilmenge im \mathbb{R}^d zu finden, die von B_d^* überdeckt wird, woraus sich dann $\dim B_d^* = d + 1$ folgern lässt.

4.3.2 Lemma (Dimension des Bereichsraumes der Halbräume). *Der Bereichsraum $B_d^* = (\mathbb{R}^d, H_d^+ \cup \mathbb{R}^d)$ der positiven Halbräume im \mathbb{R}^d hat höchstens die VC-Dimension $d + 1$, d.h. es gilt:*

$$\dim B_d^* \leq d + 1.$$

Beweis. Da zu einer Hyperebene h beide Halbräume zugelassen sind, kann es nun zu einer endlichen Menge A bis zu doppelt so viele verschiedene Bereiche $A \cap h^*$ geben. Dies sieht man leicht ein, indem man erneut das duale Bild $\mathcal{D}(A) = \{\mathcal{D}(x_1), \dots, \mathcal{D}(x_n)\}$ betrachtet. Dies sind n Hyperebenen, die den \mathbb{R}^d nach Lemma 4.1.2 (S.65) in ein Arrangement mit $m \leq \Phi_d(n)$ Zellen aufteilen. Jede Zelle ist die Menge der Punkte mit derselben Lage in Bezug auf die Hyperebenen $\mathcal{D}(x_1), \dots, \mathcal{D}(x_n)$. Es sei daran erinnert, dass ein Punkt x_i genau dann über der Hyperebene h liegt, wenn $\mathcal{D}(h)$ über $\mathcal{D}(x_i)$ liegt. Sei p ein Punkt einer Zelle z im zuvor definierten Arrangement. Dann ist wie zuvor $\mathcal{D}(p)$ eine Hyperebene h mit einem Bereich $A \cap h^+ \subset A$. Dieser Bereich ist wieder unabhängig von der Wahl des Punktes innerhalb der Zelle z . Es ist jedoch auch der Bereich $A \cap h^-$ ein zulässiger Bereich. Jede Zelle induziert bis zu zwei verschiedene Bereiche $A \cap h^+$ und $A \cap h^-$. Damit ist die Anzahl der verschiedenen Bereiche für eine endliche Menge A mit $|A| = n$ beschränkt durch $2 \cdot \Phi_d(n)$.

Da die Abschätzung $2\Phi_d(n) < 2^n$ für $n > d + 1$ leider keine allgemeingültige Aussage darstellt, wird der Beweis hier nicht durch eine Erweiterung des vorigen Arguments geführt. Stattdessen wird auf eine andere Weise gezeigt, dass eine n -elementige Menge mit $n = d + 2$ nicht überdeckt werden kann. Der Beweis wird nur skizziert, da einige technische Details unhandlich sind. Sei $A = \{x_1, x_2, \dots, x_{d+1}, x_{d+2}\}$ und werde zunächst angenommen, dass $A - \{x_{d+2}\}$ überdeckt wird. Das heißt, für beliebige $A' \subset A - \{x_{d+2}\}$ gibt es eine Hyperebene h^* mit $h^* \cap (A - \{x_{d+2}\}) = A'$. Nun wird der duale Raum untersucht. Insbesondere betrachtet man das Arrangement aus den $d + 1$ Hyperebenen $H := \{\mathcal{D}(x_1), \dots, \mathcal{D}(x_{d+1})\}$. Aus dem Lemma 4.1.3 (S.66) über die Anzahl der Schnitzellen in diesem Arrangement folgt, dass die $(d + 2)$ -te Hyperebene nicht alle Zellen schneidet. Man betrachte nun einen Punkt in einer solchen Zelle z . Die Menge der Hyperebenen $\mathcal{D}(x_i), 1 \leq i \leq d + 1$ über z bezeichne man mit H^+ und die unter z mit H^- . Dann gilt: Sei nun o.B.d.A. z über $\mathcal{D}(x_{d+2})$. Dann gibt es also nach der Definition von Zellen im Arrangement keinen Punkt $y \in \mathbb{R}^d$, für den gilt: y liegt über den Hyperebenen aus H^+ und unter der Hyperebenen $H^- \cup \{\mathcal{D}(x_{d+2})\}$. Die dualen Mengen sind Teilmengen von A , welche wir nun mit $A^{(1)}$ und $A^{(2)}$ bezeichnen. Dann folgt durch eine recht leichte, aber technisch unhandliche darzustellende Überlegung im primalen Raum:

$A^{(1)}$ und $A^{(2)}$ sind keine Bereiche, d.h. es gibt keinen Halbraum h^* mit der zugehörigen Hyperebene $h := \mathcal{D}(y)$, so dass $h^* \cap A = A^{(1)}$ bzw. $h^* \cap A = A^{(2)}$.

Also kann A nicht überdeckt werden und es folgt:

$$\dim B_d^* \leq d + 1. \quad \square$$

4.4 Kombination von Bereichsräumen

Eine weitere Eigenschaft von allgemeinen Bereichsräumen ist, dass sich aus k (identischen) Bereichsräumen der Dimension d ein neuer Bereichsraum mit „kombinierten“ Bereichen bilden lässt. Die Dimension des Bereichsraumes ist dann stets endlich und lässt sich durch einen Term mit k und d beschränken. Diese Kombination wird nun definiert.

4.4.1 Definition (k -Kombination). Sei (X, \mathcal{R}) ein Bereichsraum. Man betrachte einen Bereichsraum $(X, \tilde{\mathcal{R}})$, wobei jeder Bereich $\tilde{R} \in \tilde{\mathcal{R}}$ definiert sei durch eine Kombination von $k \geq 2$ Bereichen aus (X, \mathcal{R}) . Das ist zu verstehen als $R' = \phi(R_1, R_2, \dots, R_k)$ mit einer festen mengentheoretischen Formel ϕ (die Vereinigungen, Mengendifferenzen u.Ä. enthalten kann) und $R_1, \dots, R_k \in \mathcal{R}$. Dann heißt $(X, \tilde{\mathcal{R}})$ eine k -Kombination des Bereichsraumes (X, \mathcal{R}) .

4.4.2 Beispiel. Betrachte den Bereichsraum $(\mathbb{R}^2, \mathcal{R})$ mit $\mathcal{R} = \{h^+, h \text{ horizontale Gerade}\}$. Wähle die Formel $\phi : \mathcal{R} \times \mathcal{R} \rightarrow \mathcal{R}$ mit $\phi(R_1, R_2) = R_1 - R_2$. Deren Bildmenge ist die Menge der horizontalen Streifen der Ebene.

Sei nun also $\tilde{\mathcal{R}} = \{\phi(R_1, R_2), R_{1,2} \in \mathcal{R}\}$. Dann ist $(\mathbb{R}^2, \tilde{\mathcal{R}})$ ein Bereichsraum, der sich als 2-Kombination von $(\mathbb{R}^2, \mathcal{R})$ ergibt.

4.4.3 Lemma (Dimension von k -Kombinationen). *Sei (X, \mathcal{R}) ein Bereichsraum mit VC-Dimension d und $(X, \tilde{\mathcal{R}})$ eine k -Kombination von diesem. Dann hat der Bereichsraum $(X, \tilde{\mathcal{R}})$ eine Dimension $< 2dk \cdot \log(dk)$.*

4.4.4 Bemerkung. Diese allgemeine Formulierung als k -Kombination wird in [Mat93, Lemma 1.6] angegeben. Die von Haussler und Welzl benötigte Form ist nur der von ihnen formulierte Spezialfall von Korridoren, der im Folgenden behandelt wird. Da der Beweis nicht schwieriger ist, wird hier diese allgemeinere Form vorgezogen.

Beweis. [HW87, vgl. Lemma 4.5] Sei $A \subset X$ eine endliche Teilmenge mit $|A| =: m \geq 2$. Dann ist nach Lemma 3.3.7 (S.43) und Lemma 3.1.3 (S.34): $|A \cap \mathcal{R}| \leq \Phi_d(m)$. Man betrachte den von A induzierten Teilraum von $(X, \tilde{\mathcal{R}})$: Jeder Bereich $R \cap A$ mit $R \in \tilde{\mathcal{R}}$ hat die Form $R \cap A = \phi(R_1, R_2, \dots, R_k)$ mit $R_i \in \mathcal{R}$. Damit ist die Anzahl der Bereiche beschränkt durch $\Phi_d(m)^k < m^{dk}$. Es ist zu zeigen, dass für eine m -elementige Menge A mit $m = 2dk \log(dk)$ nie überdeckt werden kann, d.h. stets $|A \cap \tilde{\mathcal{R}}| < 2^m$ gilt. Mit $k \geq 2$ und $m \geq 2$ ergibt sich durch Standardumformungen $\Phi_d(m)^k < m^{dk} \leq 2^m$ und damit die Behauptung. \square

4.4.5 Beispiel. Der obige Bereichsraum $(\mathbb{R}^2, \mathcal{R})$ hat die VC-Dimension $d = 2$, denn seine Bereiche sind auch Bereiche in $B_2^+ = (\mathbb{R}^2, H_2^+ \cup \mathbb{R}^2)$. Es folgt, dass auch $(X, \tilde{\mathcal{R}})$ ein Bereichsraum von endlicher VC-Dimension \tilde{d} mit $\tilde{d} < 8 \cdot \log 4 = 16$ ist.

4.5 Der Bereichsraum der k -Korridore im \mathbb{R}^d

Ein k -Korridor ist eine offene Teilmenge des \mathbb{R}^d , der wie folgt durch k Halbräume h_1^*, \dots, h_k^* definiert ist:

$$K_{i=1}^k(h_i^*) := \bigcup_{i=1}^k (h_i^*) - \bigcap_{i=1}^k (\overline{h_i^*}).$$

Für die spätere Anwendung ist es entscheidend, dass Korridore offen sind, daher wurde der Abschluss $\overline{h_i^*}$ verwendet.

Der Bereichsraum $(\mathbb{R}^d, \mathcal{K}_d^k)$ mit $\mathcal{K}_d^k := \{K : K \text{ ist } k\text{-Korridor im } \mathbb{R}^d\}$ der k -Korridore ist eine k -Kombination des Bereichsraumes der Halbräume. Durch die Verwendung des Abschlusses wird ein klein wenig Technik notwendig, um den Bereichsraum formal als Kombination mit einer Formel ϕ anzugeben. Zu einem offenen Halbraum h^* sei h^\dagger stets der gegenüberliegende Halbraum, d.h. $h^* \cup h \cup h^\dagger = \mathbb{R}^d$. ϕ lässt sich dann wie folgt festlegen:

$$\phi(h_1^*, \dots, h_k^*) := \bigcup_{i=1}^k (h_i^*) - \bigcap_{i=1}^k (\mathbb{R}^d - h_i^\dagger).$$

Wegen $\mathbb{R}^d - h_i^\dagger = \overline{h_i^*}$ ist damit $\phi(h_1^*, \dots, h_k^*) = K_{i=1}^k(h_i^*)$. Da ϕ nur aus Differenz, Vereinigung und Schnitt gebildet ist, gilt also das folgende Lemma:

4.5.1 Lemma.

Der Bereichsraum $(\mathbb{R}^d, \mathcal{K}_d^k)$ ist eine k -Kombination des Bereichsraumes B_d^* .

Damit lässt sich die Dimension der k -Korridore nach oben abschätzen:

4.5.2 Satz (Dimension von k -Korridoren). Sei k fest. Dann hat der Bereichsraum $(\mathbb{R}^d, \mathcal{K}_d^k)$ der k -Korridore eine endliche Dimension und es gilt:

$$\dim(\mathbb{R}^d, \mathcal{K}_d^k) \leq (d+1)k \cdot \log(d+1)k.$$

Beweis. [HW87, Korollar 4.6] Nach dem vorigen Lemma 4.5.1 ist der Bereichsraum der k -Korridore eine k -Kombination mit der angegebenen Formel. Da der Bereichsraum B_d^* der Halbräume des \mathbb{R}^d nach Lemma 4.3.2 (S.69) höchstens die Dimension $(d+1)$ hat, folgt die Aussage also aus dem Lemma 4.4.3 (S.71) über k -Kombinationen. \square

4.6 Zum ε -Theorem für k -Korridore

In diesem Abschnitt wird zum Bereichsraum der k -Korridore der von einer endlichen Menge A induzierte Unterraum betrachtet. Das Theorem besagt, dass sich die Anzahl der Punkte in dessen Bereichen mit Hilfe von ε -Netzen beschränken lässt.

Theorem IV (Das ε -Theorem für k -Korridore) . Sei k fest. Dann gibt es eine Konstante c_d , so dass für alle $1/2 > \varepsilon > 0$ und alle n -elementigen Mengen $A \subset \mathbb{R}^d$ gilt:

1. Es gibt eine ν -Stichprobe N (als Menge) mit $\nu = c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$, so dass für alle k -Korridore K des \mathbb{R}^d gilt:

$$|K \cap N| = 0 \implies |K \cap A| \leq \varepsilon n.$$

2. Eine zufällige ν -Stichprobe N erfüllt diese Eigenschaft mindestens mit Wahrscheinlichkeit $1/2$.

Beweis. Sei $\mathcal{R} := \mathcal{K}_d^k$ die Menge der k -Korridore. Nach dem Satz 4.5.2 über die Dimension gilt für die Dimension \tilde{d} von $(\mathbb{R}^d, \mathcal{R})$:

$$\tilde{d} \leq (d+1)k \log((d+1)k).$$

Nach dem Lemma 3.3.4 (S.41) über die Vererbung der VC-Dimension hat dann auch der durch die endliche Menge A induzierte Unterraum $(A, \mathcal{R}|_A)$ eine endliche Dimension kleiner oder gleich \tilde{d} . Diese Dimension \tilde{d} ist konstant, d.h. nicht von $|A|$ abhängig. Nach Theorem III (S.59) gilt dann mit der Konstanten $c_d^* := c_{\tilde{d}}$:

1. Es gibt ein ε -Netz N (als Menge) mit $\nu = c_d^* \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$ für den Bereichsraum $(A, \mathcal{R}_{|A})$.
Für dessen Bereiche $R \subset \mathcal{R}_{|A}$ gilt:

$$|R \cap N| = 0 \implies |R \cap A| \leq \varepsilon n.$$

2. Eine zufällige ν -Stichprobe N ist mindestens mit Wahrscheinlichkeit $1/2$ ein ε -Netz für $(A, \mathcal{R}_{|A})$.

Sei N nun ein ε -Netz für $(A, \mathcal{R}_{|A})$. Nach der Definition von ε -Netz gilt dann für alle k -Korridore K :

$$|(K \cap A) \cap N| = |(K \cap N)| = 0 \implies |K \cap A| \leq \varepsilon n. \quad \square$$

Interessant für die algorithmische Anwendung sind offensichtlich nur k -Korridore, bei denen der Parameter k nicht in Abhängigkeit von $|A|$ beliebig groß werden kann, sondern durch eine Konstante beschränkt ist. Im folgenden Kapitel werden wir zeigen, dass speziell der Einsatz von $(d+1)$ -Korridoren im \mathbb{R}^d die Grundlage für die Durchführung eines effizienten Divide & Conquer-Verfahrens bietet.

Kapitel 5

Verwendung von Korridoren

In diesem Kapitel werden die Bereichsräume der $(d + 1)$ -Korridore genauer untersucht. Durch geometrische Überlegungen gelingt es folgendes zu zeigen: Die Menge der Regionen auf einer Ebene des Partitionsbaumes, für die rekursive Aufrufe notwendig werden, ist enthalten in einem geeigneten $(d + 1)$ -Korridor. Diese Aussage wird im Einbettungssatz 5.3.1 formuliert. Nach diesem Satz gibt es einen $(d + 1)$ -Korridor, der die Menge der geschnittenen Zellen enthält. Hierdurch ist die Gesamtanzahl der Punkte, die in rekursiv aufgerufenen Regionen liegen, sehr klein, falls ein ε -Netz für den Bereichsraum der $(d + 1)$ -Korridore verwendet wird. Diese Eigenschaft wird beschrieben durch eine ε -Bedingung für Hyperebenen, welche eine Aussage über die sogenannte Zone einer beliebigen Hyperebene macht.

5.1 Geometrische Grundlage: Zonen in Arrangements

Haussler und Welzl verwenden das Paradigma des *Random Sampling* (zufällige Stichproben). Für diese Stichproben N wird stets das Arrangement $\mathfrak{A}(N) \stackrel{\text{def}}{=} \mathfrak{A}(H_d(N))$ bestimmt. Rekursive Aufrufe werden für Zellen notwendig, welche eine Anfrage-Hyperebene h schneiden. Um die Aussage der ε -Bedingung für Hyperebenen zu quantifizieren, wird zunächst die Menge der geschnittenen Zellen wie in der algorithmischen Geometrie üblich als Zone von h in N definiert.

5.1.1 Definition (Zone). Seien eine Hyperebene h und ein Arrangement $\mathfrak{A}(H)$ aus einer endlichen Menge H von Hyperebenen gegeben. Dann ist die Zone $Z(h, \mathfrak{A}(H))$ die Menge der Zellen dieses Arrangements, welche die Hyperebene h schneiden:

$$Z(h, \mathfrak{A}(H)) := \{f \in F(\mathfrak{A}(H)) : f \cap h \neq \emptyset\}.$$

Es sei daran erinnert, dass $F(\mathfrak{A}(H))$ die Menge der Zellen aus $\mathfrak{A}(H)$ ist und dass Zellen f stets offen sind. Ist N eine endliche Punktmenge mit dem Arrangement $\mathfrak{A}(N)$, so schreiben wir kurz

$$Z(h, N) := \{f \in F(N) : f \cap h \neq \emptyset\}.$$

Es ist also zu unterscheiden zwischen der Notation $\mathfrak{A}(N) := \mathfrak{A}(H_d(N))$ für Punkt Mengen N und $\mathfrak{A}(H)$ für Mengen H von Geraden. Falls das betrachtete Arrangement im Kontext klar ist, kann auch einfach $Z(h)$ geschrieben werden.

5.2 Formulierung des Ziels

Ziel ist es zu jeder endlichen Punktmenge $A \subset \mathbb{R}^d$ eine kleine Teilmenge N (Stichprobe) zu finden, so dass für eine beliebige Hyperebene h nur wenige Punkte in der Menge der geschnittenen Zellen, d.h. der Zone $Z(h, N)$, liegen. Wir machen daher die folgende Definition:

5.2.1 Definition (ε -Bedingung). Sei eine endliche Menge A gegeben. Wir sagen dann, dass eine Teilmenge (Stichprobe) N eine ε -Bedingung für Halbräume mit $0 < \varepsilon < 1/2$ erfüllt, wenn gilt:

$$\sum_{f \in Z(h, N)} |f \cap A| \leq \varepsilon \cdot |A|.$$

5.3 Einbettung der Zone in einen geeigneten Korridor

5.3.1 Satz (Einbettungssatz). *Sei $N \subset \mathbb{R}^d$ eine endliche Menge. Zur Einfachheit setzen wir voraus, dass $|N| \geq d + 1$ gilt. Dann gibt es einen $(d + 1)$ -Korridor $C := K_{i=1}^{d+1} g_i^*$ mit $g_i^* \in H_d^* \cup \mathbb{R}^d$, für den gilt:*

1. $C \cap N = \emptyset$.
2. $C \supset Z(h, N) \stackrel{\text{def}}{=} Z(h, \mathfrak{A}(H_d(N)))$.

Um den Satz zu beweisen, müssen einige Untersuchungen im dualen Raum angestellt werden. Das duale Bild einer Strecke (x_1, x_2) (als Punktmenge betrachtet) spielt dabei eine besondere Rolle. Dieses ist eine Hyperebenenchar, deren Punktmenge ein Doppelkegel ist. Man erkennt in Abbildung 5.1 leicht, dass die beiden seitlichen Zellen enthalten sind, nicht aber die nach oben bzw. unten unbeschränkte. Der Doppelkegel lässt sich am instruktivsten in dem (sehr kleinen) Arrangement aus den Hyperebenen $g_1 := \mathcal{D}(x_1)$ und $g_2 := \mathcal{D}(x_2)$ ausdrücken. Dieses Arrangement hat im $\mathbb{R}^d, d \geq 2$ genau $4 = 2^2$ Zellen, von denen eine nach oben sowie eine nach unten beschränkt ist. Es hat genau einen $(d - 2)$ -dimensionalen Zusammenhangsteil z . (Im \mathbb{R}^2 ist dies der Schnittpunkt der Geraden g_1 und g_2 .) Dieser

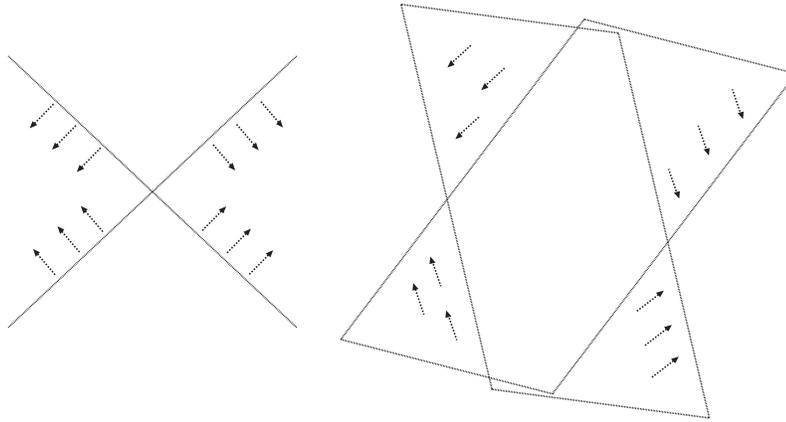


Abbildung 5.1: Doppelkegel

Zusammenhangsteil z ist das Duale der Geraden h , welche die Strecke (x_1, x_2) enthält. Das duale Bild der Strecke (x_1, x_2) ist die Schar der Hyperebenen g , welche z enthalten und in der Vereinigung von z und den beiden unbeschränkten Zellen enthalten ist. Da (x_1, x_2) eine offene Strecke ist, reicht es offenbar zu zeigen, dass eine solche Hyperebene g nicht in der Vereinigung der beschränkten Zellen enthalten ist. Dies erkennt man anhand der Abbildung leicht wie folgt: Die Vereinigung der nach oben bzw. unten beschränkten Zellen enthält vertikale Hyperebenen. Hyperebenen mit „großer Steigung“ sind solche, die sich vertikalen Geraden „annähern“. Deren duale Entsprechungen sind aber beliebig weit entfernte Punkte. Diese sind nicht Teilmenge eines Segments mit endlicher Länge. Ein Doppelkegel hat damit die Gestalt $(g_1^* \cup g_2^*) - (\overline{g_1^*} \cap \overline{g_2^*})$. Er ist also nicht nur die Vereinigung zweier Zellen, sondern zugleich ein 2-Korridor.

Das folgende Lemma beinhaltet einen Spezialfall des Einbettungssatzes, welcher sich auf direktem Weg beweisen lässt.

5.3.2 Lemma (Spezialfall des Einbettungssatzes). *Der Einbettungssatz gilt für Hyperebenen h mit $h^* \cap N = \emptyset$, d.h. wenn nur auf einer Seite von h Punkte aus N liegen.*

Beweis. In diesem Sonderfall ist der Beweis deutlich einfacher: Man betrachte eine beliebige $(d+1)$ -elementige Teilmenge $N' = \{x_1, \dots, x_{d+1}\}$, die nicht komplett auf der Hyperebene h liegt. Diese beschreibt den Rand eines unter Umständen entarteten Simplex. Ein Simplex σ^* ist nach Definition 4.0.5 (S.62) gegeben als ein Schnitt von $(d+1)$ Hyperebenen:

$$\sigma^* = \bigcap_{i=1}^{d+1} h_i^*.$$

Dabei sind die $(d+1)$ zugehörigen Hyperebenen h_i festgelegt durch die d -elementigen Teilmengen von N' . Die Anzahl der möglichen Auswahlen ist gegeben durch $\binom{d+1}{d} = d+1$.

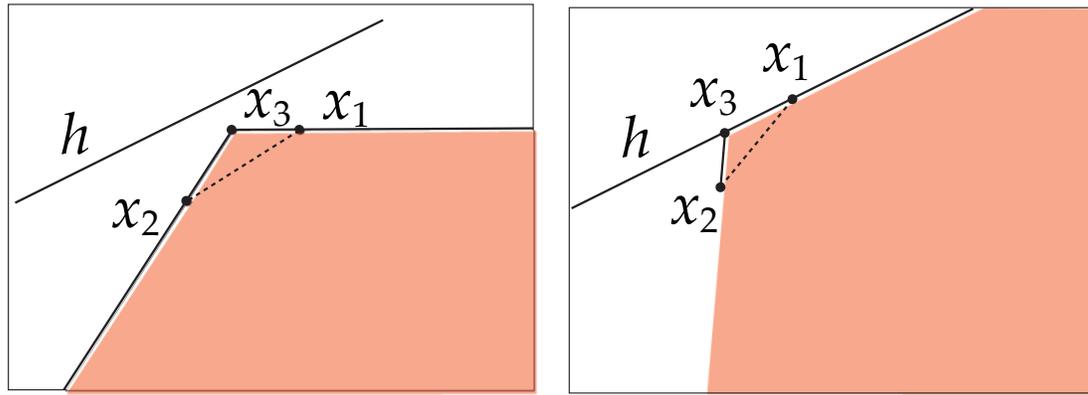


Abbildung 5.2: Spezialfall: Affiner Kegel

Sei nun x_j ein Eckpunkt von σ^* mit minimalem Abstand $\delta \geq 0$ zu h . Es lässt sich o. B. d. A. die Bezeichnung $j = d + 1$ wählen. Nun liegen genau d Hyperebenen an x_{d+1} an und genau eine Hyperebene h_{d+1} liegt x_{d+1} gegenüber. Mit dieser Bezeichnung ist der Schnitt der d abgeschlossenen Halbräume

$$C := \bigcap_{i=1}^d \overline{h_i^*}$$

ein affiner offener Kegel im \mathbb{R}^d mit der Spitze x_{d+1} (siehe Abbildung 5.2). Wie sich mit linearer Algebra zeigen lässt, kann man diesen Kegel beschreiben als die Menge der positiven Linearkombinationen von Richtungsvektoren $\overrightarrow{x_{d+1}x_i}$. Aufgrund der Wahl von x_{d+1} haben alle Punkte des Kegels einen Abstand von mehr als $\delta \geq 0$ von h . Dies trifft auch dann zu, wenn mehrere Punkte auf h liegen. Also gilt:

$$\begin{aligned} \bigcap_{i=1}^d \overline{h_i^*} \cap h &= \emptyset \\ \implies h &\subset \mathbb{R} - \bigcap_{i=1}^d \overline{h_i^*}. \end{aligned}$$

An dieser Stelle wird erkennbar, warum \mathbb{R}^d als Bereich von B_{d+1}^* zugelassen wurde: Mit $g_i^* := h_i^*$ für $i = 1, \dots, d$ und $g_{d+1}^* := \mathbb{R}^d$ gilt $\bigcap_{i=1}^{d+1} \overline{g_i^*} = C \cap \mathbb{R}^d = C$ und $\bigcup_{i=1}^{d+1} g_i^* = \mathbb{R}^d$. Damit ist C ein $(d + 1)$ -Korridor, der sich durch die $d + 1$ Halbräume g_1^*, \dots, g_{d+1}^* ergibt:

$$K_i^{d+1} g_i^* = \bigcup_{i=1}^{d+1} g_i^* - \bigcap_{i=1}^{d+1} \overline{g_i^*} = C. \quad \square$$

Da der Einbettungssatz in diesem Spezialfall nachgewiesen wurde, sei in den folgenden Lemmata als Voraussetzung gegeben, dass in beiden Halbräumen von h Punkte aus N liegen. Wir definieren diese Teilmengen wie folgt:

$$N^+ := N \cap h^+ \neq \emptyset \neq N \cap h^- \quad (\text{und } N^0 := N \cap h).$$

Nützlich sein wird die Definition der Umgebung von h . Dies ist die Menge $G^{(h)} \subset H_d(N)$ aus Hyperebenen, die „zwischen“ N^+ und N^- liegen im folgenden Sinne:

5.3.3 Definition ($G^{(h)}$, die Umgebung von h).

Als Umgebung von h bezeichnen wir $G^{(h)}$ definiert durch:

$$G^{(h)} := \{g \in H_d(N) : g^+ \cap N = N^+ \wedge g^- \cap N = N^-\}.$$

5.3.4 Lemma (über „Umgebungskorridore“). *Seien g_1, \dots, g_k Hyperebenen aus der Umgebung $G^{(h)}$. Dann enthält der k -Korridor $K_{i=1}^k g_i^+$ keine Punkte aus N , d.h.*

$$\left(K_{i=1}^k(g_i^+)\right) \cap N = \emptyset.$$

Beweis. Es ist $K_{i=1}^k(g_i^+) = \bigcup_{i=1}^k g_i^+ - \bigcap_{i=1}^k \overline{g_i^+}$.

Wir zeigen zunächst $K_{i=1}^k g_i^+ \cap (N^- \cup N^0) = \emptyset$: Betrachte ein festes g_j .

Dann ist $g_j^+ \cap (N^- \cup N^0) = \emptyset$. Also ist auch $\bigcup_{i=1}^k (g_i^+ \cap (N^- \cup N^0)) = \emptyset$. Die Differenzbildung fügt keine Punkte hinzu, es folgt die Teilaussage. Wir zeigen nun $K_{i=1}^k(g_i^+) \cap (N^+) = \emptyset$: Man betrachte erneut ein g_j . Es ist $g_j \supset N^+$. Also ist auch $\bigcap_{i=1}^k g_i^+ \subset N^+$ und damit

$$\left(\bigcup_{i=1}^k g_i^+ - \bigcap_{i=1}^k \overline{g_i^+}\right) \cap N^+ = \emptyset$$

Mit den beiden Teilaussagen folgt die Behauptung. □

Unmittelbar Aus der Definition der Umgebung von h folgt sofort, dass die Umgebung im Dualen genau die Eckpunkte der Zelle darstellt, welche h enthält:

5.3.5 Lemma (über eine Zelle des dualen Arrangements). *Man betrachte die Menge $\mathcal{D}(N)$ der zu Punkten aus N dualen Hyperebenen. Dann gilt:*

1. *Der Punkt $\mathcal{D}(h)$ ist in einer beschränkten Zelle f enthalten.*
2. *Die Menge der Eckpunkte von f ist $\mathcal{D}(G^{(h)})$.*

Beweis. Für den ersten Teil ist nur zu beachten, dass N^+ und N^- nicht-leer sind und damit $\mathcal{D}(h)$ nach oben und nach unten beschränkt. Aus der Definition der Zelle erhält man nun leicht die Aussage.

Auch die zweite Aussage folgt aus der Definition der Zelle unter Beachtung der dualen Lage, vgl.[Ede87]. □

5.3.6 Lemma (Satz von Caratheodory). *Sei $A \subset \mathbb{R}^d$ mit $|A| = n > d$ und x in der konvexen Hülle $\text{CH}(A)$ enthalten. Dann existiert eine $(d+1)$ -elementige Teilmenge $A_1 \subset A$, so dass schon $x \in A_1$ ist.*

Im \mathbb{R}^2 besagt diese Aussage (äquivalent) einfach folgendes: Gegeben sei ein Vieleck, dessen Fläche einen Punkt x enthält. Dann gibt es ein Dreieck aus Eckpunkten des Vielecks, dessen Fläche x enthält.

Beweis. Dieses Lemma ist eine bekannte Aussage über konvexe Polytope, vgl. zum Beispiel Grünbaum. [Grü67, Theorem 2.3.5] \square

5.3.7 Lemma (Beobachtung über Zellen). *Man betrachte den dualen Raum. Sei $\mathcal{D}(h)$ ein Punkt, der im Inneren einer beschränkten Zelle f mit der Eckpunktmenge $\mathcal{D}(G)$ enthalten ist. Dabei sei G eine Menge von Hyperebenen im primalen Raum. Dann ist h eine Hyperebene im primalen Raum, deren Zone $Z(h, G)$ im Korridor von G enthalten ist.*

Beweis. Man betrachte das Arrangement $\mathfrak{A}(G)$. Der Korridor von G enthält alle Zellen aus diesem Arrangement mit Ausnahme einer nach oben beschränkten und einer nach unten beschränkten Zelle. Für Punkte x aus diesen Zellen gilt: x liegt über allen Geraden $g \in G$ bzw. x liegt unter allen Geraden $g \in G$.

Da $\mathcal{D}(h)$ aber in einer beschränkten Zelle f enthalten ist, gibt es mindestens einen Punkt $\mathcal{D}(g_i)$, der über $\mathcal{D}(h)$ und mindestens einen Punkt $\mathcal{D}(g_j)$, der unter $\mathcal{D}(h)$ liegt. Damit folgt: h enthält keine Punkte, die in unbeschränkten Zellen aus $\mathfrak{A}(G)$ liegen. Mit der Definition der Zone $Z(h, G)$ folgt die Behauptung. \square

Damit lässt sich der Einbettungssatz nun beweisen.

Beweis (von Satz 5.3.1). Zu zeigen ist: Es gibt für jede Hyperebene h im \mathbb{R}^d einen $(d+1)$ -Korridor $C := K_{i=1}^{d+1}(g_i^*)$ mit $g_i^* \in H_d(N) \cup \mathbb{R}^d$, für den gilt:

1. $C \cap N = \emptyset$.

2. $C \supset Z(h, N) \stackrel{\text{def}}{=} Z(h, \mathfrak{A}(H_d(n)))$. \square

Der Spezialfall 5.3.2 wurde schon bewiesen. Es lässt sich daher nun annehmen, dass N^+ und N^- nichtleer sind.

Nach Lemma 5.3.5 ist $\mathcal{D}(h)$ in einer beschränkten Zelle mit der Eckpunktmenge $\mathcal{D}(G^{(h)})$ enthalten. Aufgrund von Lemma 5.3.6 gibt es eine Teilmenge aus $d+1$ Punkten, die $D(g_1), \dots, D(g_{d+1})$ enthalten. Nach Lemma 5.3.7 gilt für den Korridor $C := K_{i=1}^{d+1}(g_i^+)$ $Z(h, G^{(h)}) \subset \overline{C}$. Da $G^{(h)} \subset H_d(N)$, ist jede Zelle aus der Zone von h in $\mathfrak{A}H_d(N) = \mathfrak{A}N$ in einer Zelle aus der Zone von h in $\mathfrak{A}(G^{(h)})$ enthalten („Verfeinerung des Arrangements“). Es folgt also die zweite zu zeigende Aussage: $C \supset Z(h, N)$.

Die erste zu zeigende Aussage wurde in Lemma 5.3.4 über Umgebungskorridore nachgewiesen.

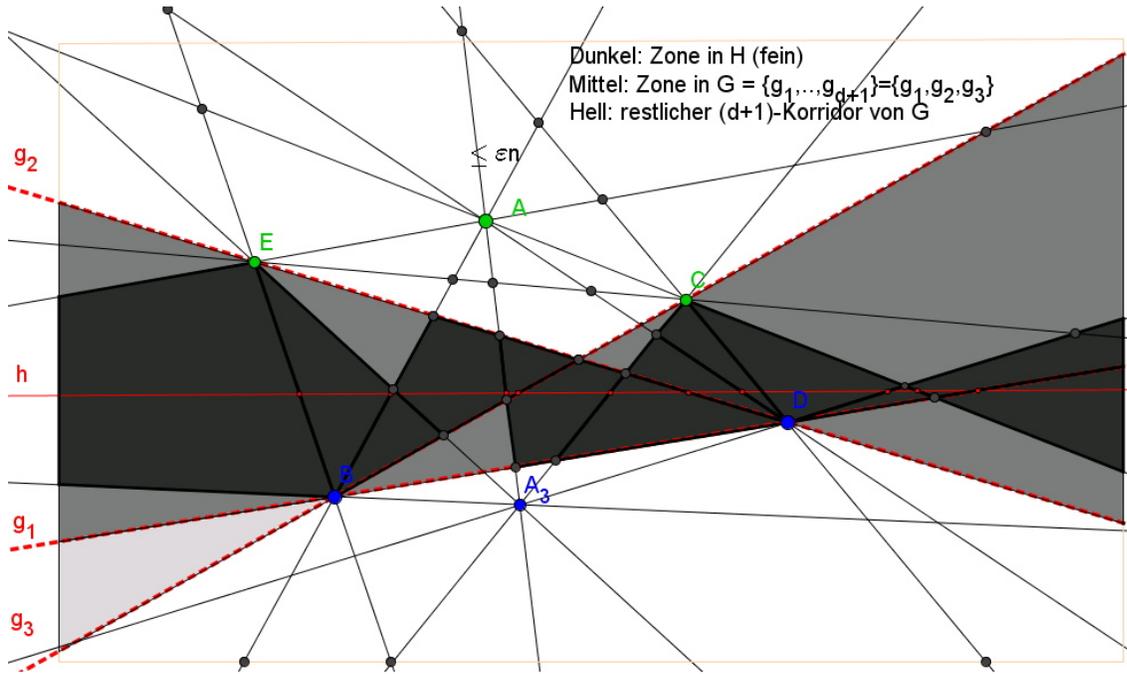


Abbildung 5.3: Korridor

5.4 Geeignete ν -Stichproben endlicher Mengen

Mit den vorigen Ergebnissen gelingt es nun, das zentrale ε -Theorem aufzustellen. Nach diesem gibt es geeignet kleine Teilmengen N von A , so dass nur ein kleiner Teil der Punkte von A in der Zone $Z(h, N)$ einer beliebigen Hyperebene h liegt.

Theorem V (Das zentrale ε -Theorem)

Es gibt zu d eine Konstante c_d^* , so dass im \mathbb{R}^d für alle $0 < \varepsilon < 1/2$ gilt:

- Existenz:** Für jede endliche Menge $A \subset \mathbb{R}^d$ gibt es eine Stichprobe (Teilmenge) $N \subset A$ mit höchstens $\nu = c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$ Elementen, welche die ε -Bedingung von A für Halbräume erfüllt. Das heißt nach Definition 5.2.1 der ε -Bedingung (S.75), dass für eine beliebige Hyperebene h der Bruchteil der Punkte aus A in Schnitzellen kleiner als ε ist. Mit Schnitzellen sind die Zellen aus der Zone $Z(h, N)$ von h in $\mathfrak{A}(N)$ gemeint. Die Aussage lässt sich formal schreiben als

$$\exists N \subset A \text{ mit } |N| \leq c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}, \text{ so dass gilt: } \sum_{f \in Z(h, N)} f \cap A \leq \varepsilon \cdot |A|.$$

- Wahrscheinlichkeit:** Eine zufällige ν -Stichprobe \bar{x} erfüllt die obige Eigenschaft mindestens mit Wahrscheinlichkeit $1/2$.

Es sei erneut auf die erstaunliche Eigenschaft hingewiesen, dass $|N|$ nicht von $|A|$ abhängt, sondern nur von der Konstanten c_d^* und dem vorgegebenen ε .

Beweis. Wir betrachten den Bereichsraum $B_d^{d+1} = (\mathbb{R}^d, \mathcal{K}_d^{d+1})$ der $(d+1)$ -Korridore. Die Konstante c_d^* wird wie im Korridor-Theorem IV (S. 72) mit $k := d+1$ gewählt. Seien $0 < \varepsilon < 1/2$ und $A \subset \mathbb{R}^d$, $|A| = n$ gegeben. Zu zeigen ist für die beiden Aussagen die Implikation:

$$N \text{ ist } \varepsilon\text{-Netz von } A \text{ für } B_d^{d+1} \implies N \text{ erfüllt die } \varepsilon\text{-Bedingung von } A.$$

Diese Folgerung ergibt sich nun aus dem Einbettungssatz 5.3.1 (S. 75): Es existiert stets ein $(d+1)$ -Korridor $C \in \mathcal{K}_d^{d+1}$, der $Z(h, N)$ enthält und für den gilt: $C \cap N = \emptyset$. Nach der Definition von ε -Netz gilt also:

$$|C \cap A| \leq \varepsilon \cdot |A| \implies \sum_{f \in Z(h, N)} |f \cap A| \leq \varepsilon \cdot |A|.$$

Die ε -Bedingung wird also erfüllt. □

Mit dem zentralen ε -Theorem ist die Grundlage für die Existenz der Datenstrukturen geschaffen.

Der folgende Abschnitt befasst sich noch mit einer Grundlage für die Konstruierbarkeit solcher Datenstrukturen. Da die Datenstruktur durch zufällige Stichproben aufgebaut werden wird, drängt sich hierbei die folgende Frage auf: Ist die Zulässigkeit einer Stichprobe verifizierbar?

5.5 Zulässigkeit von Stichproben

Der folgende Satz beantwortet diese Frage positiv, d.h. es lässt sich entscheiden, ob eine Menge $N \subset A$ die ε -Bedingung für A erfüllt.

5.5.1 Satz (Verifikation von Stichproben). *Sei eine n -elementige Menge $A \subset \mathbb{R}^d$ und eine ν -elementige Teilmenge N gegeben. ν sei nicht von n abhängig, d.h. $\nu = O_n(1)$. Dann lässt sich in $\mathcal{O}(|A|) = \mathcal{O}(n)$ Zeit überprüfen, ob die ε -Bedingung eines (ε, ν) -Baum für diesen Knoten mit der Menge N erfüllt ist.*

Beweis. Die Anzahl der Zellen des Arrangements von $\mathfrak{A}(N)$ ist $\mathcal{O}(1)$. Damit lassen sich in $\mathcal{O}(n)$ Zeit die Punkte den Kindknoten A_K zuweisen sowie die Anzahl der Punkte in einer Zelle f bestimmen. Der entscheidende Schritt ist die Überprüfung der (ε, ν) -Bedingung bei Halbraum-Bereichsanfragen. Zunächst muss jede Menge F aus Zellen des Arrangements $\mathfrak{A}(N)$, die Zone einer Hyperebene h sein kann, bestimmt werden. Dann muss überprüft werden, ob für eine solche Menge F stets $\sum_{f \in F} n_f \leq \varepsilon n$ gilt. Haussler und Welzl schreiben, dass dies in $\mathcal{O}(n)$ Zeit möglich sei, da ν und damit die Anzahl der zulässigen Mengen F konstant ist [HW87, S. 145]. Als Ergänzung sei hier betont, dass sich die Menge dieser Mengen F konstruktiv (innerhalb der Zeitgrenzen) bestimmen lässt. Sonst lässt sich

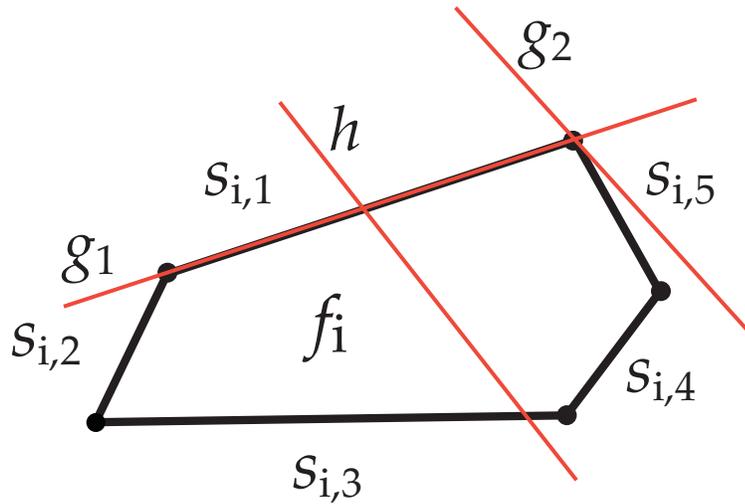


Abbildung 5.4: Schnitt einer Zelle

nicht korrekt entscheiden, ob für jede Hyperebene die ε -Bedingung erfüllt ist. Der letzte Beweisschritt ist die Aussage des nächsten Lemmas. \square

5.5.2 Lemma (Zulässige Zonen). *Seien N mit $|N| = \nu = O_n(1)$ und $\mathfrak{A}(N)$ gegeben. $F(N) = \{f_1, \dots, f_m\}$ ist die Menge der Zellen in $\mathfrak{A}(N)$. Dann lässt sich die Menge \mathcal{Z} der zulässigen Zellmengen $F \subset F(N)$ bestimmen. Eine Zellmenge F heiße zulässig, wenn es eine Hyperebene h mit $Z(h, N) = F$ gibt.*

Beweis. Die Potenzmenge von $F(N)$ hat endlich viele Elemente. Es ist also für jedes ihrer Elemente $F \subset F(N)$ zu entscheiden, ob $F \in \mathcal{Z}$ gilt (oder nicht). Seien zu jeder Zelle f_i die inzidenten (anliegenden) Kanten $s_{i,1}, s_{i,2}, \dots, s_{i,z(i)}$ gegeben. $z(i)$ ist dabei lediglich der Index für die Anzahl der berandenden Kanten der Zelle f_i .

Sei nun eine Menge $F_J = \{f_i | i \in J\}$, $J \subset \{1, \dots, m\}$ aus Zellen vorgegeben. Zu entscheiden ist, ob es eine Hyperebene h gibt, für die gilt:

$$f_i \cap h \neq \emptyset \Leftrightarrow i \in J.$$

Dazu bestimmen wir die Menge der Hyperebenen h mit dieser Eigenschaft und entscheiden, ob sie leer oder nicht-leer ist.

Wir gehen nun zum dualen Raum über und bestimmen die Menge der Punkte $\mathcal{D}(h)$ mit der dualen Eigenschaft. Eine Hyperebene schneidet die (offene) Zelle f_i , wenn eine (offene) anliegende Kante existiert, die geschnitten wird. Wir nehmen nun zur besser handhabbaren Darstellung leicht vereinfachend an, dass eine Zelle auch nur dann geschnitten wird. Eine Präzisierung findet sich dann am Ende dieses Beweistextes. Im Folgenden wird der Beweis anschaulich geführt. In der Abbildung 5.4 zum Schnitt einer Zelle finden sich eine Gerade

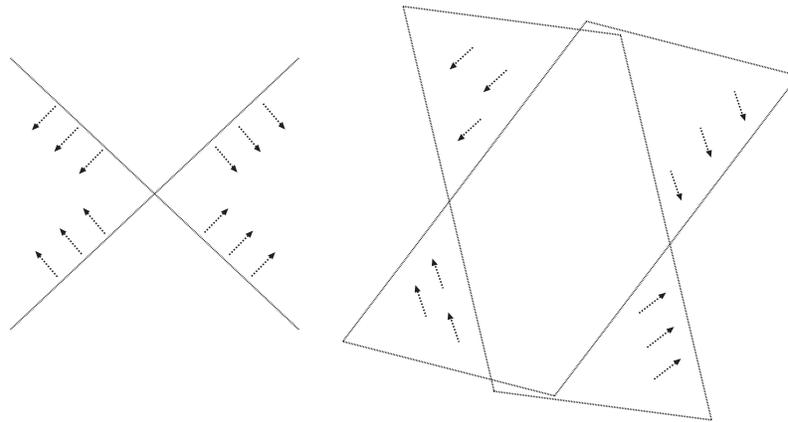


Abbildung 5.5: Doppelkegel

h , die eine Zelle schneiden und zwei Geraden g_1, g_2 , die sie nicht schneiden. In diesen drei Situationen trifft die Annahme zu. Das Duale einer Kante $s_{i,k}$ ist ein offener Doppelkegel $\mathcal{D}(s_{i,k})$, der hier nochmal abgebildet wird. In der Zeichnung links ist der Punkt $\mathcal{D}(h)$ im offenen Doppelkegel enthalten, $\mathcal{D}(g_1)$ und $\mathcal{D}(g_2)$ jedoch nicht. Zu einer Zelle f_i lässt sich nun die Menge der schneidenden Hyperebenen als Punktmenge im dualen Raum angeben:

$$f_i \cap h \neq \emptyset \Leftrightarrow \mathcal{D}(h) \text{ liegt in zu } f_i \text{ gehörigen offenen Doppelkegel} \Leftrightarrow \mathcal{D}(h) \in \mathcal{F}_i.$$

Dabei ist

$$\mathcal{F}_i := \bigcup_{1 \leq k \leq z(i)} \mathcal{D}(s_{i,k})$$

die Vereinigung der zu f_i gehörigen offenen Doppelkegel. Eine Indexmenge

$$J \subset \{i: f_i \in F(N)\} = \{1, \dots, m\} \subset \mathbb{N}$$

ermöglicht also die folgende Charakterisierung: h schneidet genau dann alle f_i mit $i \in J$ (und keine f_i mit $i \notin J$), wenn gilt:

$$\mathcal{D}(h) \in \bigcap_{i \in J} \mathcal{F}_i - \bigcup_{i \notin J} \mathcal{F}_i.$$

Es existiert mithin genau dann eine verlangte Hyperebene h zu einer Menge von Zellen $F_J \subset F(N)$, wenn im dualen Raum gilt:

$$\mathcal{F}_J := \bigcap_{i \in J} \mathcal{F}_i - \bigcup_{i \notin J} \mathcal{F}_i \neq \emptyset.$$

Begründung: Wenn \mathcal{F}_J nicht-leer ist, so enthält \mathcal{F}_J einen Punkt, den wir mit $\mathcal{D}(h)$ bezeichnen. Dieser Punkt $\mathcal{D}(h)$ ist in genau den \mathcal{F}_i mit $i \in J$ enthalten. Das Duale des Punktes ist eine Hyperebene h im primalen Raum. h schneidet im primalen Raum nach

der obigen Charakterisierung von \mathcal{F}_i genau die Zellen von \mathcal{F}_i mit $i \in J$. Wir können also zu einer Menge von Zellen F_J entscheiden, ob es eine Hyperebene h gibt, die genau diese Zellen schneidet. Es ergibt sich schließlich für die Menge \mathcal{Z} der zulässigen Zellmengen:

$$\mathcal{Z} = \{F_J | J \subset \{1, \dots, m\} \text{ mit } \mathcal{F}_J \neq \emptyset\}.$$

Zuletzt beschreiben wir, wie der Beweisweg zu modifizieren ist, wenn man die vereinfachende Annahme durch eine exakte Aussage ersetzt. Folgende geometrische Aussage gilt: Eine Hyperebene h im \mathbb{R}^d schneidet eine Zelle genau dann, wenn eine der beiden Voraussetzungen erfüllt ist:

1. h schneidet zumindest eine anliegende Kante (siehe Beweis) oder
2. h enthält mehrere Eckpunkte, die nicht alle zu derselben Seitenfläche inzident sind.

Die zweite Voraussetzung ist beispielsweise bei einer Gerade h erfüllt, die zwei nicht benachbarte Eckpunkte eines Fünfecks enthält. Sie ist auch erfüllt bei einer Ebene h , die drei Eckpunkte eines Würfels enthält, die nicht zu einer Seitenfläche gehören. (Die Zelle sei das Innere des Fünfecks bzw. Würfels.) Der Beweisweg lässt sich also präzisieren, indem man die duale Menge \mathcal{F}_i für jede Zelle F_i um solche Punkte $\mathcal{D}(h)$ erweitert, für die h die Voraussetzung 2. erfüllt ist: $\widetilde{\mathcal{F}}_i := \mathcal{F}_i \cup \widehat{\mathcal{F}}_i$. Dabei sei $\widehat{\mathcal{F}}_i := \{\mathcal{D}(h) | h \text{ erfüllt Voraussetzung 2.}\}$. Wir deuten noch kurz an, wie sich die Menge $\widehat{\mathcal{F}}_i$ bestimmen lässt. Sei \mathcal{E}_i die Menge der Eckpunkte. Deren duale Entsprechungen sind Hyperebenen. Es lassen sich alle Teilmengen aus mehr als zwei Eckpunkten betrachten, die nicht derselben Seitenfläche inzident ist. Die Vereinigung ihrer dualen Entsprechungen ist die gesuchte Menge $\widehat{\mathcal{F}}_i$. \square

Wir rekapitulieren die entscheidenden Aussagen dieses Kapitels: Nach dem zentralen ε -Theorem (Theorem V, S. 80) existieren zu jeder endlichen Menge $A \subset \mathbb{R}^d$ geeignet kleine Stichproben $N \subset A$, welche die ε -Bedingung erfüllen. Die Existenz solcher Stichproben nimmt die Schlüsselrolle für die Existenz eines Partitionsbaumes für Halbraumanfragen (s. nächstes Kapitel) ein. Nach der Aussage des Satzes 5.5.1 (S. 81) zur Verifikation lässt sich desweiteren überprüfen, ob eine vorgelegte Stichprobe $N \subset A$ die ε -Bedingung erfüllt. Unter der Annahme, dass $\nu = \mathcal{O}_n(1)$ gilt, ergibt sich als Laufzeit einer ν -Stichprobe N von A wie behauptet $\mathcal{O}(n) = \mathcal{O}(|A|)$, da die einzige Operation, die A betrifft, die Zuordnung der Punkte zu Zellen ist. Mit dieser Aussage über die Verifizierbarkeit gelingt es am Ende des nächsten Kapitels auch, einen solchen Partitionsbaum durch einen randomisierten Algorithmus zu konstruieren.

Kapitel 6

Halbraum-Bereichsanfragen

Zur Erinnerung wird im ersten Abschnitt das Grundprinzip von Partitionsbäumen aus Abschnitt 1.4 wiederholt.

Im Abschnitt 6.2 wird der (ε, ν) -Baum, eine parametrisierte Datenstruktur definiert, und die sich ergebende Laufzeit in Abhängigkeit der Parameter untersucht. Erst im folgenden Abschnitt 6.3 zur Existenz von (ε, ν) -Bäumen wird bewiesen, dass tatsächlich (ε, ν) -Bäume mit geeigneten Parametern existieren. Im zentralen Abschnitt 6.4 (S.93) wird dann das Theorem zur Existenz der Datenstruktur für Halbraum-Bereichsanfragen formuliert und bewiesen. Dieses sagt aus, dass geeignete (ε, ν) -Bäume eine Datenstruktur mit sublinearer Laufzeit und linearem Speicher darstellen.

Nach einem Zwischenabschnitt mit Plausibilitätsüberlegungen wird dann im Abschnitt 6.6 die Konstruktion durch einen Las-Vegas-Algorithmus grob skizziert. Es wird nachgewiesen, dass dessen erwartete Laufzeit $\mathcal{O}(n \log n)$ beträgt.

Zuletzt werden in Abschnitt 6.7 die Konstanten abgeschätzt, um die Praktikabilität des Algorithmus einschätzen zu können.

6.1 Wiederholung: Partitionsbäume für Bereichsanfragen

Wir wiederholen das Grundprinzip eines Partitionsbaumes aus Abschnitt 1.4 (S.3).

Gegeben sei eine endliche Menge $A \subset \mathbb{R}^d$. Grundidee ist die Partitionierung (Aufteilung) des \mathbb{R}^d in mehrere offene Regionen, d.h. offene, zusammenhängende Mengen. Diese werden wiederum rekursiv in Teilregionen unterteilt. Jeder Region $r(K)$ wird ein Knoten K zugeordnet. Zudem wird für unsere Anwendung jedem Knoten die Anzahl der in der Region enthaltenen Punkte als Attribut gespeichert. Für Regionen, die nur eine Anzahl von Punkten aus A enthalten, die kleiner oder gleich einer festgelegten Konstanten ist, sind die zugeordneten Knoten Blattknoten. In Blattknoten sind alle aus A enthaltenen Punkte gespeichert und können einzeln abgearbeitet werden.

Damit gilt für einen beliebigen Knoten K und dessen zugehörige Region $r(K)$ die folgende

Beobachtung: Die Menge $A_K := A \cap r(K)$ der Punkte aus A , die in $r(K)$ enthalten sind, besteht aus genau jenen Punkten, die im Teilbaum mit der Wurzel K enthalten sind. Desweiteren ist die Menge der Knoten der gleichen Höhe eine Partitionierung der Punktmenge A .

Um eine Bereichsanfrage mit einem Halbraum h^* zu bearbeiten, beginnt man im Wurzelknoten, dem der gesamte \mathbb{R}^d zugeordnet ist.

Bei der Bearbeitung unterscheidet man für die Kindknoten drei Typen von zugeordneten Regionen: Die Regionen außerhalb des Anfragebereichs h^* werden verworfen. Für die komplett in h^* enthaltenen Regionen addiert man unmittelbar die gespeicherte Anzahl der enthaltenen Punkte zur Ausgabe der Bereichsanfrage. Jene Regionen, die von h^* geschnitten werden, sind genauer zu untersuchen. Für deren Knoten wird die Bereichsanfrage mit h^* rekursiv aufgerufen, so dass ihre Teilregionen nach dem gleichen Schema bearbeitet werden.

6.2 Der (ε, ν) -Baum: ein Partitionsbaum mit Parametern

In diesem Abschnitt wird zunächst eine Klasse von Datenstrukturen, die der (ε, ν) -Bäume, formal definiert und deren Eigenschaften in Abhängigkeit der Parameter untersucht. Später wird gezeigt, dass (ε, ν) -Bäume mit geeigneten ε und ν existieren und eine gewünschte Datenstruktur für Halbraum-Bereichsanfragen darstellen. Die (nicht schwierige) Erweiterung zu Simplex-Bereichsanfragen wird in Kapitel 7 betrachtet.

Grundprinzip des Partitionsbaumes ist es, eine Region mit Punktmenge A durch eine Stichprobe N der Größe ν aufzuteilen. Die Teilregionen sind dabei alle Zellen des Arrangements $\mathfrak{A}(N)$. Mit größerem ν wächst also die Komplexität der Datenstruktur sehr stark an.

Der zweite Parameter ε des (ε, ν) -Baum steht gewissermaßen für die Güte der Stichprobe N : Ein Partitionsbaum wird als (ε, ν) -Baum bezeichnet, wenn in jedem Knoten die Stichprobe N die ε -Bedingung für Hyperebenen erfüllt, d.h. für jede Hyperebene höchstens $\varepsilon \cdot |A|$ Punkte in geschnittenen Regionen liegen. Das Entscheidende für effiziente (sublineare) Bereichsanfragen ist ein guter Kompromiss, die Größe der Datenstruktur bei wachsender Präzision hinreichend niedrig zu halten. Wie sich zeigen wird, bietet das zentrale ε -Theorem einen hinreichend guten Kompromiss.

6.2.1 Definition ((ε, ν)-Baum). Seien ganze Zahlen $d \geq 2$ und $\nu > d$ sowie $0 < \varepsilon < 1/2$ und eine endliche Punktmenge $\widehat{A} \subset \mathbb{R}^d$ gegeben.

Ein Partitionsbaum heißt (ε, ν) -Baum für \widehat{A} , wenn die folgenden Eigenschaften erfüllt sind:

1. Jedem Knoten K ist eine (bezüglich der Standard-Metrik) offene Region $r(K) \subset \mathbb{R}^d$ zugeordnet. (Die Regionen $r(K)$ werden in der Realisierung nur implizit durch die Teilbaumstruktur gespeichert, vgl. dazu auch den Kommentar unter dem Algorithmus 6.2). Die Teilmenge von \widehat{A} innerhalb von $r(K)$ wird mit A_K bezeichnet. (Die

Menge A_K ist genau die Teilmenge der Punkte von A , die im Teilbaum mit Wurzelknoten K enthalten ist.)

2. Der Baum hat eine Wurzel \widehat{K} , für die $r(\widehat{K}) = \mathbb{R}^d$ gilt.
3. Ein Knoten ist ein Blattknoten genau dann, wenn $|A_K| \leq \nu$ gilt.
4. Ein Blattknoten K hat folgende Attribute:
 - $a(K) = |A_K| \leq \nu$: die Anzahl der zugeordneten Punkte
 - $N(K) = A_K$: die Menge der zugeordneten Punkte (Kardinalität $\leq \nu$)
5. Ein innerer Knoten K hat folgende Attribute:
 - $a(K) = |A_K| > \nu$: die Anzahl der Punkte (im Teilbaum)
 - $N(K) \subset A_K$: eine Teilmenge (Stichprobe) von A_K der Kardinalität ν
 - $\mathfrak{A}(K)$ das Arrangement $\mathfrak{A}(N(K))$ aller Hyperebenen, die genau d der Punkte aus $N(K)$ enthalten
 - K_f für jede Zelle f aus $\mathfrak{A}(K)$, die Punkte aus A_K enthält, einen Verweis auf einen Kindknoten K_f , für den gilt: $r(K_f) = f \cap r(K)$
6. Die Teilmenge $N(K)$ genügt der ε -Bedingung für A_K nach Definition 5.2.1 (S.75), d.h. für jede Hyperebene h des \mathbb{R}^d gilt:

$$\sum_{f \in Z(h, \mathfrak{A}(K))} |f \cap A_K| \leq \varepsilon \cdot |A_K|.$$

$Z(h, \mathfrak{A}(K))$ ist dabei die Zone von h in $\mathfrak{A}(K)$, siehe Definition 5.1.1 (S.74). Dies ist die Menge der Zellen f im Arrangement $\mathfrak{A}(K) = \mathfrak{A}(N(K))$, welche h schneiden.

6.2.2 Bemerkung. Haussler und Welzl lassen auch $\varepsilon = 0$ und $1 \geq \varepsilon \geq 1/2$ zu, benötigen diese aber nicht für ihr Verfahren.

6.2.3 Beispiel. Abbildung 6.1 zeigt ein Beispiel für einen Knoten eines (ε, ν) -Baum im \mathbb{R}^2 mit einer Stichprobengröße von $\nu = 5$. Im Beispiel betrachtet wird ein innerer Knoten K mit zugehöriger Region $r(K) = f$. Wie die kommende Analyse zeigt, genügt der Knoten den Erfordernissen eines (ε, ν) -Baum für $\varepsilon = 0,25$ (und damit auch für größere ε) und $\nu = 5$. Die Attribute sind im Beispiel wie folgt belegt:

$r(K)$	das blaue Dreieck
A_K	kleine und rote Punkte, nicht: zusätzliche Geraden-Schnittpunkte
$a(K)$	$= A_K = 51 > 5$
$N(K)$	$\{V_1, V_2, V_3, V_4, V_5\} \subset A_K$
$\mathfrak{A}(K)$	das Arrangement aus den schwarzen Geraden, welche durch je zwei der roten Punkte verlaufen
Kindknoten	K_1, K_2, \dots, K_{22}

In der Abbildung sind sämtliche Zellen aus diesem Arrangement zu erkennen. Es werden

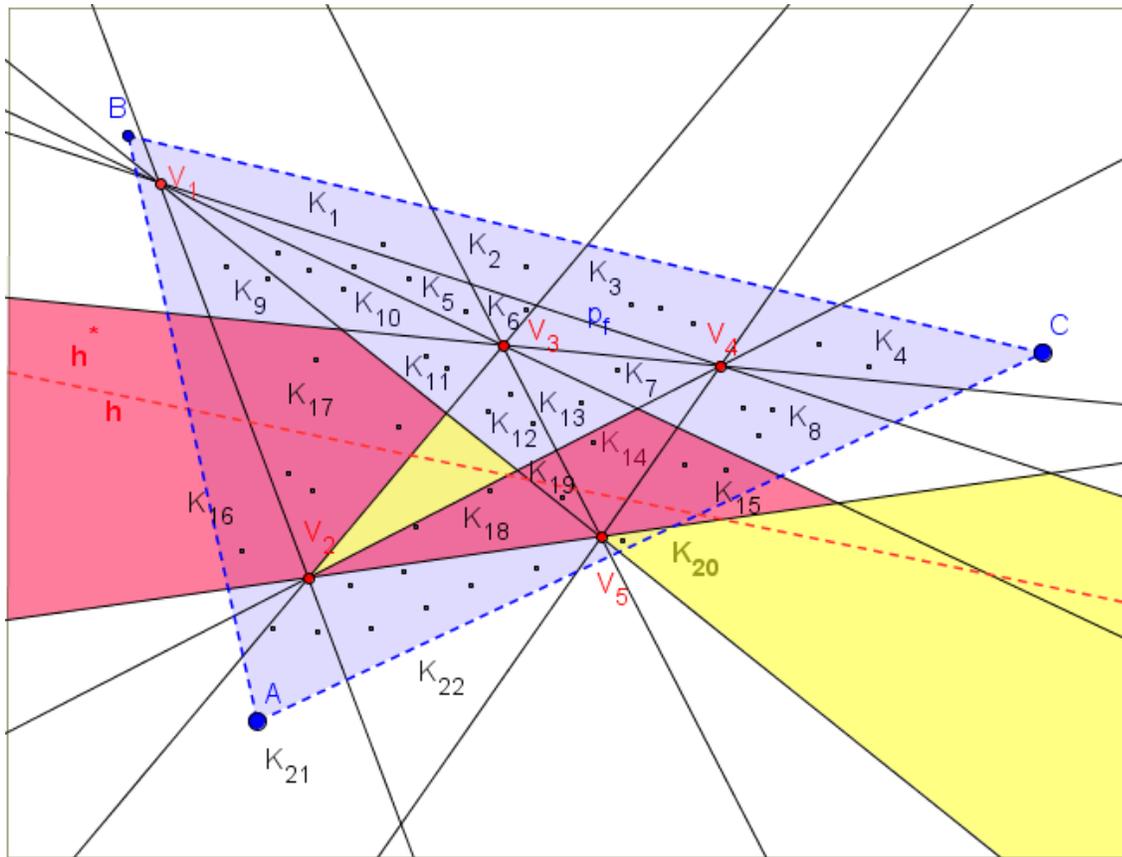


Abbildung 6.1: Aufteilung einer Region und Zone einer Geraden

aber nur jene Kindknoten K_1, K_2, \dots, K_{22} erzeugt, die zu Zellen f aus $\mathfrak{A}(K)$ gehören, welche Punkte aus A_K enthalten. Für diese gilt: $r(K_f) = f \cap r(K)$ ist ein innerer Knoten. Weiter finden sich in der Abbildung eine beispielhafte Gerade h (rot gestrichelt) und eine Halbebene h^* (oberhalb von h). Die Menge $Z(h, \mathfrak{A}(K))$ der von h geschnittenen Zellen ist gelb skizziert. Die Zellen, die tatsächlich Punkte aus h^* enthalten, ohne komplett in h^* enthalten zu sein, sind zusätzlich rot skizziert. Die ε -Bedingung muss für jede Hyperebene (Gerade) für jeden Knoten erfüllt sein, beispielsweise erhält man mit der Hyperebene h für den Knoten K :

$$11 = \sum_{f \in Z(h, \mathfrak{A}(K))} |f \cap A_K| \leq \varepsilon |A_K| = 0,25 \cdot |A_K|.$$

Wenn also diese Bedingung mit $\varepsilon = 0,25$ auch für jede andere Gerade gilt und zudem alle weiteren Knoten einem (ε, ν) -Baum gerecht werden und ein Wurzelknoten \widehat{K} existiert, so ist der vorliegende Partitionsbaum ein (ε, ν) -Baum mit $\nu = 5$ und $\varepsilon = 0,25$ für die zugrundeliegende Menge \widehat{A} .

Bevor die Existenz von (ε, ν) -Bäumen mit geeigneten Parametern ε und ν aus dem zentralen ε -Theorem hergeleitet wird, betrachten wir die Anfragefunktion HALBRAUM-TEILANFRAGE zu einem gegebenen Halbraum h^* und einem Knoten K . Die HALBRAUM-BEREICHSANFRAGE wird dann durch den Aufruf von HALBRAUM-TEILANFRAGE mit dem Wurzelknoten \widehat{K} durchgeführt. Anschließend geben wir eine Rekursionsformel für die Laufzeit an.

1: **Funktion** HALBRAUM-TEILANFRAGE(K, h^*)

Eingabe: Knoten K und Halbraum h^*

- 2: setze $\Sigma \leftarrow |N(K) \cap h^*|$
- 3: **falls** K innerer Knoten ist, **dann** ▷ Sonst in Zeile 2 alles aufgezählt
- 4: **für alle** Zellen f in $\mathfrak{A}(K)$, die Punkte aus A_K enthalten :
- 5: **falls** f in h^* enthalten ist, **dann** ▷ komplett enthalten
- 6: setze $\Sigma \leftarrow \Sigma + a(K_f)$ ▷ alle Punkte zählen
- 7: **sonst: falls** f von h geschnitten wird, **dann** ▷ Redundanz, siehe (†)
- 8: rufe HALBRAUM-TEILANFRAGE rekursiv mit dem Kindknoten K_f auf
- 9: setze $\Sigma \leftarrow \Sigma + \text{HALBRAUM-TEILANFRAGE}(K_f, h^*)$

Ausgabe: $\Sigma = |A_K \cap h^*|$

10: **Prozedur** HALBRAUM-BEREICHSANFRAGE(h^*)

Eingabe: endliche Punktmenge \widehat{A} und Halbraum h^*

- 11: rufe HALBRAUM-TEILANFRAGE mit dem Wurzelknoten \widehat{K} auf

Ausgabe: $|A_{\widehat{K}} \cap h^*| = |\widehat{A} \cap h^*|$

Algorithmus 6.1: Aufzählung der Punkte eines Knotens K , die zu einem Halbraum h^* gehören

(†) Hier wäre bei üblicher Verwendung eines Partitionsbaumes zu testen, ob $r(K)$ von h geschnitten wird. Das ist bei der Realisierung des (ε, ν) -Baum nicht möglich, da $r(K)$ nicht explizit gespeichert wird. Eine Diskussion dieses Details findet sich im Beispielaufruf (Beispiel 6.2.5).

6.2.4 Lemma (Speicherbedarf). *Der Speicherbedarf eines (ε, ν) -Baumes beträgt bei geeigneter Realisierung $\mathcal{O}(\nu^{d^2} \cdot n)$.*

Beweis. Der Speicherbedarf für einen der $\mathcal{O}(n)$ Blattknoten ist $\mathcal{O}(1)$. Da ein Baum mit $\mathcal{O}(n)$ Blattknoten nur $\mathcal{O}(n)$ innere Knoten hat, ist es hinreichend zu zeigen: Der Speicherbedarf eines inneren Knotens ist $\mathcal{O}(\nu^{d^2})$.

Dieser Speicherbedarf ergibt sich aus der Anzahl der Zellen des Arrangements aus $H_d(N)$ nach Lemma 4.1.2 im Kapitel zu geometrischen Bereichsräumen. Es ist nicht notwendig, die Regionen $r(K)$ explizit zu speichern. In einer Realisierung werden also die Attribute $a(K)$, $N(K)$ sowie $\mathfrak{A}(K)$ gespeichert. Bei einer solchen Realisierung von $\mathfrak{A}(K)$ ist der Speicherbedarf $\mathcal{O}(\nu^{d^2})$ [HW87, vgl. Abschnitt 4, S.140 u.]. Eine geeignete Realisierung eines d -dimensionalen Arrangements findet sich bei Edelsbrunner, wobei $n \leq \nu^d$ die Anzahl der Hyperebenen ist [Ede87, Kapitel 7]. \square

6.2.5 Beispiel. Man betrachte erneut die Abbildung aus Beispiel 6.2.3. Der Algorithmus 6.2 zeigt den Aufruf von HALBRAUM-TEILANFRAGE(K, h^*). h^* ist hier die offene obere Halbebene der Geraden h aus Beispiel 6.2.3. Die Kindknoten sind dort mit K_1, \dots, K_{22} bezeichnet. Seien die zugehörigen Zellen aus $\mathfrak{A}(K)$ entsprechend f_1, \dots, f_{22} .

Beim folgenden Beispielablauf (Algorithmus 6.2) Der Aufruf für K_{20} in Zeile 5 ist redun-

1: **Funktion** HALBRAUM-TEILANFRAGE(K, h^*)

Eingabe: Knoten K mit der blau eingefärbten Region $r(K)$ sowie der Halbraum h^* oberhalb der Geraden h

2: setze $\Sigma \leftarrow |N(K) \cap h^*| = |\{V_1, V_3, V_4\}| = 3$ $\triangleright K$ ist ein innerer Knoten

3: **für alle** komplett enthaltenen Zellen $f_j, j = 1, \dots, 13$:

4: Setze $\Sigma \leftarrow \Sigma + a(K_{f_j})$ \triangleright alle Punkte hinzuzählen

5: **für alle** geschnittenen Zellen $f_j, j = 16, \dots, 20$:

6: rufe HALBRAUM-TEILANFRAGE rekursiv mit dem Kindknoten K_{f_j} auf

7: setze $\Sigma \leftarrow \Sigma + \text{HALBRAUM-TEILANFRAGE}(K_{f_j}, h^*)$

Ausgabe: $\Sigma = |r(K) \cap h^*|$

Algorithmus 6.2: Beispielablauf für Algorithmus 1

dant, da HALBRAUM-TEILANFRAGE(K_{20}, h^*) den Rückgabewert 0 hat. Solche rekursiven Aufrufe entstehen durch die Bedingung „ f_j wird von h geschnitten“. Diese Bedingung ersetzt die inhaltlich konsequente Bedingung „ $r(K) \cap f_j$ “. Grund dafür ist, dass die Regionen

$r(K)$ recht komplex werden und damit zu aufwendige Schnitttests produzieren können. Durch die Verwendung dieser Bedingung kann auf die Berechnung und Speicherung der Regionen $r(K)$ verzichtet werden. Nachteile für die Laufzeit haben die auftretenden redundanten Tests nicht, da in den Analysen stets die Anzahl *aller* Zellen aus dem Arrangement zur Abschätzung verwendet wird.

Für die Anfragezeit in Abhängigkeit der Parameter ergibt sich:

6.2.6 Lemma (Anfragezeit(rekursiv)). *Sei $T(K, h^*)$ die Anfragezeit, die der Aufruf HALBRAUM-TEILANFRAGE(K, h^*) (Algorithmus 6.1) benötigt. Dann gilt:*

1. Für innere Knoten K

$$T(K, h^*) \leq \mathcal{O}(\nu^{d^2}) + \sum_{f \in Z(h, \mathfrak{A}(K))} T(K_f, h^*).$$

2. $|Z(h, \mathfrak{A}(K))| \leq \nu^{d(d-1)}$.

3. Für Blattknoten $T(K, h^*) \leq \mathcal{O}(\nu)$.

Beweis. In Blattknoten K sind alle Punkte direkt gespeichert: $N(K) = A_K$ und es ist $|A_K| \leq \nu$. Für jeden Punkt lässt sich in $\mathcal{O}(1)$ Laufzeit überprüfen, ob er in h^* enthalten ist. Damit ergibt sich die Aussage für Blattknoten. Die Anzahl der Zellen im Arrangement ist nach 4.1.2 kleiner oder gleich $\nu^{d(d-1)}$. Da die Zone eine Teilmenge dieser Zellen ist, folgt die Abschätzung von $|Z(h, \mathfrak{A}(K))| \leq \nu^{d(d-1)}$. Nun beweisen wir

$$T(K, h^*) \leq \mathcal{O}(\nu^{d^2}) + \sum_{f \in Z(h, \mathfrak{A}(K))} T(K_f, h^*).$$

Die Laufzeit für sämtliche rekursiven Aufrufe ist $\sum_{f \in Z(h, \mathfrak{A}(K))} T(K_f, h^*)$. Es ist zu zeigen, dass die übrige Laufzeit $\mathcal{O}(\nu^{d^2})$ ist. Die entscheidende Passage ist dabei die Traversierung des Arrangements in den Zeilen 4 bis 7 von Algorithmus 6.1. Bei einer geeigneten Realisierung von $\mathfrak{A}(K)$ ist die Laufzeit für diese Traversierung des Arrangements linear in der Anzahl der Zellen, d.h. $\mathcal{O}(\nu^{d(d-1)})$. Für eine genauere Untersuchung sei erneut auf Edelsbrunner verwiesen [Ede87, Kapitel 7]. Damit folgt die Behauptung. \square

Es ist nun zwingend notwendig eine explizite Abschätzung zu finden. Diese wird im Abschätzungstheorem X unter Verwendung der Parameter durchgeführt. Wegen der oberen Schranke für die Anzahl der Zellen eines Arrangements in Beobachtung 4.1.2 genügt die maximale Anzahl der rekursiven Aufrufe $|F|$ der Schranke $m \leq \Phi_{d-1} \left(\binom{\nu}{d} \right) \leq \nu^{d(d-1)}$. ν ist dabei nicht von n abhängig, also in diesem Sinne konstant. Unter Beachtung dieser parametrischen Beschränkungen gelingt es Haussler und Welzl mit sehr umfangreichem und ausgefeiltem Einsatz analytischer Mittel, eine explizite parametrisierte Schranke herzuleiten. Diese Herleitung verschieben wir auf den Anhang A, in dem das folgende Theorem

gezeigt wird.

Vorgriff: Abschätzungstheorem X (Auflösung der Rekursion, S.119)

Seien $0 < \varepsilon < 1/2$, $c \in \mathbb{N}$ und $2 \leq m \in \mathbb{N}$ gegeben.

Sei $Q: \mathbb{N} \rightarrow \mathbb{N}$ mit

$$Q(n) = c + \max \sum_{i=1}^m Q(n_i); \quad Q(0) = 0,$$

wobei das Maximum über allen m -Tupeln aus natürlichen Zahlen (n_1, \dots, n_m) mit $\sum_{i=1}^m n_i \leq \varepsilon n$ zu bilden ist. Dann gilt:

$$Q(n) = \mathcal{O}(c \cdot n^\alpha) \text{ mit } \alpha(\varepsilon, m) := 1 - \frac{1}{\log_{1/\varepsilon} m + 1}.$$

Insbesondere ergibt sich für $m \leq \Phi_{d-1} \left(\binom{\nu}{d} \right) \leq \nu^{d(d-1)}$ und $c = \mathcal{O}(\nu^{d^2})$:

$$Q(n) = \mathcal{O} \left(\nu^{d^2} \cdot n^{\alpha(\varepsilon, \nu)} \right).$$

Aus Lemma 6.2.6 und dem Abschätzungstheorem X ergibt sich dann:

6.2.7 Satz (Halbraum-Anfragezeit in (ε, ν) -Bäumen). *Sei eine n -elementige Menge $A \subset \mathbb{R}^d$ und ein (ε, ν) -Baum für A gegeben. Dann lassen sich mit diesem (ε, ν) -Baum Halbraum-Bereichsanfragen beantworten mit der Anfragezeit*

$$T(n) = \mathcal{O}(\nu^{d^2} \cdot n^{\alpha(\varepsilon, \nu)}) \text{ mit } \alpha(\varepsilon, \nu) = 1 - \frac{1}{d(d-1) \cdot (\log_{1/\varepsilon} \nu) + 1}.$$

Beweis. [HW87, Lemma 4.3]

Man betrachte den Aufruf von HALBRAUM-TEILANFRAGE(K, h^*) für einen inneren Knoten K , in dessen Region $n' = |A_K|$ Punkte liegen. Nach dem Lemma 6.2.6 über die rekursive Anfragezeit gilt:

$$T(K, h^*) \leq \mathcal{O} \left(\nu^{d^2} \right) + \sum_{f \in F} T(K_f, h^*).$$

Seien n_1, \dots, n_m die Anzahlen der Punkte in den geschnittenen Regionen, die durch Aufrufe HALBRAUM-TEILANFRAGE(K_f, h^*) bearbeitet werden. Dann gilt nach der ε -Bedingung in der Definition 6.2.1 von (ε, ν) -Bäumen (S.86): $\sum_{i=1}^m n_i \leq \varepsilon n'$ mit $m \leq \nu^{d(d-1)}$. Damit genügt $T(n)$ den Voraussetzungen aus dem Abschätzungstheorem X mit $m \leq \nu^{d(d-1)}$ und einem geeignet großen $c = \mathcal{O}(\nu^{d^2})$. c sei dabei so groß, dass der Aufruf von HALBRAUM-TEILANFRAGE(K, h^*) mit beliebigen Blattknoten K eine Laufzeit kleiner als c hat. Dies ist möglich, da sich nach Lemma 6.2.6 Blattknoten in $\mathcal{O}(\nu)$ Zeit bearbeiten lassen. Damit folgt die Aussage. \square

Damit haben wir die Anfragezeit eines (ε, ν) -Baumes hinreichend gut beschrieben. Mit hinreichend ist hier gemeint, dass sich bei Verwendung von geeigneten Parametern eine Halbraum-Bereichsanfrage mit sublinearer Laufzeit durchführen lässt. Die Existenz von (ε, ν) -Bäumen mit derartig geeigneten Parametern wird im Folgenden diskutiert.

6.3 Existenz von (ε, ν) -Bäumen

In diesem Abschnitt wird die Existenz von (ε, ν) -Bäumen für $0 < \varepsilon < 1/2$ und geeignet kleine ν aus dem zentralen ε -Theorem V (S.80) hergeleitet. Wir wiederholen den ersten Teil des zentralen ε -Theorems:

Es gibt zu d eine Konstante c_d^* , so dass für $0 < \varepsilon < 1/2$ gilt:

Für jede endliche Menge $A \subset \mathbb{R}^d$ gibt es eine ν -elementige Teilmenge (Stichprobe) $N \subset A$ mit $\nu = c_d^* \cdot \frac{1}{\varepsilon} \cdot \log \frac{1}{\varepsilon}$, welche die ε -Bedingung von A für Halbräume erfüllt, d.h.:

$$\sum_{f \in Z(h, N)} |f \cap A| \leq \varepsilon \cdot |A|.$$

Aus dem zentralen ε -Theorem folgt:

6.3.1 Satz (Existenz von kleinen (ε, ν) -Bäumen).

Zu jedem $0 < \varepsilon < 1/2$ und jeder endlichen Menge $\hat{A} \subset \mathbb{R}^d$ existiert ein (ε, ν) -Baum mit

$$\nu = \nu(\varepsilon) := c_d^* \cdot \frac{1}{\varepsilon} \cdot \log \frac{1}{\varepsilon}.$$

Beweis. Es ist nur zu zeigen, dass sich in jedem inneren Knoten K eine Teilmenge $N(K)$ aus ν Elementen wählen lässt, welche die ε -Bedingung für A_K erfüllt.

Da $\nu = \nu(\varepsilon)$ wie im zentralen ε -Theorem V gewählt ist, folgt die Behauptung aus dem zentralen ε -Theorem V (S.80). \square

Im folgenden Abschnitt zeigen wir, dass (ε, ν) -Bäume mit $\nu = \nu(\varepsilon)$ für kleine ε eine $\mathcal{O}(n)$ -Datenstruktur mit der gewünschten Anfragezeit sind. Dabei ist nur das zentrale ε -Theorem V anzuwenden.

6.4 Zur Existenz der Datenstruktur

In diesem Abschnitt weisen wir nach, dass es eine Datenstruktur mit linearem Speicher und sublinearer Anfragezeit für Halbraum-Bereichsanfragen im \mathbb{R}^d gibt. Genauer formuliert wird dies im folgenden Theorem:

Theorem VI (Existenz der Datenstruktur) *Für alle $\gamma > 0$ gilt: Es gibt eine $\mathcal{O}(n)$ -Datenstruktur, die zu allen n -elementigen Punktmengen \hat{A} Halbraum-Bereichsanfragen mit Anfragezeit $T(n) = \mathcal{O}(n^{\beta+\gamma})$ ermöglicht. Dabei ist $\beta = 1 - \frac{1}{d(d-1)+1}$.*

Beweis. [HW87, Theorem 4.9]

Sei $\gamma > 0$. Nach dem vorigen Satz 6.3.1 gibt es für alle $0 < \varepsilon < 1/2$ einen (ε, ν) -Baum mit $\nu = c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$. Gezeigt wird, dass ein solcher (ε, ν) -Baum für hinreichend kleines ε eine geforderte Datenstruktur ist. Sei also $0 < \varepsilon < 1/2$ fest und $\nu = \nu(\varepsilon) := c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} = \mathcal{O}_n(1)$. Aufgrund von Lemma 6.2.4 (S.90) ist der Speicherbedarf des (ε, ν) -Baum gegeben durch

$\mathcal{O}(\nu^{d^2} n) = \mathcal{O}(n)$. Nach Lemma 6.2.7 (S.92) ist die Halbraum-Anfragezeit $\mathcal{O}(\nu^{d^2} \cdot n^{\alpha(\varepsilon)}) = \mathcal{O}(n^{\alpha(\varepsilon)})$ mit

$$\alpha(\varepsilon) := \alpha(\varepsilon, \nu(\varepsilon)) = 1 - \frac{1}{d(d-1) \log_{1/\varepsilon} \nu(\varepsilon) + 1}.$$

Den entscheidenden logarithmischen Term bezeichnen wir mit

$$w(\varepsilon) := \log_{1/\varepsilon} \nu(\varepsilon)$$

und zeigen $\lim_{\varepsilon \rightarrow 0} w(\varepsilon) = 1$. Dazu formen wir wie folgt um:

$$\begin{aligned} w(\varepsilon) &= \log_{1/\varepsilon} \nu(\varepsilon) = \log_{1/\varepsilon} \left[c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \right] \geq \log_{1/\varepsilon} c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \\ &= \log_{1/\varepsilon} c_d^* + \log_{1/\varepsilon} \frac{1}{\varepsilon} + \log_{1/\varepsilon} \log \frac{1}{\varepsilon} \\ &= \log_{1/\varepsilon} c_d^* + 1 + \log_{1/\varepsilon} \log \frac{1}{\varepsilon} \\ \implies \lim_{\varepsilon \rightarrow 0} w(\varepsilon) &= 1 + \lim_{\varepsilon \rightarrow 0} \log_{1/\varepsilon} c_d^* + \lim_{\varepsilon \rightarrow 0} \log_{1/\varepsilon} \log \frac{1}{\varepsilon} = 1 + 0 + 0 = 1. \end{aligned}$$

Es ist also $\lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) = 1 - \frac{1}{d(d-1)+1}$. Aus Stetigkeitsgründen kann nun ε so gewählt werden, dass gilt:

$$\alpha(\varepsilon) < 1 - \frac{1}{d(d-1)+1} + \gamma.$$

Damit ist ein (ε, ν) -Baum mit einem solchen ε und $\nu = \nu(\varepsilon)$ eine derartige Datenstruktur. \square

6.5 Plausibilitätsüberlegungen

In diesem Abschnitt wird rückblickend plausibel gemacht, warum sich ε -Netze für sublineare Halbraum-Bereichsanfragen eignen. Es wird erörtert, dass mit der hier durchgeführten Verwendung von ε -Netzen keine weitere Laufzeitverbesserung möglich ist. Für das Erreichen der Laufzeit ist notwendig, dass $w(\varepsilon)$ gegen 1 konvergiert. Bezeichne mit $\nu(\varepsilon)$ das kleinste ν , so dass zu ε ein (ε, ν) -Baum existiert. Aus der Definition ergibt sich, dass $w(\varepsilon)$ dann gegen 1 konvergiert, wenn $w(\varepsilon)$ außer einem konstantem Faktor und einem linearen Faktor $\frac{1}{\varepsilon}$ nur einen polylogarithmischen Faktor $\log^{\mathcal{O}(1)} \frac{1}{\varepsilon}$ enthält. Nach Kapitel 3 existieren stets ε -Approximationen mit $\mathcal{O}(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon})$ Elementen. Dieser Wert ist also zu groß, da ein quadratischer Faktor $(\frac{1}{\varepsilon})^2$ vorliegt. Es ist denkbar, dass Haussler und Welzl zunächst versucht haben, mit ε -Approximationen Halbraumbereichsanfragen mit sublinearer Laufzeit durchzuführen, und dann erkannt haben, dass ε -Approximationen eine solche aufgrund ihrer Größe nicht zulassen.

Entscheidend ist jedenfalls die Idee, mit allgemeinen Bereichsräumen (k -Korridoren) für beliebig kleine ε eine ε -Bedingung zu garantieren und dabei die Anzahl der Teilregionen durch eine Potenz zu beschränken. Die Basis soll dabei eine konstante (von A unabhängige) Größe haben, welche nur schwach mit $1/\varepsilon$ wächst. Der Exponent sollte durch einen

möglichst kleine (von der Dimension abhängige) Konstante beschränkt sein.

Als Laufzeit ergibt sich dann wegen der Rekursionsgleichung eine Laufzeit $\mathcal{O}(n^{\tilde{\alpha}})$ mit

$$\tilde{\alpha} = 1 - \frac{1}{1 + \log_{1/\varepsilon} \mu^k} = 1 - \frac{1}{1 + k \cdot \log_{1/\varepsilon} \mu}.$$

Falls $\log_{1/\varepsilon} \mu$ gegen 1 konvergiert, erhält man dann eine beliebige Annäherung an die Laufzeit $\mathcal{O}(n^{\tilde{\alpha}})$ mit $\tilde{\alpha} = 1 - \frac{1}{1+k}$.

Bei der Durchführung von Haussler und Welzl mit *Random Sampling* wird eine Stichprobe N der Größe ν verwendet und das Arrangement $\mathfrak{A}(N)$ konstruiert. Für die Anzahl der Teilregionen gilt die Abschätzung $m \leq \nu^{d(d-1)}$. Der Term $d(d-1)$ in der Laufzeit $\mathcal{O}(n^\alpha)$ ergibt sich also aus dieser Abschätzung.

Eine bessere Laufzeit, die näher an der unteren Schranke $\mathcal{O}(n^{1-1/d}/\log n)$ liegt, lässt sich also mit dieser Durchführung auch nicht erreichen, wenn man kleinere ε -Netze findet.

Matoušek erreicht die Laufzeit $\mathcal{O}(n^{1-1/d})$ mit einer linearen Datenstruktur bis auf polylogarithmische Faktoren durch einen verbesserten Partitionsbaum [Mat92]. Das Verfahren verwendet unter anderem auch ε -Netze, interessanterweise aber auch das ältere Stichprobenkonzept der ε -Approximationen, um eine der ε -Bedingung ähnliche Eigenschaft nachzuweisen [Mat92, Cutting Lemma, Partition Theorem].

Zentral ist dabei ein weiteres Stichprobenkonzept der ε -*Cuttings*. Matoušek gelingt es im Gegensatz zum hier untersuchten Verfahren auch, eine deterministische Konstruktion mit der Laufzeit $\mathcal{O}(n \log n)$ anzugeben. Dieses geschieht durch eine Derandomisierung.

Es sei an dieser Stelle nochmals darauf hingewiesen, dass für die Laufzeit nach Haussler und Welzl ganz entscheidend ist, nur k -Korridore mit einem konstanten (d.h. von n und ε unabhängigen) k zu verwenden, da ansonsten die Dimension des Bereichsraumes mit k wächst und damit auch die notwendige Stichprobengröße ν .

6.6 Konstruktion der Datenstruktur

Eine Konstruktion eines (ε, ν) -Baumes mit $\mathcal{O}(n \log n)$ erwarteter Laufzeit ergibt sich durch Algorithmus 6.3. Bei diesem Las-Vegas-Algorithmus wird in jedem Knoten K so lange eine zufällige Stichprobe gewählt, bis diese ein ε -Netz von A_K ist.

6.6.1 Lemma (Laufzeit des Konstruktions-Algorithmus). *Sei ein Knoten K und eine endliche Menge A_K gegeben. Dann hat der Aufruf des Algorithmus 6.3 mit dem Knoten K eine erwartete Laufzeit von $\mathcal{O}(|A_K| \log |A_K|)$.*

Die Laufzeit ist nur abhängig von A_K . Der Parameter ν ist vorgegeben mit $\mathcal{O}(\nu) = \mathcal{O}(1)$.

Beweis. Wir zeigen: Die Konstruktion eines Knotens mit der vorgegebenen Punktmenge A_K ist in $\mathcal{O}(|A_K|)$ Laufzeit möglich. Nur der erste Teil bis Zeile 10 ist randomisiert. Die

1: **Prozedur** KONSTRUKTION(K)
Eingabe: endliche Punktmenge A_K

2: setze $a(K) \leftarrow |A_K|$

3: **falls** $|A_K| < \nu$, **dann** ▷ Der Knoten ist ein Blattknoten

4: setze $N(K) \leftarrow A_K$

5: **sonst:**

6: **wiederhole**

7: erzeuge aus A_K eine zufällige ν -Stichprobe $N(K)$ ▷ $\mathcal{O}(\nu) = \mathcal{O}(1)$

8: entferne die Duplikate aus $N(K)$

9: erzeuge das Arrangement $\mathfrak{A}(N(K))$ aus $N(K)$

10: **bis** die ε -Bedingung im Knoten K durch die Stichprobe $N(K)$ erfüllt wird, d.h. für jeden Halbraum h^* die Summe der Punkte aus geschnittenen Zellen höchstens $\varepsilon \cdot |A_K|$ ist

11: erzeuge das Arrangement $\mathfrak{A}(K)$

12: teile A_K auf in $N(K)$ und die Teilmengen A_{K_f} für alle Zellen im Arrangement

13: **für alle** Zellen f mit $A_{K_f} \neq \emptyset$:

14: erzeuge einen Verweis auf einen anzulegenden Kindknoten K_f

15: erzeuge den Knoten K_f durch den rekursiven Aufruf KONSTRUKTION(K_f)

Algorithmus 6.3: probabilistische Konstruktion eines Knotens K

Überprüfung, ob die ε -Bedingung erfüllt wird, lässt sich nach dem Satz 5.5.1 zur Verifikation (S.81) in $\mathcal{O}(|A_K|)$ Zeit durchführen.

Im zweiten Teil ab Zeile 11 wird für den Knoten K mit der gefundenen Stichprobe $N(K)$ die Datenstruktur realisiert.

Wie oben bemerkt, ist $\mathcal{O}(\nu) = 1$ und das Arrangement hat eine konstante Größe. Die einzige Operation, die mehr als $\mathcal{O}(1)$ Zeit in Anspruch nimmt, ist also die Aufteilung der Menge A_K . Ein Element lässt sich in $\mathcal{O}(1)$ Zeit zuweisen, denn das Arrangement lässt sich in konstanter Zeit traversieren und der Test, ob ein Punkt in einer Zelle f liegt, benötigt nur konstante Zeit.

Insgesamt ist die Laufzeit des Aufrufes mit einem Knoten K also $\mathcal{O}(|A_K|)$. □

Mit dem obigen Lemma lässt sich das folgende Theorem beweisen:

Theorem VII (Konstruktion der Datenstruktur) . Sei ein $\gamma > 0$ sowie eine n -elementige Menge $\hat{A} \subset \mathbb{R}^d$ vorgelegt. Dann lässt sich eine Datenstruktur mit linearem Speicher zur Beantwortung von Halbraum-Bereichsanfragen mit einer Anfragezeit von $\mathcal{O}(n^{\beta+\gamma})$ konstruieren, wobei $\beta = (1 - \frac{1}{d(d-1)+1})$ ist. Dies ist möglich durch einen Las-Vegas-Algorithmus mit $\mathcal{O}(n \log n)$ erwarteter Laufzeit.

Beweis (von Theorem VII). Haussler und Welzl begründen die Konstruktionszeit nur sehr knapp. Die folgenden Ausführungen sind umfangreicher.

Wähle c_d^* so wie im zentralen ε -Theorem V (80). Dann erfüllt eine zufällige ν -Stichprobe von A_K stets mit einer Wahrscheinlichkeit von mindestens $1/2$ die ε -Bedingung.

Wähle $0 < \varepsilon < 1/2$ so, dass ein (ε, ν) -Baum mit $\nu = c_d^* \cdot 1/\varepsilon \log 1/\varepsilon$ nach Theorem VI (S.93) die Laufzeit $\mathcal{O}(n^{\beta+\gamma})$ mit $\beta = 1 - \frac{1}{d(d-1)+1}$ hat. Zur Konstruktion des (ε, ν) -Baumes wird der Konstruktionsalgorithmus mit dem Wurzelknoten \widehat{K} und der Menge \widehat{A} aufgerufen.

Laufzeit des Algorithmus: Die Laufzeit für Blattknoten ist $\mathcal{O}(1)$ und braucht nicht weiter betrachtet zu werden. In jedem inneren Knoten K ist die erwartete Laufzeit $\mathcal{O}(|A_K|)$. Die Behauptung folgt nun durch Aufsummieren: Auf jedem Level des Partitionsbaumes sind die Mengen A_{K_i} disjunkt. Also ist die erwartete Laufzeit für die Konstruktion aller Knoten K_i eines Levels gegeben durch

$$\sum_i \mathcal{O}(|A_{K_i}|) \subset \mathcal{O}(|\widehat{A}|) = \mathcal{O}(n).$$

Da ein Baum mit $\mathcal{O}(n)$ Blattknoten höchstens die Höhe $\log n$ hat, folgt die Behauptung: Der Konstruktionsalgorithmus hat eine erwartete Laufzeit von $\mathcal{O}(n \log n)$. \square

Wir beenden den Abschnitt zur Konstruktion der Datenstruktur für Halbraum-Bereichsanfragen mit zwei Bemerkungen.

6.6.2 Bemerkung. Es ist zu beachten, dass der formale Erwartungswert für die Anzahl der Wiederholungen nur auf einem abzählbaren Produktraum definiert ist. Dieses haben wir im Grundlagenkapitel nicht behandelt. Desweiteren lässt sich der formale Erwartungswert für die Gesamtlaufzeit nicht leicht erklären, da die jeweiligen erwarteten Laufzeiten von den vorherigen Wahlen der Stichproben abhängen. Für eine formale Definition des Erwartungswertes benötigt man daher ein Wahrscheinlichkeitsmaß für Bäume. Da die erwartete Laufzeit von Algorithmen auch in den meisten Büchern wie von Raghavan und Motwani nicht formal definiert wird, wird dieses Wahrscheinlichkeitsmaß hier nicht eingeführt, sondern nur an dieser Stelle auf die Problematik aufmerksam gemacht.

6.6.3 Bemerkung. Da ν mit $1/\varepsilon$ wächst, hat man eine enorme Vergrößerung der Laufzeit für klein gewählte γ . Dies ist nicht weiter relevant, da ein direkter praktischer Einsatz des Algorithmus ohnehin nicht realistisch ist. Dies wird durch die Abschätzung der Konstanten im letzten Abschnitt ersichtlich. Da stets zufällige Stichproben gezogen werden, ist nicht gesichert, dass der Algorithmus determiniert. (Die Wahrscheinlichkeit, dass er determiniert, ist jedoch 1.) Um das Determinieren zu garantieren, gibt es die Möglichkeit, nach k unzulässigen zufälligen Stichproben abzubrechen und systematisch alle Stichproben zu überprüfen. Die Wahrscheinlichkeit, nach k Stichproben keine zulässige zu erhalten, ist kleiner oder gleich $\frac{1}{2^k}$. Wählt man k groß genug, so bleibt die erwartete Laufzeit gleich.

6.7 Abschätzung der Konstanten

In diesem Abschnitt wird im Wesentlichen erörtert, dass die vorkommenden Konstanten sowohl für den Speicherbedarf als auch für die Anfragezeit zu groß sind. Schon bei sehr kleinen Datenmengen sind der notwendige Speicherbedarf und die notwendige Anfragezeit enorm groß. Dies gilt, obwohl die asymptotische Laufzeit in $O(n^\alpha) = O(n^{\beta+\gamma})$ mit $\beta = 1 - \frac{1}{d(d-1)+1}$ nur unwesentlich kleiner 1 ist. Insofern könnte sie erst bei Datenmengen A mit sehr großem $n = |A|$ einen Vorteil ermöglichen. Dies trifft erst recht für Dimensionen $d > 2$ (oder $d > 3$) zu, da β dann näher an 1 liegt. Für den speziellen Fall $d = 2$ sind außerdem deutlich bessere angepasste Algorithmen möglich.

Zudem ist ein geeigneter (ε, ν) -Baum enorm schwierig zu implementieren. Für den Aufwand zur Konstruktion und Traversierung eines d -dimensionalen Arrangements sei auf [Ede87, Kapitel 7] verwiesen.

Insofern ist die praktische Anwendung dieses Verfahrens zur Bearbeitung von Halbraum-Bereichsanfragen keineswegs realistisch. Die Datenstruktur stellt eher eine obere Schranke für die Komplexität des Problems dar. Es sei an dieser Stelle noch angemerkt, dass Matoušek [Mat92] diese obere Schranke bis auf polylogarithmische Faktoren zu $O(n^{1-1/d})$ verbessert hat.

Die vorkommenden Konstanten hängen vom gewählten $\gamma > 0$ ab und werden für entsprechend klein gewählte γ beliebig groß. Es sind jedoch nur solche γ sinnvoll, für die $\alpha < 1$ ist, da sonst selbst der triviale Scan eine bessere Laufzeit darstellt. Zu einem gewählten γ sind ein $\varepsilon = \varepsilon(\gamma)$ und zu diesem ε ein $\nu = \nu(\varepsilon)$ derart zu bestimmen, dass der (ε, ν) -Baum nach Theorem VI (S.93) die Anfragezeit $O(n^{\beta+\gamma})$ realisiert. Die Abschätzung der vorkommenden Konstanten kann nach verschiedenen Maßstäben geschehen. Insbesondere kann man verbesserte Resultate für die kleinsten ε -Netze verwenden, vgl. Kapitel 8. Es ist auch völlig ungewiss, welche VC-Dimension die verwendeten $(d+1)$ -Korridore tatsächlich haben. Bei Verwendung der kleinsten Mengen N , welche die ε -Bedingung erfüllen, ließen sich also (ε, ν) -Bäume mit kleineren als den nach Haussler und Welzl benötigten Konstanten finden.

Andererseits werden zahlreiche bei der Realisierung auftretende Konstanten bei dieser Untersuchung gar nicht berücksichtigt.

In diesem Abschnitt werden jedoch nur rudimentäre Überlegungen durchgeführt. Sie sind also mit einiger Vorsicht zu betrachten.

6.7.1 Abschätzung der Stichprobengröße

Sei eine endliche Punktmenge im \mathbb{R}^d gegeben. Zunächst ist die Dimension \tilde{d} des Bereichsraumes der $(d+1)$ -Korridore abzuschätzen. Es gilt $\tilde{d} \leq d(d+1) \log(d(d+1))$. Auch wenn die

tatsächliche Dimension deutlich kleiner sein, nehmen wir als Ausgangshypothese Gleichheit an. Dann gilt:

$$\tilde{d} = d(d+1) \log(d(d+1)) \geq d^2 \log(d^2) \geq 2d^2.$$

Wir verwenden für die Stichprobengröße $\nu = \nu(\varepsilon)$ die Abschätzung $\nu(\varepsilon) > \frac{8\tilde{d}}{\varepsilon}$ (und verzichten auf den logarithmischen Faktor). Es ist zu beobachten, dass für ein vorgelegtes $\gamma > 0$ ein $\varepsilon > 0$ so zu wählen ist, dass die Abschätzungen aus der Grenzwertbetrachtung im Theorem VI zur Existenz der Datenstruktur (S.93) erfüllt sind. Wir verzichten an dieser Stelle auf eine Berechnung. ε muss auch für nicht allzu kleine $\gamma > 0$ sehr klein gewählt werden. Wir nehmen nun beispielhaft $\varepsilon = \frac{1}{100}$ an.

Damit ergibt sich

$$\nu(\varepsilon) > 8 \cdot 2 \cdot 2^2 \cdot 100 = 6400 \text{ für } d = 2 \text{ oder}$$

$$\nu(\varepsilon) > 8 \cdot 2 \cdot 3^2 \cdot 100 = 144 \cdot 100 = 14400$$

für $d = 3$. Man beachte, dass bei tatsächlicher Berechnung des Wertes ε zu einem $\gamma > 0$ eine Abhängigkeit von der Dimension d des \mathbb{R}^d vorliegt.

6.7.2 Abschätzung der Zellen im Arrangement

Da bei einer Halbraum-Bereichsanfrage d -dimensionale Arrangements $\mathfrak{A}(N)$ mit Stichproben N , $|N| = \nu$ traversiert werden, ist zunächst eine Abschätzung der Anzahl der Zellen in einem solchen Arrangement zu bestimmen. Wir verwenden für die Anzahl der Zellen im d -dimensionalen Arrangement $\mathfrak{A}(N)$ mit $|N| = \nu$ den Wert

$$m := \Phi_d \left(\binom{\nu}{d} \right) \approx \nu^{d(d-1)}.$$

Auch wenn dieser Term eine Abschätzung nach oben ist, stimmt die „Größenordnung“ für große ν ungefähr. Mit der obigen Hypothese für ν erhält man nun für die Anzahl der Zellen im Arrangement:

$$m > 6400^2 \text{ für } d = 2, \quad m > 14400^6 \text{ für } d = 3 \text{ sowie } m > 14000^{10 \cdot 9} \text{ für } d = 10.$$

Die Anzahl der Zellen im Arrangement taucht sowohl beim Speicherbedarf als auch bei der Anfragezeit als Konstante auf. Obwohl auf nahezu alle Berechnungen verzichtet wurde und keine sauberen Abschätzungen durchgeführt wurden, wird an dieser Stelle klar, dass der Algorithmus nicht durchführbar ist. Bei einer genauen Untersuchung treten viele weitere zum Teil exponentielle Abhängigkeiten auf.

Kapitel 7

Simplex-Bereichsanfragen

Da ein Simplex σ^* nach Definition 4.0.5 (S.62) durch $(d+1)$ Halbräume beschrieben wird, lassen sich Simplex-Bereichsanfragen auf Halbraum-Bereichsanfragen zurückführen. Dort wurde auch der Rand σ eines Simplex σ^* definiert; es gilt

$$\sigma = \bigcap_{i=1}^{d+1} h_i$$

mit bestimmten Hyperbenen h_i . Wie sich zeigt, lassen sich auch Simplex-Bereichsanfragen bei geeigneter Untersuchung mit (ε, ν) -Bäumen derart durchführen, dass für beliebige Simplex der Bruchteil der Punkte in geschnittenen Zellen hinreichend klein ist.

7.1 Einleitung zur verwendeten Datenstruktur

Bei geeigneter Durchführung erhält man wegen $d+1 = \mathcal{O}(1)$ gleiche asymptotische Laufzeiten. Eine simple Hintereinanderausführung von Halbraum-Bereichsanfragen ist nicht möglich, da als Resultate nur Anzahlen bestimmt werden. Aus diesen lässt sich nicht die Anzahl der Punkte in einem Simplex kombinieren. Stattdessen wird also der (ε, ν) -Baum als Datenstruktur mit einer angepassten Parameterwahl betrachtet. Man beobachtet dabei, dass für Simplex die ε -Bedingung aus der Definition 6.2.1 (S. 86) des (ε, ν) -Baum nicht mehr erfüllt ist. Da das Simplex durch $(d+1)$ Halbräume beschrieben wird, gilt jedoch eine modifizierte ε -Bedingung mit Faktor $(d+1)$ auf der rechten Seite. Nun werden die Abschätzungen im Satz 6.2.7 (S.92) über die Anfragezeit untersucht. Man sieht leicht, dass im Grenzwertprozess die zusätzliche Konstante $(d+1)$ verschwindet. Es gibt damit wieder ein hinreichend kleines ε und die Laufzeit ist wie bei Halbraum-Bereichsanfragen $\mathcal{O}(n^{\beta+\gamma})$ mit beliebig kleinem γ .

Zunächst definieren wir nun die Zone für Simplex analog zu Hyperebenen: Sei σ^* ein Simplex im \mathbb{R}^d und σ der Rand von σ^* . Die Hyperebenen sind nicht notwendig verschieden. Sei weiter eine Menge N von Punkten gegeben und das Arrangement $\mathfrak{A}(N)$ betrachtet.

7.1.1 Definition ($Z(\sigma, N)$). Sei σ^* ein (offenes) Simplex mit dem Rand σ . Dann ist die Zone vom Rand des Simplex die Menge der Zellen in $\mathfrak{A}(N)$, welche den Rand σ schneiden:

$$Z(\sigma, N) := \{f \in F(N) : f \cap \sigma \neq \emptyset\}.$$

Es gibt $d + 1$ Hyperebenen h_1, \dots, h_{d+1} , durch die σ^* festgelegt ist. Da $\sigma \subset \bigcup_{i=1}^{d+1} h_i$, folgt mit der Definition 5.1.1 (S. 74) der Zone einer Hyperebene h sofort:

$$Z(\sigma, N) \subset \bigcup_{i=1}^{d+1} Z(h, N).$$

7.2 Analyse der Parameter des (ε, ν) -Baumes

Die Grundlage für die Verwendbarkeit eines (ε, ν) -Baumes für Simplex-Bereichsanfragen ist die Beobachtung, dass die Erfüllung der ε -Bedingung auch eine (schwächere) Aussage über die Anzahl der Punkte in der Zone eines Simplex liefert.

7.2.1 Lemma (über die Zone eines Simplex). *Sei ein (ε, ν) -Baum für eine endliche Menge \hat{A} gegeben. Dann gilt in jedem Knoten K die folgende Aussage für alle Hyperebenen h des \mathbb{R}^d :*

$$\sum_{f \in Z(\sigma, \mathfrak{A}(K))} |f \cap A_K| \leq (d + 1)\varepsilon \cdot |A_K|.$$

7.3 Der Anfragealgorithmus

Nun wird der Anfragealgorithmus für Simplex-Bereichsanfragen beschrieben. Er ist analog zum Anfragealgorithmus für Halbraum-Bereichsanfragen.

Die Überprüfung, ob f von σ geschnitten wird, ist dabei analog zu Halbraum-Bereichsanfragen bei Verwendung einer geeigneten Datenstruktur in insgesamt $\mathcal{O}(\nu^{d^2})$ Zeit möglich. Grund dafür ist, dass σ aus $\mathcal{O}(d) = \mathcal{O}(1)$ Hyperebenen besteht und damit ein einzelner Schnittpunkt von f gegen σ in $\mathcal{O}(1)$ Zeit möglich ist.

7.3.1 Satz (Simplex-Anfragezeit in (ε, ν) -Bäumen). *Sei eine n -elementige Menge $\hat{A} \subset \mathbb{R}^d$ und ein (ε, ν) -Baum für \hat{A} gegeben. Dann lassen sich mit diesem (ε, ν) -Baum Simplex-Bereichsanfragen beantworten mit der Anfragezeit*

$$T(n) = \mathcal{O}\left(\nu^{d^2} \cdot n^{\alpha(\varepsilon, \nu)}\right) \text{ mit } \alpha(\varepsilon, \nu) = 1 - \frac{1}{d(d-1) \cdot (\log_{1/\tilde{\varepsilon}} \nu) + 1}.$$

Dabei ist $\tilde{\varepsilon} = \varepsilon(d + 1)$.

1: **Funktion** SIMPLEX-TEILANFRAGE(K, σ^*)

Eingabe: Knoten K und Simplex σ^*

2: setze $\Sigma \leftarrow |N(K) \cap \sigma^*|$

3: **falls** K innerer Knoten ist , **dann** ▷ Sonst in Zeile 2 alles aufgezählt

4: **für alle** Zellen f in $\mathfrak{A}(K)$, die Punkte aus A_K enthalten :

5: **falls** f in σ^* enthalten ist , **dann** ▷ komplett enthalten

6: setze $\Sigma \leftarrow \Sigma + a(K_f)$ ▷ alle Punkte zählen

7: **sonst: falls** f von σ geschnitten wird , **dann** ▷ Redundanz, siehe (†)

8: rufe SIMPLEX-TEILANFRAGE rekursiv mit dem Kindknoten K_f auf

9: setze $\Sigma \leftarrow \Sigma + \text{SIMPLEX-TEILANFRAGE}(K_f, \sigma^*)$

Ausgabe: $\Sigma = |A_K \cap \sigma^*|$

10: **Prozedur** SIMPLEX-BEREICHSANFRAGE(h^*)

Eingabe: endliche Punktmenge \widehat{A} und Simplex σ^*

11: rufe SIMPLEX-TEILANFRAGE mit dem Wurzelknoten \widehat{K} auf

Ausgabe: $|A_{\widehat{K}} \cap \sigma^*| = |\widehat{A} \cap \sigma^*|$

Algorithmus 7.1: Aufzählung der Punkte eines Knotens K , die zu einem Simplex h^* gehören

Beweis (vgl. den entsprechenden Satz 6.2.7 (S. 92)). Man betrachte den Aufruf von SIMPLEX-TEILANFRAGE(K, σ^*) für einen inneren Knoten K , in dessen Region $|A_K|$ Punkte liegen. Nach dem Lemma 6.2.6 (S.91) über die rekursive Anfragezeit gilt:

$$T(K, \sigma^*) \leq \mathcal{O}(\nu^{d^2}) + \sum_{f \in Z(\sigma, \mathfrak{A}(K))} T(K_f, \sigma^*).$$

Seien n_1, \dots, n_m die Anzahlen der Punkte in den geschnittenen Regionen, die durch Aufrufe SIMPLEX-TEILANFRAGE(K_f, σ^*) bearbeitet werden. Dann gilt nach der ε -Bedingung:

$$\sum_{i=1}^m n_i \leq (d+1)\varepsilon|A_K| = \widetilde{\varepsilon}|A_K| \text{ mit } m \leq \nu^{d(d-1)}.$$

Damit genügt $T(n)$ den Voraussetzungen für das Abschätzungstheorem (Theorem X, S.119) mit $m \leq \nu^{d(d-1)}$ und einem geeignet großen $c = \mathcal{O}(\nu^{d^2})$. Dies gilt auch für den Aufruf von Blattknoten, da sich diese nach Lemma 6.2.6 (S. 91) in $\mathcal{O}(\nu)$ Zeit bearbeiten lassen. Es folgt also die Aussage. □

7.4 zur Existenz der Datenstruktur

Für Simplex-Bereichsanfragen existiert eine Datenstruktur mit asymptotisch zu Halbraumbereichsanfragen identischem Speicher und Laufzeit. Das Theorem ist analog zu Theorem VI.

Theorem VIII (Existenz der Datenstruktur für Simplexes) *Es gibt für alle $\gamma > 0$ eine $\mathcal{O}(n)$ -Datenstruktur für Simplex-Bereichsanfragen mit Anfragezeit $T(n) = \mathcal{O}(n^{\beta+\gamma})$, wobei $\beta = 1 - \frac{1}{d(d-1)+1}$ gilt.*

Beweis. Der Beweis verläuft analog zum Beweis von Theorem VI (S. 93). Es ist nur die Substitution $\tilde{\varepsilon} = \varepsilon(d+1)$ zu verwenden. Sei $\gamma > 0$. Nach dem Satz 6.3.1 zur Existenz von (ε, ν) -Bäumen gibt es für alle $0 < \varepsilon < 1/2$ einen (ε, ν) -Baum mit $\nu = c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$. Gezeigt wird, dass ein solcher (ε, ν) -Baum für hinreichend kleines ε eine geforderte Datenstruktur ist. Sei also $0 < \varepsilon < 1/2$ fest und $\nu = \nu(\varepsilon) := c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} = \mathcal{O}_n(1)$.

Aufgrund von Lemma 6.2.4 (S. 90) ist der Speicherbedarf des (ε, ν) -Baum $\mathcal{O}(\nu^{d^2} n) = \mathcal{O}(n)$. Nach Lemma 6.2.7 (S. 92) ist die Simplex-Anfragezeit $\mathcal{O}(\nu^{d^2} \cdot n^{\alpha(\tilde{\varepsilon}, \nu)}) = \mathcal{O}(n^{\alpha(\tilde{\varepsilon}, \nu)})$ mit

$$\alpha(\tilde{\varepsilon}, \nu) = 1 - \frac{1}{d(d-1) \log_{1/\tilde{\varepsilon}} \nu + 1}.$$

Den entscheidenden logarithmischen Term bezeichnen wir mit $\tilde{w}(\varepsilon) := \log_{1/\tilde{\varepsilon}} \nu(\varepsilon)$. Wir zeigen $\lim_{\varepsilon \rightarrow 0} \tilde{w}(\varepsilon) = 1$. Wir substituieren $\varepsilon = \frac{\tilde{\varepsilon}}{d+1}$ und zeigen die äquivalente Aussage $\lim_{\tilde{\varepsilon} \rightarrow 0} w(\varepsilon) = 1$:

$$\begin{aligned} \tilde{w}(\varepsilon) &= \log_{1/\tilde{\varepsilon}} \nu(\varepsilon) \\ &= \log_{1/\tilde{\varepsilon}} \left[c_d^* \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \right] \\ &= \log_{1/\tilde{\varepsilon}} \left[c_d^* \cdot \frac{d+1}{\tilde{\varepsilon}} \log \frac{d+1}{\tilde{\varepsilon}} \right] \\ &\geq \log_{1/\tilde{\varepsilon}} c_d^* \cdot \frac{d+1}{\tilde{\varepsilon}} \log \frac{d+1}{\tilde{\varepsilon}} \\ &= \log_{1/\tilde{\varepsilon}} c_d^* \cdot (d+1) \log(d+1) + \log_{1/\tilde{\varepsilon}} \frac{1}{\tilde{\varepsilon}} + \log_{1/\tilde{\varepsilon}} \log \frac{1}{\tilde{\varepsilon}} \\ &= \log_{1/\tilde{\varepsilon}} c_d^* \cdot (d+1) \log(d+1) + 1 + \log_{1/\tilde{\varepsilon}} \log \frac{1}{\tilde{\varepsilon}} \\ \implies \lim_{\tilde{\varepsilon} \rightarrow 0} w(\varepsilon) &= 1 + \lim_{\tilde{\varepsilon} \rightarrow 0} \log_{1/\tilde{\varepsilon}} c_d^* \cdot (d+1) \log(d+1) + \lim_{\tilde{\varepsilon} \rightarrow 0} \log_{1/\tilde{\varepsilon}} \log \frac{1}{\tilde{\varepsilon}} \\ &= 1 + 0 + 0 = 1. \end{aligned}$$

Es ist also $\lim_{\tilde{\varepsilon} \rightarrow 0} \alpha(\tilde{\varepsilon}, \nu) = 1 - \frac{1}{d(d-1)+1}$. Aus Stetigkeitsgründen kann nun $\tilde{\varepsilon}$ so gewählt werden, dass

$$\alpha(\varepsilon) < 1 - \frac{1}{d(d-1)+1} + \gamma.$$

Also lässt sich auch $\varepsilon = \frac{\tilde{\varepsilon}}{d+1} > 0$ geeignet wählen. Damit ist ein (ε, ν) -Baum mit einem solchen ε und $\nu = \nu(\varepsilon)$ eine Datenstruktur, für die gilt: Simplex-Bereichsanfragen werden in $T(n) = \mathcal{O}(n^{\beta+\gamma})$ Anfragezeit mit $\beta = 1 - \frac{1}{d(d-1)+1}$ durchgeführt. \square

Mit dem Beweis dieses Theorems wurde gezeigt, dass sich (ε, ν) -Bäume gleichermaßen für Simplex-Bereichsanfragen wie für Halbraum-Bereichsanfragen eignen. Aus Vollständigkeitsgründen formulieren wir auch ein entsprechendes Theorem zur Konstruktion.

7.5 zur Konstruktion der Datenstruktur

Für Simplexbereichsanfragen ist die erwartete Laufzeit für eine Datenstruktur asymptotisch gleich zu jener für Halbraumbereichsanfragen. Das Theorem ist analog zu Theorem VII.

Theorem IX (Konstruktion von (ε, ν) -Bäumen für Simplizes) *Sei ein $\gamma > 0$ sowie $A \subset \mathbb{R}^d$ vorgelegt. Dann lässt sich eine Datenstruktur mit linearem Speicher zur Beantwortung von Simplex-Bereichsanfragen mit Anfragezeit $\mathcal{O}(n^{\beta+\gamma})$, $\beta = 1 - \frac{1}{d(d-1)+1}$ konstruieren. Dies ist möglich durch einen Las-Vegas-Algorithmus mit $\mathcal{O}(n \log n)$ erwarteter Laufzeit.*

Beweis. Da die gleiche Datenstruktur wie für Halbraum-Bereichsanfragen verwendet wird, wurde die Aussage schon im Beweis von Theorem VII (S.96) zur Konstruktion der Datenstruktur für Halbraum-Bereichsanfragen gezeigt. \square

Da schon die Konstanten für die Datenstruktur zur Beantwortung von Halbraum-Bereichsanfragen exorbitant groß sind, sei für eine Betrachtung der auftretenden Konstanten auf den letzten Abschnitt des Kapitels zu Halbraum-Bereichsanfragen verwiesen.

Mit dem Ende dieses Kapitels ist das gesetzte Ziel erreicht:

Es wurde für jedes $\gamma > 0$ die Existenz und effiziente randomisierte Konstruierbarkeit einer $\mathcal{O}(n)$ -Datenstruktur angegeben, die für beliebige n -elementige Mengen Halbraum- und Simplex-Bereichsanfragen mit einer Laufzeit von $T(n) = \mathcal{O}(n^{\beta+\gamma})$ für $\beta = 1 - \frac{1}{d(d-1)+1}$ durchführt.

Kapitel 8

Ergänzungen

Wir beschäftigen uns in diesem Kapitel mit weiterführenden Aussagen zu ε -Netzen in Bereichsräumen mit einer endlichen Dimension d . (d steht hier nicht für die Dimension des \mathbb{R}^d .)

8.1 Größe und Konstruktion von ε -Netzen im Allgemeinen

Für (ε, ν) -Bäume und deren Konstruktion durch einen Las-Vegas-Algorithmus hat sich im Kapitel zu Halbraum-Bereichsanfragen gezeigt: Es ist ausreichend, für eine endliche Menge A die Existenz eines ε -Netzes N mit $\mathcal{O}\left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right)$ Elementen nachzuweisen und (nur) randomisiert zu konstruieren.

Die genaueste Abschätzung, die hier nachgewiesen wurde, ist nach Satz 3.5.13 (S.57):

$$|N| \leq \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon}. \quad (8.1)$$

Es sind zwei Fragen für endliche Bereichsräume von grundsätzlichem Interesse:

1. Welche Größenordnung bzw. welche Größe hat das kleinste ε -Netz?
2. Wie lassen sich ε -Netze deterministisch konstruieren?

Zur ersten Frage sei zunächst $f_d(\varepsilon)$ für $d \geq 1$ definiert als die kleinste natürliche Zahl f derart, dass für jeden endlichen(!) Bereichsraum der Dimension d ein ε -Netz mit f Elementen existiert (vgl. [HW87, S.137] bzw. [KPW92, Abstract]). Haussler und Welzl geben bereits eine Abschätzung nach oben und unten an [HW87, Theorem 3.8]:

$$\left\lfloor \frac{d}{2} \right\rfloor \frac{1}{\varepsilon} - 1 \leq f_d(\varepsilon) \leq \left\lceil \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon} \right\rceil.$$

Die Abschätzung nach oben ist in dieser Arbeit der Satz 3.5.13, die Abschätzung nach unten wurde nicht verwendet.

Matoušek macht im Artikel *Epsilon-Nets and Computational Geometry* [Mat93, Bemerkung vor Theorem 1.3] die folgenden Bemerkungen: Die obere Schranke $\frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon}$ sei von Blumer et al. [BEHW89] und anschließend von Komlos [KPW92] verbessert worden zu $(1 + o(1)) \frac{d}{\varepsilon} \log \frac{1}{\varepsilon}$. Der logarithmische Faktor komme scheinbar von der Beweistechnik. Jedoch hätten Pach und Wöginger durch ein Beispiel gezeigt, dass er im allgemeinen Fall notwendig sei. Größenordnungsmäßig ist die erste Frage also für den allgemeinen Fall beantwortet. Die folgende fast exakte Abschätzung nach oben und unten für $d \geq 2$ aus dem gemeinsamen Artikel von Komlos, Pach und Wöginger beinhaltet die besten bekannten Ergebnisse [KPW92]:

$$d - 2 + \frac{2}{d + 2} \leq \liminf_{\varepsilon \rightarrow 0} \frac{f_d(\varepsilon)}{(1/\varepsilon) \log(1/\varepsilon)} \leq \limsup_{\varepsilon \rightarrow 0} \frac{f_d(\varepsilon)}{(1/\varepsilon) \log(1/\varepsilon)} \leq d.$$

Hierbei werden der \liminf und der \limsup statt verwendet, da für den notierten Bruch die Existenz des \lim nicht nachgewiesen ist. Die Ungleichungskette ist eine Aussage über die asymptotischen Schranken sowie die in den Schranken auftretenden Konstanten. In jenem Artikel werden auch weitere Überlegungen zu speziellen Bereichsräumen angestellt.

Zur zweiten Frage sei auf Ausführungen bei Matoušek und Brönnimann verwiesen [Mat91, Brö95]. Insbesondere wurden zur deterministischen Konstruktion derandomisierte Algorithmen entwickelt.

ε -Netze für Rechtecke in Bezug auf Rechtecke

Wir betrachten hier statt Punkten achsenparallele Rechtecke als Objekte. Sei Q die Menge der achsenparallelen offenen Rechtecke im \mathbb{R}^2 . Sei $\varepsilon > 0$ und $S \subset Q$ eine Menge von n Rechtecken im \mathbb{R}^2 und R ein Bereich in Form eines beliebigen Rechtecks im \mathbb{R}^2 .

Wie groß ist die kleinste Teilmenge $N \subset S$, für die gilt: Jeder Bereich R , der mehr als εn Rechtecke aus S schneidet, schneidet auch ein Rechteck aus N ? Zwei Rechtecke schneiden sich hierbei per Definition genau dann, wenn sie (als Flächen) einen nicht-leeren Schnitt haben. Insbesondere schneiden sich Rechtecke also, wenn eines im anderen enthalten ist.

Eine solche Teilmenge N ist ein ε -Netz in einem Bereichsraum, der sich auch formal angeben lässt. Es ist der Bereichsraum (Q, \mathcal{R}) mit

$$\mathcal{R} = \bigcup_{p \in Q} \{q \in Q : q \cap p \neq \emptyset\}.$$

Die Frage, ob für endliche Unterräume von (Q, \mathcal{R}) ein ε -Netz der Größenordnung $\mathcal{O}(\frac{1}{\varepsilon})$ existiert, ist zur Zeit noch unbeantwortet. Haverkort [Hav06] gibt eine Zerlegung der Problemstellung an und löst eines der Teilprobleme. Dieses Teilproblem besteht darin, ein ε -Netz der Größenordnung $\mathcal{O}(\frac{1}{\varepsilon})$ für einen speziellen Bereichsraum (siehe weiter unten) zu finden. Diese Lösung wird im folgenden nachvollzogen. Da das gesuchte ε -Netz konstruktiv

angegeben wird, ergibt sich für diesen Bereichsraum also auch eine Antwort auf die Frage nach der deterministischen Konstruktion.

8.2 ε -Netze für horizontale Segmente in Bezug auf vertikale Segmente

Man betrachte den folgenden speziellen endlichen Bereichsraum (S, \mathcal{R}) :

S sei eine Menge von n horizontalen Segmenten im \mathbb{R}^2 . Die Bereiche seien vertikale Segmente des \mathbb{R}^2 . Als Teilmengen von S aufgefasst ist \mathcal{R} das System aller Teilmengen $R \subset X$, so dass für R gilt: Es gibt ein vertikales Segment des \mathbb{R}^2 , welches die horizontalen Segmente aus R alle schneidet und keines der horizontalen Segmente aus $X - R$ schneidet. Auch dieser Bereichsraum lässt sich formal angeben:

Sei W die Menge der horizontalen Segmente im \mathbb{R}^2 und V die Menge der vertikalen Segmente im \mathbb{R}^2 . Dann ist $(W, \tilde{\mathcal{R}})$ mit $\tilde{\mathcal{R}} := \bigcup_{v \in V} \{w \in W : w \cap v \neq \emptyset\}$. Dann ist (S, \mathcal{R}) der von S induzierte Unterraum von $(W, \tilde{\mathcal{R}})$, d.h. $(S, \mathcal{R}) = (S, \tilde{\mathcal{R}}|_S)$. Nach Haverkort [Hav06] gibt es in diesem Bereichsraum ein ε -Netz der Größenordnung $\mathcal{O}(\frac{1}{\varepsilon})$. Desweiteren lässt sich ein solches deterministisch konstruieren mit der Laufzeit $\mathcal{O}(n \log n)$.

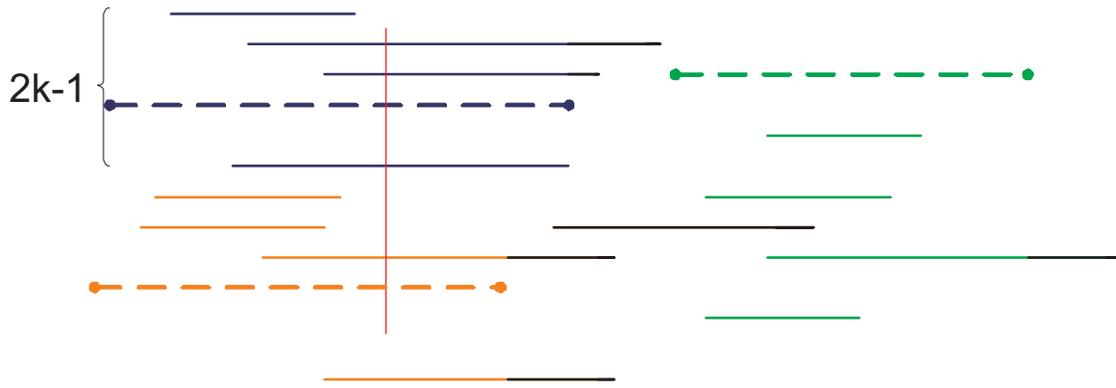
8.2.1 Satz (über den Schnitt horizontaler und vertikaler Segmente). *Sei S eine Menge horizontaler Segmente in der Ebene und $\varepsilon > 0$. Bezeichne jedes vertikale Segment mit mehr als εn Schnitten mit S als kritisch. Dann gibt es eine Teilmenge $N \subset S$ mit $|N| < 8/\varepsilon$, für die jedes kritische Segment ein Segment aus N schneidet. D.h. es gibt für den Bereichsraum (S, \mathcal{R}) ein ε -Netz N mit weniger als $8/\varepsilon = \mathcal{O}(\frac{1}{\varepsilon})$ Elementen.*

Für $8/\varepsilon \geq n$ ist nichts zu zeigen. Sei also $8/\varepsilon \leq n$. Im Folgenden wird eine solche Teilmenge N konstruktiv erzeugt.

Grob-Skizze der Konstruktion:

Die Segmente aus S sollen weniger als $8/\varepsilon$ Bereichen zugeteilt werden. Diese sind geeignet „flach“ zu wählen, damit jedes kritische Segment mindestens einen Bereich komplett überschneiden muss (vgl. Abb 8.2). Ein solcher Bereich wird zur anschaulichen Deutung als Stamm T (*tribe*) bezeichnet. Weiter soll aus jedem Stamm genau ein Element von S , welches die vollständige(!) horizontale Ausdehnung des Stammes hat, zur Menge N gehören. Damit würde gelten: Jedes kritische Segment (in Abb. 8.1 gestrichelt dargestellt) schneidet wie gefordert mindestens ein solches Element von N . Dabei tritt zunächst das folgende Problem auf:

Es gibt in der Regel Segmente mit weiterer Ausdehnung nach rechts (schwarz). Die Problembehandlung im folgenden Algorithmus verlangt, dass ein Segment zu verschiedenen Zeitpunkten verschiedenen Stämmen angehören kann. Interpretiert man die horizontale Achse als Zeit, so bezeichnet ein Stamm anschaulich also eine Zusammengehörigkeit von Segmenten über einen gewissen Zeitraum.



Gestrichelt: Segmente aus N

Problem: Es gibt in der Regel Segmente mit weiterer Ausdehnung nach hinten (schwarz).
 Problembehandlung durch Einführung von „Stämmen“ derart, dass ein Segment zu verschiedenen Zeitpunkten verschiedenen solchen Bereichen angehören kann

Abbildung 8.1: Grobskizze

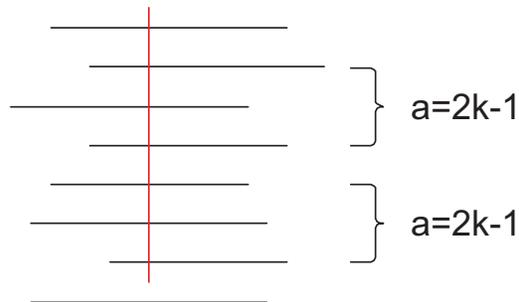


Abbildung 8.2: Beobachtung zu kritischen Segmenten

8.2.1 Ein Plane-Sweep-Algorithmus zur Konstruktion eines ε -Netzes

Der Algorithmus verwendet eine vertikale Sweepline von links nach rechts. Sei $k = \lfloor \varepsilon n / 4 \rfloor$.

Invarianten: Die folgenden Eigenschaften seien zu jedem Zeitpunkt erfüllt.

- (A) Jedes aktive Segment ist zu einem Zeitpunkt jeweils einem Stamm zugewiesen.
- (B) Zu jedem Stamm gehören mindestens eines und höchstens $2k - 1$ Segmente.
 Mit der Invariante (B) bezeichnen wir auch die unmittelbare Folgerung, dass jedes kritische Segment mindestens einen Stamm komplett überdeckt. (Dies folgt, da es mehr als $\varepsilon n \geq 4k - 4$ Segmente schneidet.)
- (C) In jedem Stamm sind mindestens ein und höchstens k Mitglieder als „Gründungsmitglieder“ markiert.
- (D) Die Stämme überlappen sich nicht, d.h. der Bereich zwischen zwei Segmenten eines Stammes enthält keine Segmente anderer Stämme.

Zudem gilt Eigenschaft **(E)**: Ein Segment kann nur bei Eintritt in einen neuen Stamm als Gründungsmitglied markiert werden; eine solche Markierung wird nie aufgehoben.

Initialisierung: $E = \{\}$.

Aus technischen Gründen wird zu S ein horizontales Beobachter-Segment hinzugefügt, dessen Anfang sich links unterhalb und dessen Ende sich rechts unterhalb aller anderen Segmente befindet. Zu Beginn existiert nur ein Stamm, dessen einziges Mitglied dieses Beobachter-Segment ist.

Ereignisbearbeitung für einen Segment-Anfang:

Füge das Segment dem Stamm zu, dem das direkt darunter liegende Segment aktuell angehört. Bei Überschreitung der Maximalgröße $2k - 1$ wird der Stamm in zwei Stämme aufgeteilt (siehe Split).

Ereignisbearbeitung für ein Segment-Ende:

Das Segment wird von dem zugehörigen Stamm entfernt. Falls es das letzte verbliebene als Gründungsmitglied markierte Element ist, wird dieses Segment in N aufgenommen. Die Invariante (C) ist dann verletzt. In diesem Fall werden alle Mitglieder des Stammes dem darunterliegenden Stamm zugeordnet und der alte Stamm aufgelöst. Falls dabei Invariante (B) verletzt wird, wird schließlich noch ein Split durchgeführt.

Ereignisbearbeitung für einen Split:

Teile den betroffenen Stamm T in die oberen $\lceil |T|/2 \rceil$ und die unteren $\lfloor |T|/2 \rfloor$ auf. Derjenige Teil, der weniger (oder zumindest nicht mehr) Gründungsmitglieder besitzt, werde zu einem neuen Stamm T' zusammengefasst. Der andere Teil wird nun als T bezeichnet und im Folgenden nicht verändert. Nach Invariante (E) bleiben die Gründungsmitglieder aus T auch in T' Gründungsmitglieder. Zudem werden auch weitere (beliebige) Mitglieder als Gründungsmitglieder markiert, so dass T' insgesamt genau k Gründungsmitglieder besitzt.

Korrektheit des Algorithmus:

Die Invarianten (A) bis (D) bleiben aufrechterhalten, wie im Detail leicht nachzuprüfen ist. Die Aufnahme des letzten (am weitesten nach rechts reichenden) Gründungsmitgliedes in N und anschließende Auflösung des Stammes garantieren zudem, dass vom linken bis zum rechten Ende des gesamten Stammes ein Segment aus N verläuft. Jedes kritische Segment schneidet aufgrund von Invariante (B) also einen Stamm und damit also wie gefordert ein in N liegendes Segment.

Das zur technischen Erleichterung der Konstruktion eingeführte Beobachter-Element muss zur Einhaltung der Invarianten zu N hinzugefügt werden. Es lässt sich allerdings zeigen, dass auch ohne dieses Element N ein ε -Netz bleibt: Nehmen wir an, ν sei ein kritisches Segment für S , welches das Beobachter-Element schneidet. Kürzt man dieses Segment so, dass dieses Beobachter-Element nicht mehr geschnitten wird, aber weiterhin alle übrigen Segmente aus S , so bleibt ν ein kritisches Segment. Nach der Korrektheit des Algorithmus ist N also ein ε -Netz.

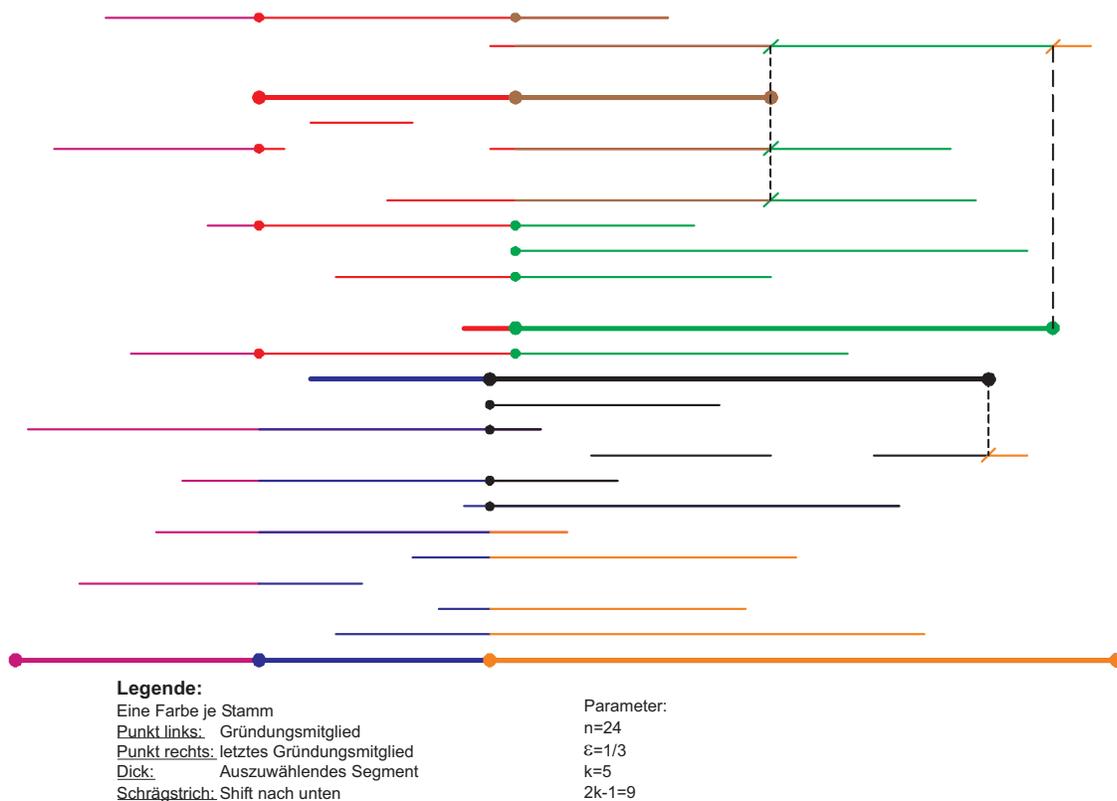


Abbildung 8.3: Abbildung zum Algorithmus

Beweis (von Satz 8.2.1, S.107).

Es ist noch zu zeigen, dass für die Größe des ε -Netzes N gilt: $|N| < 8/\varepsilon$.

Da zu jedem Stamm nur ein Segment aus N gehört und zu Beginn des Algorithmus nur der unechte Stamm existiert, entspricht die Größe von N der Anzahl der Split-Events. Es bleibt zu zeigen, dass dies weniger als $8/\varepsilon$ sind.

Bei einem Split-Event sind aus dem Teil, der in einen neuen Stamm aufgenommen wird, höchstens die Hälfte schon als Gründungsmitglieder markiert, d.h. es werden mindestens $k/2$ Markierungen vorgenommen. Die Abschätzung gegen $8/\varepsilon$ liefert nun eine einfache amortisierte Analyse: Insgesamt kann höchstens n Mal ein gewisses Segment markiert werden. Da Markierungen nie aufgehoben werden und bei jedem Split-Event zumindest $k/2$ Markierungen vorgenommen werden, gibt es höchstens $n/(k/2) < 8/\varepsilon$ Split-Events. Damit wurde der Satz bewiesen:

Es gibt eine Menge $N \subset S$ aus weniger als $8/\varepsilon$ (horizontalen) Segmenten, für die jedes kritische (vertikale) Segment ein Segment aus N schneidet. \square

8.2.2 Deterministische Konstruktion

Der folgende Satz bietet ein Resultat für die deterministische Konstruktion von N .

8.2.2 Satz (zur deterministischen Konstruktion). *Der zuvor beschriebene Algorithmus hat bei geeigneter Realisierung die Laufzeit $\mathcal{O}(n \log n)$.*

Zum Beweis wird eine solche Realisierung beschrieben und anschließend untersucht.

8.2.3 Realisierung des Konstruktionsalgorithmus

Wir bezeichnen die Koordinatenachsen der Ebene \mathbb{R}^2 als x -Achse und y -Achse.

Zunächst wird die bei einem Plain-Sweep-Algorithmus verwendete y -TABLE beschrieben.

y -TABLE:

Diese ist ein höhen-balancierter binärer Suchbaum für die Stämme in y -sortierter Reihenfolge. Ein solcher Baum wird als AVL-Baum bezeichnet und ermöglicht das Suchen, Einfügen und Löschen von Elementen in $\mathcal{O}(\log n)$ Zeit. Diese Ausführungen finden sich beispielsweise in dem Lehrbuch von Ottmann und Widmayer [OW02].

Jeder Stamm T hat in seiner Repräsentation in der y -Table einen KEY y und einen eigentlichen Inhalt. Auch der Inhalt eines Eintrags wird im Weiteren einfach mit T bezeichnet.

KEY y maximale y -Koordinate der Segmente

ELEMENT T inhomogener AVL-Baum

Es wird ein inhomogener Baum verwendet, damit das Einfügen eines Knotens v bei einem Split-Ereignis möglich ist. Desweiteren wird der AVL-Baum um die Funktionalität einer CONCATENABLE QUEUE für die Segmente in y -sortierter Reihenfolge erweitert. Diese Erweiterung wird im Skript zur Vorlesung *Algorithmische Geometrie II (WWU Münster, SS05)* erläutert [Vah05, Kapitel 12.1]. Bei dem in diesem Abschnitt entworfenen Algorithmus wird die Funktionalität zusätzlich um gewisse Marker ergänzt, wie weiter unten erläutert wird.

Die Attribute in einem inneren Knotens eines AVL-Baumes T sind die folgenden Marker:

n_l Anzahl der Segmente im linken Teilbaum

n_r Anzahl der Segmente im rechten Teilbaum

f_l Anzahl der Gründungsmitglieder im linken Teilbaum

f_r Anzahl der Gründungsmitglieder im rechten Teilbaum

y_{\max} maximale y -Koordinate im Teilbaum

Das Attribut eines Blattknotens ist ein Segment $s = (x_1, x_2, y_1)$. (Die Information, ob s ein Gründungsmitglied ist, erhält man implizit durch den Elterknoten.)

Realisierung des AVL-Baums:

Der AVL-Baum ist so zu realisieren, dass bei den Operationen INSERT, DELETE, SPLIT und SPLICE die fünf Marker sowie der KEY y stets aktualisiert werden. Der KEY y ist genau dann zu aktualisieren, wenn eine Aktualisierung von y_{\max} in der Wurzel erfolgt, daher wird im Folgenden die Aktualisierung des KEYS nicht explizit erwähnt, sie erfolgt implizit. Offensichtlich ist eine Aktualisierung der Marker bei einer lokalen Rebalancierung in $\mathcal{O}(1)$

möglich (die Marker der Elemente oberhalb der Rebalancierung sind nicht betroffen) und kann daher stets durchgeführt werden.

Realisierung der Operation `INSERT(s)`:

- 1: füge das Segment wie in einem üblichen AVL-Baum ein, jedoch noch ohne Rebalancierung
- 2: aktualisiere die Marker (beginnend im Knoten v , der s repräsentiert, auf dem gesamten Pfad bis zur Wurzel)
- 3: rebalanciere

Realisierung der Operation `DELETE(s)`:

- 1: lösche das Segment wie in einem üblichen AVL-Baum ein, jedoch noch ohne Rebalancierung)
- 2: aktualisiere die Marker (beginnend im entfernten Knoten v auf dem gesamten Pfad von v bis zur Wurzel)
- 3: rebalanciere

Realisierung der Operation `SPLICE` wie bei Vahrenhold [Vah05, Kapitel 12.1]

Die Realisierung wird hier um die Aktualisierung der Marker ergänzt.

Als Voraussetzung sind zwei Stämme T_1 und T_2 gegeben, wobei alle Elemente aus T_1 kleiner als Elemente aus T_2 sind. O. B. d. A. gelte für die Höhen der Bäume $h_1 \leq h_2$.

- 1: betrachte den Pfad Π in T_1 zum am weitesten rechts liegenden Blatt
- 2: sei v der höchste Knoten auf Π , der Wurzel eines Teilbaums T' der ist
- 3: entferne T' aus T_1 und füge stattdessen den Knoten v' mit den Kindknoten der Wurzel von T_2 ein.
- 4: aktualisiere beginnend mit v' die fünf Marker bis hin zur Wurzel
- 5: rebalanciere T_1 von v' bis zur Wurzel

Realisierung der Operation `SPLIT`:

Wie bei Vahrenhold. Die Aktualisierungen erfolgt stets in den hier aufgerufenen Methoden, daher ist keine Anpassung erforderlich.

Pre-Processing:

Sortiere die Punkte der Segmente bezüglich der x -Koordinaten zur Bearbeitung der Ereignisse Segmentanfang sowie Segmentende in richtiger Reihenfolge.

STARTEVENT(s): Segmentanfang mit einem Segment $s = (x_1, x_2, y_1)$:

- 1: suche in der y -TABLE nach y_1 und erhalte den Stamm T , in den das Segment einzufügen ist

- 2: $T.\text{INSERT}(s)$, d.h. füge das Segment s in den AVL-Baum T ein
- 3: aktualisiere im neuen Element $T.s$ die vier Marker
- 4: aktualisiere den KEY y , falls $y < y_{min}$
- 5: verfolge beginnend mit $T.s$ den Pfad im AVL-Baum bis zur Wurzel und aktualisiere auch hier jeweils die vier Marker
- 6: IF($n_l + n_r \geq 2k$) SPLITEVENT(T)

ENDEVENT(s): Segmentende mit einem Segment $s = (x_1, x_2, y_1)$:

- 1: suche in der y -TABLE nach y_1 und erhalte den Stamm T , in dem das Segment aufzufinden ist
- 2: $T.\text{FIND}(y_1)$ // suche das Segment s im AVL-Baum T
- 3: verfolge mit diesem Element $T.s$ beginnend den Pfad im AVL-Baum bis zur Wurzel und bringe jeweils die vier Marker auf den Zustand, der nach dem Entfernen des Segmentes s korrekt ist
- 4: $T.\text{DELETE}(s)$
- 5: IF($f_l + f_r = 0$) SHIFTEVENT(T)

SPLITEVENT(T): Split-Ereignis mit einem Stamm T :

- 1: $m := (n_l + n_r)/2$ (mit n_l und n_r aus der Wurzel) // das m -te Segment von T ist das mediane Element
- 2: $median := \text{FIND}(m)$ // Finde das mediane Segment durch die Hilfsmethode $\text{find}(m)$
- 3: $T.\text{SPLIT}(median)$ /* führe die Operation Split im AVL-Baum $T.s$ in Bezug auf das mediane Segment aus. Die Elemente kleiner oder gleich dem medianen Segment seien dabei im Stamm T ; die anderen Elemente in einem Stamm T_2 */
- 4: Füge den zusätzlichen Stamm T_2 in die y -TABLE ein
- 5: Bestimme $f_l + f_r$ für die Wurzeln aus T sowie T_2 und lege in dem Stamm mit kleinerer Summe beliebige Segmente als Gründungsmitglieder fest

SHIFTEVENT(T): Shift-Ereignis mit einem Stamm T :

- 1: $T_2 :=$ der direkt unter T liegende Stamm
- 2: $T_2.\text{SPLICE}(T)$, d.h. verschmelze den Stamm T_2 mit dem Stamm T
- 3: Lösche den aufzulösenden Stamm T aus der y -TABLE
- 4: IF $n_l + n_r \geq 2k$ (SPLITEVENT(T_2))

Hilfsmethode $T.\text{FIND}(m)$:

- 1: IF($n_l \geq m$) $T.\text{rightSon}.\text{FIND}(m)$ wende rekursiv $\text{find}(m)$ auf den linken Teilbaum an (das mediane Segment befindet sich im linken Teilbaum)
- 2: ELSE($T.\text{rightSon}.\text{FIND}(m - n_l)$) wende rekursiv $\text{FIND}(m - n_l)$ auf den rechten Teilbaum

an (das mediane Segment befindet sich im rechten Teilbaum und ist dort an der Position m abzüglich der n_l Elemente im linken Teilbaum)

8.2.4 Laufzeit des Algorithmus

Wir beweisen hier, dass die Laufzeit der obigen Realisierung $\mathcal{O}(n \log n)$ beträgt.

8.2.3 Bemerkung. Wegen $8/\varepsilon \leq n$ ist $\mathcal{O}(1/\varepsilon) \subset \mathcal{O}(n)$. Man beachte, dass dagegen Laufzeiten wie $\mathcal{O}(n/\varepsilon)$ oder $\mathcal{O}(1/\varepsilon^2)$ nicht in $\mathcal{O}(n)$ sind.

Die Anzahl der Elemente eines AVL-Baumes aus der y -TABLE ist stets $\mathcal{O}(n)$, die Anzahl der Elemente der y -TABLE stets $\mathcal{O}(1/\varepsilon) \subset \mathcal{O}(n)$.

Zunächst machen wir die folgende Behauptung:

8.2.4 Lemma. *Die Operationen FIND, INSERT, DELETE, SPLICE und SPLIT im AVL-Baum haben Laufzeit $\mathcal{O}(\log n)$.*

Beweis. Die Anzahl der Elemente in einem AVL-Baum ist $\mathcal{O}(n)$, die Höhe des Baumes ist also $\mathcal{O}(\log n)$.

Nach dem Vorlesungsskript von Vahrenhold lassen sich damit die Operationen ohne Aktualisierung der Marker jeweils in $\mathcal{O}(\log n)$ durchführen [Vah05, Kapitel 12.1]. Die Aktualisierung der Marker ist in den Operationen INSERT, DELETE und SPLICE höchstens einmalig erforderlich und erfolgt vom betroffenen Element bis zur Wurzel, damit ist auch dies in $\mathcal{O}(\log n)$ Zeit möglich. In den Operationen FIND und SPLIT sind keine zusätzlichen Aktualisierungen notwendig.

Damit ist die Behauptung nachgewiesen. □

Mit Hilfe des Lemmas lässt sich nun der Satz zur deterministischen Konstruktion zeigen.

Beweis (von Satz 8.2.2, S.110). Wir führen den Beweis durch Untersuchung der einzelnen Teile des Algorithmus.

Laufzeit des Pre-Processing:

Das Sortieren der $2n$ x -Koordinaten der Segmente erfolgt in $\mathcal{O}(n \log n)$ Laufzeit.

Laufzeit für einen Segmentanfang:

Das Suchen des zugehörigen Stammes erfolgt in $\mathcal{O}(\log n)$, das Einfügen in den zugehörigen Stamm in $\mathcal{O}(\log n)$. Die Aktualisierung der Marker entlang des Pfades bis zur Wurzel ist in $\mathcal{O}(\log n)$ Laufzeit möglich. Die Überprüfung in der Wurzel, ob ein SPLIT notwendig ist, benötigt $\mathcal{O}(1)$ Laufzeit. Die Laufzeit für einen Segmentanfang ist also $\mathcal{O}(\log n)$. Da insgesamt $\mathcal{O}(n)$ Segmentanfänge auftreten, tragen diese Operationen nicht mehr als $\mathcal{O}(n \log n)$ zur Gesamtlaufzeit bei.

Laufzeit für ein Segmentende:

Das Suchen des zugehörigen Stammes erfolgt in $\mathcal{O}(\log n)$, das Einfügen in den zugehörigen Stamm in $\mathcal{O}(\log n)$. Die Aktualisierung der Marker entlang des Pfades bis zur Wurzel ist in $\mathcal{O}(\log n)$ Laufzeit möglich. Die Überprüfung in der Wurzel, ob ein SHIFT notwendig ist, benötigt $\mathcal{O}(1)$ Laufzeit. Die Laufzeit für einen Segmentanfang ist also $\mathcal{O}(\log n)$. Da insgesamt $\mathcal{O}(n)$ Segmentanfänge auftreten, tragen diese Operationen nicht mehr als $\mathcal{O}(n \log n)$ zur Gesamtlaufzeit bei.

SHIFT:

Die Durchführung der SPLICE-Operation benötigt $\mathcal{O}(\log n)$ Laufzeit; die Überprüfung, ob ein SPLIT ausgelöst werden muss, erfolgt in $\mathcal{O}(1)$ Zeit.

Insgesamt treten $\mathcal{O}(n)$ Segmentenden auf, d.h. die SHIFT-Operationen tragen insgesamt nicht mehr als $\mathcal{O}(n \log n)$ zur Gesamtlaufzeit bei.

Laufzeit für einen SPLIT:

Das Finden des medianen Elements und die SPLIT-Operation entlang des Median erfolgen in $\mathcal{O}(\log n)$ Laufzeit; das Dekrementieren von Markern in $\mathcal{O}(1)$.

Nach der Korrektheit des Algorithmus gibt es höchstens $8/\varepsilon \leq n = \mathcal{O}(n)$ SPLIT-Operationen. Damit tragen auch die SPLIT-Operationen insgesamt nicht mehr als $\mathcal{O}(n \log n)$ zur Gesamtlaufzeit bei.

Kein Teil des Algorithmus trägt mehr als $\mathcal{O}(n \log n)$ zur Laufzeit bei, damit ist die Laufzeit dieser Realisierung des Plain-Sweep-Algorithmus also wie behauptet $\mathcal{O}(n \log n)$. \square

Es wurde in diesem Kapitel gezeigt, dass ein ε -Netz der Größenordnung $\mathcal{O}\left(\frac{1}{\varepsilon}\right)$ für den speziellen Bereichsraum von horizontalen Segmenten in Bezug auf vertikale Segmente existiert und deterministisch in $\mathcal{O}(n \log n)$ Laufzeit konstruiert werden kann. Die Fragestellung, ob eine analoge Aussage auch im Bereichsraum der Rechtecke in Bezug auf Rechtecke gilt, ist bisher unbeantwortet.

Kapitel 9

Zusammenfassung und Ausblick

Wir haben einen Beweis von Haussler und Welzl [HW87] ausgearbeitet, nach dem zu einem vorgegebenen $\gamma > 0$ eine $\mathcal{O}(n)$ -Datenstruktur existiert, die Halbraum- und Simplex-Bereichsanfragen mit einer Laufzeit von $\mathcal{O}(n^\alpha)$ beantwortet, wobei $\alpha = 1 - \frac{1}{d(d-1)+1} + \gamma$ ist. Mit Hilfe eines Stichprobentheorems und der Methode der bedingten Wahrscheinlichkeiten wurde die Existenz und randomisierte Konstruierbarkeit von ε -Netzen nachgewiesen. Das von Haussler und Welzl eingeführte Konzept der ε -Netze wurde zunächst in allgemeinen Bereichsräumen eingeführt. Anschließend wurden geometrische Bereichsräume untersucht und mit Hilfe probabilistischer, kombinatorischer und geometrischer Betrachtungen das zentrale ε -Theorem formuliert. Dieses macht mit Hilfe von ε -Netzen eine Aussage über die Existenz geeignet kleiner Stichproben N für endliche Punktmenge A . Betrachtet wurde das Arrangement $\mathfrak{A}(N)$. Nach dem zentralen ε -Theorem ist für beliebige Hyperebenen h der Bruchteil der Punkte aus A in der Zone $Z(h, N)$ höchstens ε . Mit dieser Aussage konnte ein Partitionsschema für endliche Punktmenge \hat{A} entworfen werden, das Teilmengen A rekursiv in Zellen aus dem Arrangement $\mathfrak{A}(N)$ geeigneter Stichproben N partitioniert. Das Partitionsschema wurde realisiert durch einen als (ε, ν) -Baum bezeichneten Partitionsbaum. Für die Anfragezeit mit (ε, ν) -Bäumen ließ sich eine rekursive Laufzeit herleiten. Die aus recht großen Stichproben aufgebauten d -dimensionalen Arrangements weisen eine enorme Komplexität auf. Dennoch ergibt sich bei Verwendung einer geeigneten Stichprobengröße und genauer Analyse der Rekursion (siehe Anhang A) eine sublineare Laufzeit für Halbraum-Bereichsanfragen. Es wurde desweiteren eine randomisierte Konstruktion in $\mathcal{O}(n \log n)$ erwarteter Laufzeit durch einen Las-Vegas-Algorithmus angegeben. Die Aussagen über Simplex-Bereichsanfragen ergaben sich bei Verwendung derselben Datenstruktur mit abgeänderten Konstanten. Eine Übersicht zur Beweisstruktur findet sich in Anhang B auf Seite 131.

Da die Konstanten so groß sind, dass eine praktische Realisierung ausgeschlossen ist, sollte man schlicht von einer oberen Schranke für die Anfragezeit sprechen. Wie schon zuvor bemerkt, hat Matoušek im Artikel *Efficient Partition Trees* [Mat92] die obere Schranke zu

$\mathcal{O}\left(n^{1-1/d} \log^{\mathcal{O}(1)} n\right)$ verbessert. Damit ist die obere Schranke bis auf einen polylogarithmischen Faktor an die untere Schranke für die Anfragezeit mit linearem Speicher angenähert. Offen bleibt jedoch die Frage, ob sich (möglicherweise mit ε -Netzen oder ähnlichen Konzepten) auch praktikable Datenstrukturen entwerfen lassen, die sich in der Nähe der unteren Schranke befinden. Eine weitere Frage ist, wie sich ε -Netze und verwandte Mengensysteme wie ε -Approximationen deterministisch konstruieren lassen. In Bezug auf diese Frage zeigt Brönnimann [Brö95] in seiner Doktorarbeit auf, wie sich randomisierte Konstruktionen *derandomisieren* lassen. Er bezeichnet das hier betrachtete allgemeine Konzept von Bereichsräumen und dessen VC-Dimension als „powerful tool which has been sharpened by the work of many researchers“ [Brö95, S.137] und nennt dabei zuerst den in dieser Arbeit behandelten Artikel von Haussler und Welzl sowie diverse Artikel von Matoušek et al. Brönnimann betrachtet das den ε -Netzen verwandte Mengensystem der sogenannten *Semicuttings*. Mit Hilfe von *Semicuttings* geeigneter Größe gelingt es ihm durch eine Derandomisierung, einen in jeder Dimension d optimalen deterministischen Algorithmus zur Bestimmung der konvexen Hülle einer Punktmenge im \mathbb{R}^d anzugeben. Die Größe von ε -Netzen und verwandten Mengensystemen in verschiedenen Typen von Bereichsräumen zu bestimmen, ist daher weiterhin ein relevanter Themenbereich.

Desweiteren ist der Versuch, durch ε -Netze und ähnliche Mengensysteme Verbesserungen für weitere geometrische Algorithmen zu finden, eine vielversprechende aktuelle Problemstellung.

Anhang A

Abschätzung der Rekursion

Die Abschätzung der Rekursionsgleichung für Satz 6.2.7 (S.92) wird bei Haussler und Welzl im Anhang bewiesen [HW87, Anhang]. In dieser Diplomarbeit wird die Abschätzung als das Abschätzungstheorem bezeichnet. Der Beweis wird hier in Sätze und Lemmata gegliedert und in einigen Details ausführlicher durchgeführt.

Im ersten Abschnitt wird eine rekursive Funktion $Q(n)$ mit Parametern c , ε und m definiert und das Abschätzungstheorem (Theorem X) für $Q(n)$ formuliert. Am Ende des Abschnitts findet sich eine Bemerkung zu den Größenordnungen in der Verwendung des Theorems in Satz 6.2.7.

Der zweite Abschnitt beginnt dann mit einer Beweisskizze, aus der sich die Unterteilung in den dritten und vierten Abschnitt ergibt.

Im fünften Abschnitt wird das Abschätzungstheorem schließlich bewiesen.

A.1 Einleitung: Formulierung des Abschätzungstheorems

Wir beginnen mit der Definition von $Q(n)$:

A.1.1 Definition ($Q(n)$). Sei $c > 0$, $0 < \varepsilon < 1$ und $m \geq 2$. $Q : \mathbb{N} \rightarrow \mathbb{N}$ ist eine rekursiv definierte Funktion mit $Q(0) := 0$ und

$$Q(n) := c + \max_{\bar{n} \in \mathcal{N}(n)} \sum_{i=1}^m Q(n_i).$$

Dabei ist $\bar{n} = (n_1, \dots, n_m)$ ein m -Tupel und $\mathcal{N}(n)$ die Menge der m -Tupel mit Werten aus \mathbb{N} , für welche die Summe der Elemente kleiner oder gleich εn ist:

$$\mathcal{N}(n) := \left\{ \bar{n} \in \mathbb{N}^m : \sum_{i=1}^m n_i \leq \varepsilon n = \frac{n}{r} \right\}.$$

$Q(n)$ ist wohldefiniert, da für $\bar{n} \in \mathcal{N}(n)$ und $i \in \{1, \dots, m\}$ wegen $\varepsilon n < n$ stets $n_i < n$ gilt.

A.1.2 Bemerkung. Es wird statt ε auch die Bezeichnung $r := 1/\varepsilon$ mit $1 < r \in \mathbb{R}$ verwendet, wenn dies eine angenehmere Darstellung der Terme ermöglicht. Wenn ε eine sehr kleine vorgelegte Zahl ist, hat man entsprechend eine sehr große vorgelegte Zahl r . Für die m -Tupel $\bar{n} \in \mathcal{N}(n)$ lässt sich bei der Bestimmung des Maximums wegen der Symmetrie in der Summation o.B.d.A. $n_1 \leq n_2 \leq \dots \leq n_m$ verlangen. Man beachte auch, dass $Q(n)$ monoton ist, da $\bar{n} \in \mathcal{N}(n) \implies \bar{n} \in \mathcal{N}(n+1)$. Desweiteren ist $Q(n)$ „monoton in m “, da ein m -Tupel \bar{z} durch Nullen zu einem $m' > m$ -Tupel ergänzt werden kann.

Das Theorem, nach welchem sich die Funktion $Q(n)$ explizit abschätzen lässt, lautet wie folgt:

Theorem X (Abschätzungstheorem: $Q(n) \in \mathcal{O}(c \cdot n^\alpha)$)

$$Q(n) \leq 4cn^\alpha \text{ mit } \alpha := \frac{\log_{1/\varepsilon} m}{1 + \log_{1/\varepsilon} m} = 1 - \frac{1}{1 + \log_{1/\varepsilon} m}.$$

Dieses Theorem besagt, dass $Q(n) \in \mathcal{O}(c \cdot n^\alpha)$ ist. Der Beweis findet sich am Ende des Kapitels.

A.1.3 Bemerkung (Größenordnungen). In der Verwendung des Theorems im Satz 6.2.7 (S.92) hat man die Ungleichung $m \leq \nu^{d(d-1)}$ und $\nu \leq \tilde{c} \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$ mit einer großen Konstanten \tilde{c} . Es ist also m deutlich größer als $1/\varepsilon$ und sowohl $1/\varepsilon$ als auch m sind sehr große Zahlen. $1/\varepsilon$ ist dabei eine gegebene Konstante, während m in der Anwendung ein Parameter ist, der nicht zu stark exponentiell wachsen darf, damit α nicht zu groß ist.

A.2 Untersuchung von $Q(n)$

Bevor wir $Q(n)$ untersuchen, folgt hier eine prosaische Beweisskizze sowie die Definition einer natürlichen Zahl n^* als „Grenze“, die in mehreren Fallunterscheidungen auftritt.

A.2.1 Bemerkung (prosaische Beweisskizze und Definition der „Grenze“ n^*).

Es wird eine geeignetere Funktion $\gamma(n)$ definiert, die „um den Faktor c gestaucht“ und unterhalb einer „Grenze“ linear ist. Wie sich zeigt, ist der Bereich unterhalb der Grenze $n^* := \lceil mr \rceil = \lceil m/\varepsilon \rceil$ in einigen Lemmata unproblematischer. Man kann $Q(n)$ erst gegen $\gamma(n)$ und anschließend gegen eine geeignete, auf positiven reellen Zahlen definierte, konkave Funktion $f(x)$ abschätzen.

A.2.2 Lemma (Steigung von $Q(n)$).

$$Q(n+1) \leq c + Q(n) \text{ für alle } n.$$

Beweis (mit vollständiger Induktion nach n).

Induktionsanfang: $Q(0) = 0$ und $Q(1) = c$ erfüllen die Behauptung. Sei nun n fest.

Induktionsvoraussetzung: $Q(k) \leq c + Q(k-1)$ für alle $1 \leq k \leq n$.

Induktionsbehauptung: $Q(n+1) \leq c + Q(n)$

Beweis: Wähle $\bar{z} \in \mathcal{N}(n+1)$ so, dass gilt:

$$\max_{\bar{n} \in \mathcal{N}(n+1)} \sum_{i=1}^m Q(n_i) = \sum_{i=1}^m Q(z_i).$$

Annahme: $\bar{z} = (0, \dots, 0)$. Dann gilt $Q(n+1) = c \leq Q(n) + c$.

Sei nun also $\bar{z} \neq (0, \dots, 0)$. Da $Q(n)$ symmetrisch bezüglich der n_i definiert ist, lässt sich o.B.d.A. $z_1 \neq 0$ verlangen. Mit der Definition von $Q(n+1)$ erhält man durch Einsetzen der Induktionsvoraussetzung für $Q(z_1)$:

$$\begin{aligned} Q(n+1) &= c + \sum_{i=1}^m Q(z_i) \\ &\leq c + Q(z_1 - 1) + c + \sum_{i=2}^m Q(z_i) \\ &\leq c + (c + \max_{\bar{n} \in \mathcal{N}(n)} \sum_{i=1}^m Q(n_i)) \\ &= c + Q(n). \end{aligned}$$

In der vorletzten Umformung wurde dabei die Definition von $\mathcal{N}(n)$ verwendet und aufgezeigt, dass ein darin enthaltenes m -Tupel \bar{z}' vorliegt:

$$\begin{aligned} \bar{z} &= (z_1, z_2, \dots, z_m) \in \mathcal{N}(n+1) \\ \implies \sum_{i=1}^m z_i \leq \varepsilon(n+1) &\implies (z_1 - 1) + z_2 + \dots + z_m \leq \varepsilon(n+1) - 1 \leq \varepsilon n \\ \implies \bar{z}' := (z_1 - 1, z_2, \dots, z_m) &\in \mathcal{N}(n). \end{aligned}$$

Die letzte Umformung zu $c + Q(n)$ folgt aus der Definition von $Q(n)$. □

A.2.3 Lemma (Wert von $Q(n)$ für $n < (m+1)r$).

$$Q(n) = c \cdot (\lfloor \varepsilon n \rfloor) \text{ für } n < (m+1)r.$$

Beweis. Sei $n < (m+1)r$. Aus Lemma A.2.2 „ $Q(n+1) \leq Q(n) + 1$ “ folgt für alle $k \in \mathbb{N}$:

$$Q(k+1) + Q(0) \leq c + Q(k) + 0 = Q(k) + Q(1).$$

Diese Eigenschaft beschreibt gewissermaßen eine Konkavität der Funktion $Q(n)$. Sie lässt sich ausnutzen, um ein m -Tupel schrittweise zu reduzieren.

Sei $\bar{z} \in \mathcal{N}(n)$ so, dass \bar{z} die maximale Summe ergibt:

$$c + \sum_{i=1}^m Q(z_i) = c + \max_{\bar{n} \in \mathcal{N}(n)} \sum_{i=1}^m Q(n_i) = Q(n).$$

Nach der Definition von $\mathcal{N}(n)$ gilt $\sum_{i=1}^m z_i \leq \lfloor \varepsilon n \rfloor$. Für die Elemente von \bar{z} kommen die folgenden beiden Fälle in Frage:

I. Fall: $z_i \leq 1$ für alle i . Es ist $Q(0) = 0$ und $Q(1) = c > Q(0)$. Nach der Wahl von \bar{z} haben maximal viele z_i den Wert 1. Da $\sum_{i=1}^m z_i \leq \lfloor \varepsilon n \rfloor$, ist diese maximale Anzahl von $z_i = 1$ gegeben durch $\lfloor \varepsilon n \rfloor$. Es folgt:

$$Q(n) = c + \sum_{i=1}^m Q(z_i) = c + \lfloor \varepsilon n \rfloor Q(1) + k \cdot 0 = c(1 + \lfloor \varepsilon n \rfloor).$$

II. Fall: Es gibt ein z_i mit $z_i \geq 2$.

Annahme: Es gibt kein z_j mit $z_j = 0$. Dann gilt: $\sum_{i=1}^m z_i \geq (m-1) \cdot 1 + 1 \cdot 2 = m+1 > \varepsilon n \frac{1}{2}$. Sei also nun $z_j = 0$. o. B. d. A. setzen wir $j = 1$ und $i = 2$, d.h. $z_1 = 0$ und $z_2 \geq 2 \Leftrightarrow z_2 - 1 > 0$. Dieser Fall lässt sich nun reduzieren durch

$$\bar{z}' := (z_1 + 1, z_2 - 1, z_3, \dots, z_m) \text{ mit } \sum_{i=1}^m z'_i = \sum_{i=1}^m z_i.$$

Es gilt:

$$\begin{aligned} \sum_{i=1}^m Q(z'_i) &= Q(1) + Q(z_2 - 1) + \sum_{i=3}^m Q(z_i) \\ &\leq Q(0) + Q(z_2) + \sum_{i=3}^m Q(z_i) = \sum_{i=1}^m Q(z_i). \end{aligned}$$

Falls \bar{z} das Maximum im zweiten Fall annimmt mit $z_i \geq 2$, so gibt es wie gezeigt stets ein z_j mit $z_j = 0$. Daraufhin wird in einem Schritt \bar{z} durch \bar{z}' ersetzt mit $z'_i = z_i - 1 > 0$ und $0 < z'_j = 1 < 2$. Dann wird das Maximum wie gezeigt auch von \bar{z}' angenommen. Dieser Ersetzungsprozess ist endlich, da \bar{z}' stets weniger Nullen hat als \bar{z} . Letzendlich wird der erste Fall erreicht, sobald jedes $z_i \geq 2$ durch 1 ersetzt wird. Damit ist es für die Bestimmung des Maximums ausreichend, nur den ersten Fall zu betrachten.

Es gilt also tatsächlich $Q(n) = c \cdot (\lfloor \varepsilon n \rfloor)$ für $n < (m+1)r$. \square

A.3 Untersuchung der Hilfsfunktion $\gamma: \mathbb{N} \rightarrow \mathbb{R}^+$

In diesem Abschnitt definieren und untersuchen wir eine etwas angenehmer zu behandelnde Funktion γ und schätzen am Ende des Abschnittes $Q(n)$ gegen $\gamma(n)$ ab.

A.3.1 Definition ($\gamma(n)$). γ ist eine Funktion von \mathbb{N} nach \mathbb{R}^+ definiert durch:

$$\gamma(n) = \begin{cases} 0 & \text{für } n = 0 \\ 1 + \frac{m}{\lceil m \cdot r \rceil - 1} (n - 1) & \text{für } 1 \leq n \leq n^* \\ 1 + \max_{\bar{n} \in \mathcal{N}(n)} \sum_{i=1}^m \gamma(n_i) & \text{für } n > n^*. \end{cases}$$

Es sei an die obige Definition $n^* = \lceil mr \rceil > 2$ erinnert. Offensichtlich wächst γ monoton.

Unmittelbar aus der Definition von γ folgt:

A.3.2 Lemma (Steigung von γ unterhalb der Grenze n^*).

$$\gamma(n+1) \leq 1 + \gamma(n) \text{ für } n < n^*$$

Beweis. Es ist $\gamma(1) = 1 \leq 1 + \gamma(0)$ und $\gamma(n^*) = m + 1$ nach Definition.

Weiter ist γ linear für $1 \leq n \leq n^*$ mit der Steigung

$$\frac{m}{\lceil m \cdot r \rceil - 1}.$$

Für den Beweis genügt es nachzuweisen, dass diese Steigung wegen $r > 1$ kleiner oder gleich 1 ist:

$$m < mr \implies m + 1 \leq \lceil mr \rceil \implies m \leq \lceil mr \rceil - 1 \implies \frac{m}{\lceil mr \rceil - 1} \leq 1. \quad \square$$

Nun wird gezeigt, dass auch für $n = n^*$ die Steigung höchstens 1 ist:

A.3.3 Lemma (Steigung von γ an der Grenze).

$$\gamma(n^* + 1) \leq 1 + \gamma(n^*)$$

Beweis. Wegen $\gamma(n^*) = m + 1$ ist $\gamma(n^* + 1) \leq m + 2$ nachzuweisen. Zunächst gilt:

$$m \leq \left\lfloor \frac{n^* + 1}{r} \right\rfloor \leq m + 1.$$

Dies folgt leicht aus $r > 1$ und $mr \leq \lceil mr \rceil = n^* \leq mr + 1$ durch Abschätzung:

$$m \leq \left\lfloor \frac{mr + 1}{r} \right\rfloor \leq \left\lfloor \frac{n^* + 1}{r} \right\rfloor \leq \left\lfloor \frac{mr + 2}{r} \right\rfloor = \left\lfloor m + \frac{2}{r} \right\rfloor = m + \left\lfloor \frac{2}{r} \right\rfloor \leq m + 1.$$

$\gamma(n^* + 1)$ ist definiert als

$$\gamma(n^* + 1) = 1 + \max_{\bar{z} \in \mathcal{N}(n^* + 1)} \sum_{i=1}^m \gamma(z_i).$$

Sei nun $\bar{z} \in \mathcal{N}(n^* + 1)$ so, dass das Maximum angenommen wird:

$$\gamma(n^* + 1) = 1 + \sum_{i=1}^m \gamma(z_i).$$

Dann ist $\sum_{i=1}^m z_i \geq m$, da $\gamma(n)$ monoton wächst und $\gamma(n^*) = m + 1$.

Außerdem folgt aus $r > 1$ direkt $\frac{n^*+1}{r} < n^* + 1$. Also ist

$$\frac{n^* + 1}{r} < n^* + 1 \implies \sum_{i=1}^m z_i \leq n^*.$$

Für die Elemente von \bar{z} können höchstens die folgenden vier Fälle auftreten:

I.Fall: Es gibt ein $z_j \geq n^*$.

Dieser Fall tritt höchstens ein, wenn $z_j = n^*$ und $z_i = 0$ für $i \neq j$.

Wegen $\gamma(0) = 0$ und $\gamma(n^*) = m + 1$ folgt

$$\gamma(n^* + 1) = 1 + \sum_{i=1}^m \gamma(z_i) = 1 + \gamma(n^*) + (m-1)\gamma(0) = 1 + m + 1 = m + 2, \text{ was zu zeigen war.}$$

II.Fall: Es gibt keine $z_j > 1$.

Wegen $\gamma(0) = 0 \leq \gamma(1) = 1$ folgt $\gamma(n^* + 1) \leq 1 + m \cdot \gamma(1) = m + 1 \leq m + 2$.

III.Fall: Es gibt ein $z_j > 1$ und es sind alle $z_i \neq 0$.

Wegen $\sum_{i=1}^m z_i \leq m + 1$ folgt dann $z_j = 2$ und $z_i = 1$, $i \neq j$. Dann folgt:

$$\gamma(n^* + 1) = 1 + (m-1)\gamma(1) + \gamma(2) = 1 + (m-1) + \gamma(2) = m + \gamma(2) \leq m + \gamma(1) + 1 = m + 2.$$

IV.Fall: Es gibt ein $z_j := k$ mit $n^* > k > 1$ und ein $z_i = 0$.

Ohne Einschränkung sei $z_1 = k$ und $z_2 = 0$. \bar{z} lässt sich nun durch \bar{z}' ersetzen, wobei \bar{z}' eine 0 weniger als Element hat. Sei $\bar{z}' = (k-1, 1, z_3, \dots, z_m)$. Dann ist $k-1 \neq 0$ und

$$\sum_{i=1}^m z'_i = \sum_{i=3}^m z_i.$$

Also ist auch $\bar{z}' \in \mathcal{N}(n^* + 1)$. Es wird gezeigt, dass dann auch für \bar{z}' das Maximum angenommen werden muss. Es wird benutzt, dass nach Lemma A.3.2 „Steigung von $\gamma(n)$ unter der Grenze“ gilt:

$$\gamma(k) + \gamma(0) \leq \gamma(k-1) + \gamma(1) \text{ für } n^* > k > 1.$$

Damit folgt dann:

$$\begin{aligned}
 \gamma(n^* + 1) &= 1 + \sum_{i=1}^m \gamma(z_i) \\
 &= 1 + \gamma(k) + \gamma(0) + \sum_{i=3}^m \gamma(z_i) \\
 &\leq 1 + \gamma(k-1) + \gamma(1) + \sum_{i=3}^m \gamma(z'_i) \\
 &= 1 + \sum_{i=1}^m \gamma(z'_i).
 \end{aligned}$$

Da also in diesem Fall jedes \bar{z} ohne Einschränkung durch ein \bar{z}' mit weniger Nullen ersetzt werden kann, erhält man schließlich ein m -Tupel, in dem überhaupt keine Nullen auftreten. Der Fall, dass keine Nullen auftreten, wurde jedoch schon zuvor behandelt. Damit ist die Fallunterscheidung abgeschlossen und es wurde gezeigt:

$$\gamma(n^* + 1) \leq m + 2 = 1 + \gamma(n^*). \quad \square$$

Der folgende Satz besagt, dass die Steigung von γ sogar für alle n höchstens 1 ist:

A.3.4 Satz (Steigung von γ).

Die Steigung von γ ist für alle n höchstens 1, d.h. es gilt für alle $n \geq 0$:

$$\gamma(n+1) \leq 1 + \gamma(n).$$

Beweis.

Der Satz wird mit vollständiger Induktion nach n bewiesen.

Induktionsanfang: Für $n \leq n^*$ wurde die Behauptung soeben bewiesen.

Induktionsvoraussetzung: Für festes $k \geq n^*$ gilt: $\gamma(n+1) \leq 1 + \gamma(n)$ für alle $n \leq k$

Induktionsbehauptung: $\gamma(n+1) \leq 1 + \gamma(n)$ für $n = k+1$.

Induktionsschritt: Sei $\bar{z} \in \mathcal{N}(n+1)$ so, dass die maximale Summe angenommen wird, d.h. $\gamma(n+1) = 1 + \sum_{i=1}^m \gamma(z_i)$. Wie beim Beweis von Lemma A.2.3 (S.120) führen wir mit der Eigenschaft $\forall k: \gamma(k+1) + \gamma(0) \leq \gamma(k) + \gamma(1)$ einen Reduktionsschritt $\bar{z} \mapsto \bar{z}'$ durch. Falls das m -Tupel \bar{z} nur aus Nullen besteht, d.h. $\bar{z} = (0, \dots, 0)$, so gilt $\gamma(n+1) = 1+0 \leq \gamma(n)+1$. Sei nun also $\bar{z} \neq (0, \dots, 0)$. Da $\gamma(n)$ symmetrisch bezüglich der z_i definiert ist, lässt sich

o.B.d.A. $z_1 \neq 0$ verlangen. Mit der Definition von $\gamma(n+1)$ erhält man durch Einsetzen der Induktionsvoraussetzung für $\gamma(z_1)$:

$$\begin{aligned}\gamma(n+1) &= 1 + \sum_{i=1}^m \gamma(z_i) \\ &= 1 + \gamma(z_1) + \sum_{i=2}^m \gamma(z_i) \\ &\leq 1 + 1 + \gamma(z_1 - 1) + \sum_{i=2}^m \gamma(z_i) \\ &\leq 1 + (1 + \max_{\bar{n} \in \mathcal{N}(n)} \sum_{i=1}^m \gamma(n_i)) \\ &= 1 + \gamma(n).\end{aligned}$$

In der vorletzten Umformung wurde dabei die Definition von $\mathcal{N}(n)$ verwendet und aufgezeigt, dass ein darin enthaltenes m -Tupel \bar{z}' vorliegt. (Vgl. Beweis von A.2.2, S.119).

Die letzte Umformung zu $1 + \gamma(n)$ folgt aus der Definition von $\gamma(n)$ für $n \geq \lceil mr \rceil$. Diese Voraussetzung ist für $n = k+1 = n^*+1 = \lceil mr \rceil + 1$ erfüllt. Damit ist der Induktionsbeweis vollständig und es gilt für alle $n \geq 0$:

$$\gamma(n+1) \leq 1 + \gamma(n). \quad \square$$

Es folgt die angekündigte Abschätzung von $Q(n)$ gegen $\gamma(n)$:

A.3.5 Satz ($Q(n)$ abgeschätzt gegen $\gamma(n)$).

$$Q(n) \leq c \cdot (1 + \varepsilon) \cdot \gamma(n).$$

Beweis. I.Fall: $1 \leq n \leq n^* = \lceil \frac{m}{r} \rceil$.

Wir zeigen, dass $n < (m+1)r$ gilt, so dass sich Lemma A.2.3 „Wert von $Q(n)$ für $n < (m+1)r$ “ anwenden lässt um den Wert von $Q(n)$ auszudrücken:

$$\frac{n}{r} \leq \frac{1}{r} \cdot \lceil mr \rceil < \frac{1}{r} (mr + 1) = m + \frac{1}{r} < m + 1 \implies n < (m+1)r.$$

Also ist nach Lemma A.2.3 zu zeigen:

$$c(\lceil \varepsilon n \rceil + 1) \leq c(1 + \varepsilon)\gamma(n)$$

Es wird zunächst $\gamma(n)$ nach unten abgeschätzt:

$$\gamma(n) = \frac{m}{\lceil m/\varepsilon \rceil - 1} (n-1) + 1 \geq \frac{m(n-1)}{m/\varepsilon} + 1 \geq \varepsilon(n-1) + 1 = \varepsilon n + 1 - \varepsilon.$$

Jetzt lässt sich der rechte Term in Gleichung 59 nach unten abschätzen:

$$\begin{aligned} c(1 + \varepsilon)\gamma(n) &\geq c(1 + \varepsilon)(\varepsilon n + 1 - \varepsilon) \\ &\geq c[(1 + \varepsilon)\varepsilon n + (1 + \varepsilon)(1 - \varepsilon)] \geq c(\varepsilon n + 1^2) \geq c(\lfloor \varepsilon n \rfloor + 1). \end{aligned}$$

II.Fall: $n > n^*$. Beweis mit vollständiger Induktion nach n :

Für kleine n gilt die Aussage nach dem I. Fall. Sei nun n fest und sei $z \in \mathcal{N}(n)$ so, dass das Maximum angenommen wird, d.h.:

$$Q(n) = c + \sum_{i=1}^m Q(z_i).$$

Nach der Induktionsvoraussetzung gilt dann die Abschätzung für alle $Q(z_i)$, da $z_i < n$. Mithilfe dieser erhält man die Abschätzung wie folgt:

$$\begin{aligned} Q(n) &= c + \sum_{i=1}^m Q(z_i) \leq c + c(1 + \varepsilon) \sum_{i=1}^m \gamma(z_i) \\ &\leq c + c(1 + \varepsilon) \cdot \max_{\bar{n} \in \mathcal{N}(n)} \sum_{i=1}^m \gamma(n_i) \\ &\leq c(1 + \varepsilon) \cdot \left(1 + \max_{\bar{n} \in \mathcal{N}(n)} \sum_{i=1}^m \gamma(n_i) \right) \\ &\stackrel{\text{def}}{=} c(1 + \varepsilon)\gamma(n). \quad \square \end{aligned}$$

A.4 Untersuchung der Hilfsfunktion $f: (0, \infty) \rightarrow \mathbb{R}$

A.4.1 Definition ($f: (0, \infty) \rightarrow \mathbb{R}$).

$$f(x) := \frac{1}{m-1}(m \cdot x^\alpha - 1)$$

Dabei ist $0 < \alpha < 1$ wie zuvor festgelegt:

$$\alpha = \frac{\log_r m}{1 + \log_r m} = \frac{\log_r m}{\log_r mr} = \log_{mr} m.$$

In der letzten Gleichung der Definition wurde die Rechenregel $\log_b a = \frac{\log_n a}{\log_n b}$ angewendet. Außerdem ist $\alpha > 0$ wegen $m \geq 2$ und $\alpha < 1$ wegen $mr > m$. Wie gleich nachgewiesen wird, ist $f(x)$ für $x > 0$ eine konkave Funktion. Man beachte, dass bei Haussler und Welzl die Begriffe „konkav“ und „konvex“ gegenüber der in der Mathematik üblichen Definition vertauscht sind [HW87, Anhang]. Die folgende Definition ist äquivalent zu einer Definition bei Forster [For01, Seite 158]:

A.4.2 Definition (Konkavität). Eine Funktion $f: (a, b) \rightarrow \mathbb{R}$ heißt konkav, wenn für alle $x, y \in (a, b)$ der Graph von f oberhalb der Geraden durch $(x, f(x))$ und $(y, f(y))$ verläuft.

Der folgende Satz gehört zur grundlegenden Analysis. Ein äquivalenter Satz findet sich bei Forster [For01, Seite 159, Satz 6].

A.4.3 Satz (über Konkavität). *Eine auf einem offenen Intervall definierte, zweimal stetig differenzierbare reelle Funktion ist genau dann konkav, wenn für alle x aus dem Definitionsbereich $f''(x) \leq 0$ ist.*

Um $\gamma(n)$ nach oben abzuschätzen, wird eine Funktion $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ definiert und gezeigt, dass $\gamma(n) \leq f(n)$ für alle $n \in \mathbb{N}$ gilt. Wir kommen nun zum angekündigten Nachweis, dass $f(x)$ konkav ist:

A.4.4 Lemma (Konkavität von $f(x)$). *$f(x)$ ist im Intervall $(0, \infty)$ eine monoton wachsende konkave Funktion.*

Beweis.

$$\begin{aligned} f'(x) &= \alpha \frac{m}{m-1} (x^{\alpha-1}) > 0 && \forall x > 0. \\ f''(x) &= (\alpha-1) \alpha \frac{1}{m-1} (x^{\alpha-2}) < 0 && \forall x > 0. \quad \square \end{aligned}$$

Wir zeigen zunächst, dass $\gamma(n)$ für $n \leq n^*$ durch $f(n)$ nach oben beschränkt ist:

A.4.5 Lemma ($\gamma(n)$ kleiner gleich $f(n)$ unter (und an) der Grenze n^*).

$$\gamma(n) \leq f(n) \text{ für } 1 \leq n \leq n^* = \lceil mr \rceil.$$

Beweis. Es ist $\gamma(1) = 1 = f(1)$ nach Definition. Weiter gilt $\gamma(n^*) = m + 1$. Da $f(x)$ nach Lemma A.4.4 monoton wächst, gilt:

$$\begin{aligned} f(n^*) &= f(\lceil mr \rceil) \\ &\geq f(mr) = \frac{1}{m-1} (m \cdot (mr)^\alpha - 1) \\ &= \frac{1}{m-1} (m \cdot (mr)^{\log_{mr} m} - 1) \\ &= \frac{1}{m-1} (m^2 - 1) = m + 1 = \gamma(n^*). \end{aligned}$$

Da γ auf $\{1, \dots, n^*\}$ linear ist und f an den Randpunkten 1 und n^* nicht unterhalb von γ liegt, folgt aus dem Satz über Konkavität: f liegt im gesamten Intervall nicht unterhalb von γ , d.h.:

$$\gamma(n) \leq f(n) \text{ für } 1 \leq n \leq n^*. \quad \square$$

Der folgende Satz besagt, dass γ überall durch f nach oben beschränkt wird. Zusammen mit dem Satz A.3.5, der $Q(n)$ gegen $\gamma(n)$ abschätzt, ergibt sich im Anschluss der Beweis für die das Abschätzungstheorem.

A.4.6 Satz ($\gamma(n)$ kleiner oder gleich $f(n)$).

$$\gamma(n) \leq f(n) \quad \text{für } n \geq 1.$$

Beweis. Nach Lemma A.4.5 wurde die Behauptung für $n \leq \lceil mr \rceil$ gezeigt. Sei nun also

$$n > \lceil mr \rceil.$$

Insbesondere ist dann $m < \frac{n}{r}$. Sei $\bar{z} \in \mathcal{N}(n)$ so, dass \bar{z} das Maximum ergibt, d.h.

$$\gamma(n) = 1 + \sum_{i=1}^m \gamma(z_i).$$

Behauptung (\dagger , siehe weiter unten im Fließtext) o.B.d.A. gilt $z_i \neq 0$ für alle i .

Angenommen, dies sei nicht der Fall. Sei dann o.B.d.A. $z_1 = 0$. Falls $z_j \leq 1$ für alle j , so erhält man sicher nicht das Maximum: Ersetzt man \bar{z} durch \bar{z}' mit $z'_1 := 1$ und $z'_j = z_j$ für $j \neq 1$, so ist $\sum_{i=1}^m z'(i) \leq m < \frac{n}{r}$ und damit auch $\bar{z}' \in \mathcal{N}(n)$. Es gilt $\sum_{i=1}^m \gamma(z'_i) > \sum_{i=1}^m \gamma(z_i)$ im Widerspruch zur Wahl von \bar{z} .

Sei nun $n_1 = 0$ und o.B.d.A. $n_2 \geq 2$. Dann führt man erneut ein Reduktionsargument durch. Dazu verwendet man den Satz A.3.4, nach dem $\gamma(k+1) + 0 \leq \gamma(k) + \gamma(1)$ für alle k ist. Sei \bar{z}' festgelegt durch $z'_1 = z_1 + 1 = 1$, $z'_2 = z_2 - 1 > 0$ und $z'_i = z_i$ sonst. Dann ist

$$\begin{aligned} \gamma(n) &= 1 + \sum_{i=1}^m \gamma(z'_i) \\ &= 1 + \gamma(0) + \gamma(z_2 - 1) + \sum_{i=3}^m \gamma(z'_i) \\ &\geq 1 + \gamma(1) + \gamma(z_2) + \sum_{i=3}^m \gamma(z_i) \\ &= 1 + \sum_{i=1}^m \gamma(z_i). \end{aligned}$$

Also wird das Maximum auch von \bar{z}' angenommen, wobei \bar{z}' eine Null weniger hat.

Damit ist die obige Behauptung (*) bewiesen und es wird nun o.B.d.A. angenommen, dass $z_i \neq 0$ für alle i .

Wir führen das Ende des Beweises nun mit vollständiger Induktion nach n .

Induktionsanfang: Für $0 < n < n^*$ gilt $\gamma(n) \leq f(n)$ nach Lemma A.4.5.

Induktionsvoraussetzung: $\gamma(k) \leq f(k)$ für alle $0 < k < n$.

Induktionsbehauptung: $\gamma(n) \leq f(n)$.

Sei $\bar{n} \in \mathcal{N}(n)$ nun so, dass $\gamma(n) = 1 + \sum_{i=1}^m \gamma(n_i)$. Wegen $\frac{n}{r} < n$ kann dann die Induktionsvoraussetzung (IV) für alle $\gamma(n_i)$ verwendet werden.

$$\begin{aligned}
 \gamma(n) &= 1 + \sum_{i=1}^m \gamma(n_i) \\
 &\leq 1 + \sum_{i=1}^m \frac{-1 + m \cdot n_i^\alpha}{m-1} \text{ nach der IV} \\
 &= \frac{-1}{m-1} + \frac{m}{m-1} \cdot \sum_{i=1}^m n_i^\alpha \\
 &\leq \frac{-1}{m-1} + \frac{m^2}{m-1} \cdot \left(\frac{\sum_{i=1}^m n_i}{m} \right)^\alpha \text{ wegen } \alpha < 1 \\
 &\leq \frac{-1}{m-1} + \frac{m^2}{m-1} \cdot \left(\frac{n}{mr} \right)^\alpha \\
 &= \frac{-1}{m-1} + \frac{m^2}{m-1} \cdot \frac{n^\alpha}{(mr)^\alpha} \\
 &= \frac{-1}{m-1} + \frac{m^2}{m-1} \cdot \frac{n^\alpha}{(mr)^{\log_{mr} m}} \\
 &= \frac{-1}{m-1} + \frac{m^2}{m-1} \cdot \frac{n^\alpha}{m} \\
 &= \frac{-1}{m-1} + \frac{m}{m-1} \cdot n^\alpha \\
 &= \frac{1}{m-1} (m \cdot n^\alpha - 1).
 \end{aligned}$$

□

A.5 Beweis des Abschätzungstheorems

Schließlich lässt sich das Theorem beweisen, nach dem gilt: $Q(n) \leq 4cn^\alpha$.

Beweis (des Abschätzungstheorems, S.119). Es werden die beiden Abschätzungen nach Satz A.3.5 („Abschätzung von $Q(n)$ gegen $\gamma(n)$ “) und Satz A.4.6 („ $\gamma(n)$ kleiner oder gleich $f(n)$ “) verwendet. Für $Q(0) = 0$ ist nichts zu zeigen. Für $n > 0$ ergibt sich:

$$\begin{aligned} Q(n) &\leq c \cdot (1 + \varepsilon) \cdot \gamma(n) \\ &\leq \frac{c(1 + \varepsilon)}{m - 1} (m \cdot n^\alpha - 1) \\ &\leq \frac{c(1 + \varepsilon)}{m - 1} \cdot m \cdot n^\alpha \\ &= c(1 + \varepsilon) \cdot \frac{m}{m - 1} \cdot n^\alpha \\ &\leq c \cdot 2 \cdot 2 \cdot n^\alpha = 4cn^\alpha. \end{aligned}$$

Damit ist das Theorem bewiesen. □

Erst mit diesem Beweis des Abschätzungstheorem ist die postulierte Laufzeit für Halbraum- und Simplex-Bereichsanfragen nach Theorem VI (S.93) bzw. Theorem VIII (S.103) vollständig (vgl. Beweisstruktur in Anhang B).

Anhang B

Beweisdiagramm

In Abbildung B.1 wird die Abhängigkeit der Theoreme sowie wichtiger Sätze dargestellt:

Theorem I (Stichprobentheorem, S.22)

Theorem II (über die maximale Wahrscheinlichkeit, S.28)

3.5.1 Satz (Wahrscheinlichkeit von ε -Netzen, S.47)

3.3.4 Lemma (Vererbung der VC-Dimension, S.41)

3.5.13 Satz (über die Existenz von kleinen ε -Netzen, S.57)

Theorem III (ε -Netze konstanter Größe, S.59)

4.5.2 Satz (Dimension von k -Korridoren, S.72)

Theorem IV (Das ε -Theorem für k -Korridore, S.72)

5.3.1 Satz (Einbettungssatz, S.75)

Theorem V (Das zentrale ε -Theorem, S.80)

5.5.1 Satz (Verifikation von Stichproben, S.81)

6.2.7 Satz (Halbraum-Anfragezeit in (ε, ν) -Bäumen, S.92)

6.3.1 Satz (Existenz von kleinen (ε, ν) -Bäumen, S.93)

Theorem VI (Existenz der Datenstruktur, S.93)

Theorem VII (Konstruktion der Datenstruktur, S.96)

Theorem VIII (Existenz der Datenstruktur für Simplizes, S.103)

Theorem IX (Konstruktion von (ε, ν) -Bäumen für Simplizes, S.104)

Theorem X (Abschätzungstheorem, S.119)

Bemerkung zum Beweisdiagramm:

Das Theorem VIII zur Existenz der Datenstruktur für Simplizes wird mit den gleichen Aussagen das Theorem VI zur Existenz der Datenstruktur für Halbräume bewiesen. Diese Analogie wird im Beweisdiagramm mit dem verwendeten Pfeil ausgedrückt.

Das Beweisdiagramm ist chronologisch in Kapitel aufgeteilt, auch die vertikale Position der Aussagen im Diagramm erfolgt chronologisch.

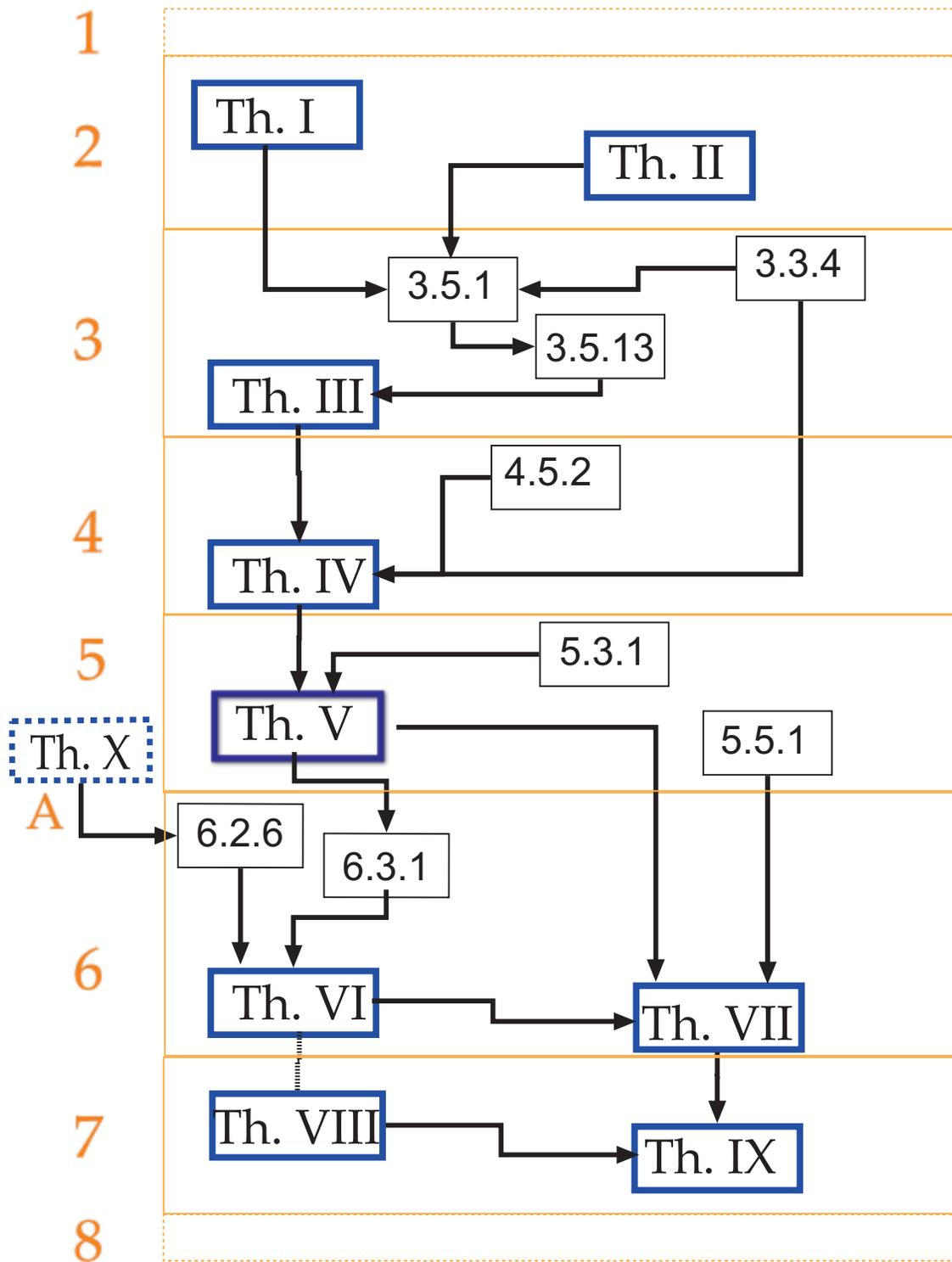


Abbildung B.1: Beweisstruktur

Literaturverzeichnis

- [ASP91] ALON, NOGA, JOEL H. SPENCER und PAUL ERDŐS: *The Probabilistic Method*. John Wiley & Sons, Inc., New York, NY, USA, 1991.
- [BEHW89] BLUMER, ANSELM, ANDRZEJ EHRENFEUCHT, DAVID HAUSSLER und MANFRED K. WARMUTH: *Classifying learnable geometric concepts with the Vapnik-Chervonenkis dimension*. Journal of the ACM, 36(4), 1989.
- [Ben75] BENTLEY, JON LOUIS: *Multidimensional Binary Search Trees Used for Associative Searching*. Communications of the ACM, 18(9), 1975.
- [Brö95] BRÖNNIMANN, HERVÉ: *Derandomization of Geometric Algorithms*. Doktorarbeit, Princeton University, 1995.
- [Cha89] CHAZELLE, BERNARD MARIE: *Lower bounds on the complexity of polytope range searching*. Journal of the American Mathematical Society, 2(4):637–666, 1989.
- [Ede87] EDELSBRUNNER, HERBERT: *Algorithms in combinatorial geometry*. Springer-Verlag, New York, NY, USA, New York, NY, USA, 1987.
- [ES74] ERDŐS, PAUL und JOEL H. SPENCER: *Probabilistic Methods in Combinatorics*. Academic Press Inc., New York, NY, USA, 1974.
- [EW86] EDELSBRUNNER, HERBERT und EMO WELZL: *Halfplanar range search in linear space and $O(n^{0.695})$ query time*. Information Processing Letters, 23(6):289–293, 1986.
- [For01] FORSTER, OTTO: *Analysis 1, 6. verbesserte Auflage*. Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, Braunschweig/Wiesbaden, 2001.
- [Grü67] GRÜNBAUM, BRANKO: *Convex Polytopes*. Interscience, New York, NY, USA, New York, 1967.
- [Hav06] HAVERKORT, HERMAN: *Small ε -nets for axis-parallel rectangles*. interner Bericht, 2006.

- [HW87] HAUSSLER, DAVID und EMO WELZL: *epsilon-nets and simplex range queries*. Discrete & Computational Geometry, Seiten 127–151, 1987.
- [KPW92] KOMLOS, JANOS, JANOS PACH und GERHARD WOEGINGER: *Almost tight bounds for epsilon-nets*. Discrete & Computational Geometry, 7(2):163–173, 1992.
- [Mat89] MATOUŠEK, JIŘÍ: *Construction of epsilon nets*. SCG '89: Proceedings of the fifth annual symposium on Computational geometry, 5:1–10, 1989.
- [Mat91] MATOUŠEK, JIŘÍ: *Approximations and Optimal Geometric Divide-and-Conquer*. Proceedings of the twenty-third annual ACM symposium on Theory of computing, Seiten 505–511, 1991.
- [Mat92] MATOUŠEK, JIŘÍ: *Efficient Partition Trees*. Discrete & Computational Geometry, 8:315–334, 1992.
- [Mat93] MATOUŠEK, JIŘÍ: *Epsilon-Nets and Computational Geometry*. New trends in discrete and computational geometry, 10:69–89, 1993.
- [Mat96] MATOUŠEK, JIŘÍ: *Derandomization in Computational Geometry*. Journal of Algorithms, 20:545–580, 1996.
- [Mer01] MERKLE, WOLFGANG: *Randomized Algorithms*. <http://www.math.uni-heidelberg.de/logic/merkle/ps/esslli2001.ps>, 2001. Vorlesungsnote.
- [MR95] MOTWANI, RAJEEV und PRABHAKAR RAGHAVAN: *Randomized Algorithms*. Cambridge University Press, Cambridge, UK, 1995.
- [OW02] OTTMANN, THOMAS und PETER WIDMAYER: *Algorithmen und Datenstrukturen*. Spektrum Akademischer Verlag, Berlin/Heidelberg, 2002.
- [Sch88] SCHMITZ, NORBERT: *Vorlesungen über Wahrscheinlichkeitstheorie*. Gesellschaft zur Förderung der Mathematischen Statistik, Münster, 1988.
- [Vah05] VAHRENHOLD, JAN: *Algorithmische Geometrie II*. unveröffentlichtes Skript zur Vorlesung an der Westfälischen Wilhelmsuniversität Münster, SS 2005.
- [VC71] VAPNIK, VLADIMIR und ALEXEY CHERVONENKIS: *On the uniform convergence of relative frequencies of events to their probabilities*. Theory of Probability and its Applications, 16(2):264–280, 1971.
- [Wil82] WILLARD, DAN: *Polygon Retrieval*. SIAM Journal on Computing, 11:149–166, 1982.

- [YY85] YAO, ANDREW CHI-CHIH und F. FRANCES YAO: *A general approach to d-dimensional geometric queries*. Proceedings of the 15th Symposium on Theory of Computation, Seiten 163–169, 1985.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Münster, 30. Januar 2008
