

**Diplomarbeit**

Probabilistische Analyse  
ganzzahliger Programmierung

Heiko Röglin



Diplomarbeit  
am Fachbereich Informatik  
der Universität Dortmund

23.08.2004

**Gutachter:**

Prof. Dr. Berthold Vöcking  
PD Dr. Detlef Sieling

---



# Zusammenfassung

Trotz ihrer fruchtbaren Resultate ist die worst-case Komplexitätstheorie nur bedingt dazu geeignet, die Effizienz von Algorithmen in der Praxis zu beurteilen, denn viele Algorithmen sind auf typischen Eingaben deutlich effizienter als auf worst-case Eingaben. Ein Schritt hin zu realitätsnäheren Eingabemodellen ist Spielman und Teng gelungen, als sie den Begriff der geglätteten Komplexität eingeführt und damit die Effizienz des Simplex-Algorithmus in praktischen Anwendungen erklärt haben [9]. Der geglätteten Komplexität liegt ein semi-zufälliges Eingabemodell zu Grunde, das zunächst einem Gegner erlaubt, eine möglichst schlechte Eingabe vorzugeben, die dann im zweiten Schritt einer leichten zufälligen Änderung (Perturbation) unterworfen wird, um zufällige und willkürliche Faktoren zu repräsentieren, die in praktischen Anwendungen stets vorhanden sind. Die größte erwartete Laufzeit auf einer solchen semi-zufälligen Eingabe, die der Gegner erreichen kann, wird als die geglättete Laufzeit des Algorithmus bezeichnet.

Beier und Vöcking haben für eine große Klasse binärer Optimierungsprobleme einen Zusammenhang zwischen der worst-case und der geglätteten Komplexität gefunden [2]. Sie haben gezeigt, dass ein Problem aus der untersuchten Klasse genau dann eine polynomielle geglättete Komplexität besitzt, wenn es einen Algorithmus für das Problem gibt, der im Worst-Case eine pseudopolynomielle Laufzeit besitzt. In dieser Arbeit verallgemeinern wir das Resultat von Beier und Vöcking auf den Fall ganzzahliger Optimierungsprobleme. Sind die Variablen nicht mehr binär, so gehen viele Eigenschaften verloren, die bei der Analyse des binären Falls ausgenutzt werden. Beispielsweise gibt es für eine lineare Funktion der Form  $w_1x_1 + \dots + w_nx_n$  und für zwei verschiedene Lösungen  $x, y \in \{0, 1\}^n$  stets einen Koeffizienten  $w_i$ , von der der Funktionswert genau einer der beiden Lösungen unabhängig ist. Diese Eigenschaft ist nicht erfüllt, wenn der Wertebereich der ganzzahligen Variablen eine beliebige Teilmenge  $\mathcal{D} \subseteq \mathbb{Z}$  ist. Ein weiteres Beispiel für eine Eigenschaft, die bei der Analyse des binären Falls verwendet wird und die im allgemeinen Fall nicht mehr erfüllt ist, betrifft das Runden von Koeffizienten. Rundet man in einer linearen Nebenbedingung der Form  $x_1w_1 + \dots + x_nw_n \leq t$  alle Koeffizienten  $w_i$  ab, so reduziert man im binären Fall damit das Gewicht  $x_1w_1 + \dots + x_nw_n$  oder verändert es nicht. Sobald  $\mathcal{D}$  auch negative Werte enthält, kann das Gewicht jedoch durch das Abrunden der Koeffizienten auch größer werden.

Entscheidend für unsere Analyse ganzzahliger linearer Optimierungsprobleme sind die strukturellen Eigenschaften Gewinner-, Verlierer- und Gültigkeits-Gap. Das Gewinner-Gap gibt an, um wie viel besser die beste erlaubte Lösung  $x^*$  im Vergleich zur zweitbesten erlaubten Lösung ist. Das Gültigkeits-Gap gibt an, wie weit das Gewicht von  $x^*$  von der Schranke  $t$  entfernt ist, und das Verlierer-Gap gibt an, wie weit die Gewichte der nicht erlaubten Lösungen, die besser als  $x^*$  sind, über der Schranke  $t$  liegen. Die Größe dieser Gaps kann als Maß dafür

aufgefasst werden, wie anfällig eine Instanz eines ganzzahligen Optimierungsproblems  $\Pi$  für kleine Änderungen der Koeffizienten ist. Sind alle drei Gaps nicht zu klein, so können alle Koeffizienten in den linearen Nebenbedingungen und der linearen Zielfunktion leicht abgeändert werden, ohne dadurch die optimale Lösung zu verändern. Wir werden zeigen, dass diese Gaps bei zufälligen Probleminstanzen mit hoher Wahrscheinlichkeit so groß sind, dass man die Koeffizienten geeignet runden kann, ohne dabei die optimale Lösung zu verändern. Diese Eigenschaft werden wir ausnutzen, um aufbauend auf einem pseudopolynomiellen Algorithmus für  $\Pi$  einen Algorithmus mit polynomieller geglätteter Laufzeit für  $\Pi$  zu konstruieren. Das Resultat über das Gewinner-Gap gab es bereits in einer unveröffentlichten Version der Arbeit von Beier und Vöcking [2]. Die Resultate über das Verlierer- und das Gültigkeits-Gap lassen sich aus o. g. Gründen nicht kanonisch vom binären auf den allgemeinen Fall übertragen. Sie sind im Rahmen dieser Diplomarbeit entstanden.

Ich danke Berthold Vöcking für die hervorragende Betreuung dieser Diplomarbeit. Außerdem bedanke ich mich bei Nicola Beume und Matthias Englert für Hinweise auf Fehler und Verbesserungsvorschläge.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Problemstellung und Motivation . . . . .	1
1.2	Ganzzahlige Optimierungsprobleme . . . . .	3
1.3	Geglättete Komplexität . . . . .	6
1.4	Überblick . . . . .	11
<b>2</b>	<b>Gewinner-, Verlierer- und Gültigkeits-Gap</b>	<b>15</b>
2.1	Vorbereitungen . . . . .	15
2.2	Analyse des Gewinner-Gaps . . . . .	16
2.3	Verlierer- und Gültigkeits-Gap im Falle einer stochastischen Nebenbedingung	18
2.3.1	Verlierer- und Gültigkeits-Gap für Dichten mit beschränkten Trägern	21
2.3.2	Verlierer- und Gültigkeits-Gap für allgemeine Dichten . . . . .	31
2.4	Verlierer- und Gültigkeits-Gap im Falle mehrerer stochastischer Nebenbedingungen . . . . .	33
<b>3</b>	<b>Die geglättete Komplexität ganzzahliger Optimierungsprobleme</b>	<b>37</b>
3.1	Beweis von Theorem 1.3.7 . . . . .	37
3.1.1	Ein Zertifizierer für stochastische Zielfunktionen im binären Fall . . . . .	38
3.1.2	Ein Zertifizierer für stochastische Zielfunktionen im allgemeinen Fall . . . . .	39
3.1.3	Ein Zertifizierer für stochastische Nebenbedingungen . . . . .	41
3.1.4	Ein Zertifizierer für stochastische Zielfunktionen und stochastische Nebenbedingungen . . . . .	43
3.1.5	Vom pseudopolynomiellen Algorithmus zur polynomiellen geglätteten Komplexität . . . . .	44
3.1.6	Von polynomieller geglätteter Komplexität zum pseudopolynomiellen Algorithmus . . . . .	48
3.2	Allgemeinere Perturbationsmodelle . . . . .	49

3.2.1	Unterschiedlich perturbierte Koeffizienten . . . . .	49
3.2.2	Nullerhaltende Perturbationen . . . . .	50
3.3	Ganzzahlige Packungs- und Überdeckungsprobleme . . . . .	51
<b>4</b>	<b>Ausblick</b>	<b>55</b>
<b>A</b>	<b>Grundlagen der Wahrscheinlichkeitstheorie</b>	<b>57</b>
A.1	Wahrscheinlichkeitsräume . . . . .	57
A.2	Verteilungen und Dichten . . . . .	59
A.3	Eindimensionale Zufallsvariablen . . . . .	60
A.4	Zweidimensionale Zufallsvariablen . . . . .	62
A.5	Bedingte Dichten und bedingte Verteilungen . . . . .	63
<b>B</b>	<b>Ergänzung zum Beweis von Lemma 2.3.10</b>	<b>65</b>
	<b>Literaturverzeichnis</b>	<b>69</b>
	<b>Symbolverzeichnis</b>	<b>70</b>
	<b>Index</b>	<b>72</b>

# Einleitung

## 1.1 Problemstellung und Motivation

Ein grundlegendes Ziel einer guten Theorie sollte es sein, Beobachtungen, die in der Praxis gemacht werden, erklären und vorhersagen zu können. Bezogen auf die Komplexitätstheorie und die Theorie vom Entwurf effizienter Algorithmen bedeutet das, dass zuverlässige Aussagen darüber getroffen werden sollten, welche Probleme in praktischen Anwendungen effizient gelöst werden können und wie effiziente Algorithmen für diese Probleme aussehen.

Um die Effizienz von Algorithmen zu messen, muss zunächst ein *Komplexitätsmaß* festgelegt werden. Das in der Theorie mit Abstand am häufigsten verwendete Maß ist das der *maximalen Rechenzeit* (*worst-case Komplexität*), dabei wird für jede Eingabelänge die (erwartete) Laufzeit auf der für den Algorithmus schlechtesten Eingabe als Referenz genommen. Die auf diesem Komplexitätsmaß aufbauende Komplexitätstheorie zählt zweifelsfrei zu den größten Errungenschaften der theoretischen Informatik. Sie erreichte 1971 einen ersten Höhepunkt als es erstmals gelang, die NP-Vollständigkeit eines Problems nachzuweisen [4], und hat seitdem viele weitere interessante und fruchtbare Resultate hervorgebracht.

Trotzdem muss man kritisch feststellen, dass das Ziel, diejenigen Probleme und Algorithmen zu identifizieren, die effizient lösbar bzw. effizient durchführbar sind, damit nur zum Teil erreicht wurde. Auch im Falle  $P \neq NP$  gibt es keinen Grund zu resignieren, wenn das betrachtete Problem NP-hart ist. Es gibt dann zwar keinen Algorithmus, der auf allen Eingaben eine polynomielle Laufzeit hat, es kann jedoch Algorithmen geben, die auf den Eingaben, die in praktischen Anwendungen vorkommen, effizient sind oder die im Erwartungswert eine polynomielle Laufzeit haben, wenn eine Eingabe zufällig ausgewählt wird. Außerdem gibt es für viele NP-harte Probleme effiziente und gute Approximationsalgorithmen, auf die wir in dieser Arbeit jedoch nicht weiter eingehen werden.

Wenn die Laufzeit eines Algorithmus nicht mehr nur essentiell von der Eingabelänge sondern auch von der konkreten Eingabe abhängt, stellt das Komplexitätsmaß der maximalen Rechenzeit folgerichtig kein realistisches Maß zur Beurteilung der Effizienz des Algorithmus dar, denn dann kann es passieren, dass der Algorithmus trotz einer großen Laufzeit auf der schlimmsten Eingabe auf fast allen Eingaben effizient ist. Ein bekanntes Beispiel hierfür ist der *Simplex-Algorithmus* zur Lösung *linearer Optimierungsprobleme*. Der Simplex-Algorithmus hat im schlimmsten Falle eine exponentielle Laufzeit, er ist jedoch in praktischen Anwendungen den bekannten polynomiellen Algorithmen ebenbürtig und oftmals sogar überlegen.

Ein Komplexitätsmaß, das die Laufzeiten in praktischen Anwendungen unter gewissen Umständen besser widerspiegeln kann, ist die *durchschnittliche Komplexität* (*average-case*)

*Komplexität*). Dazu muss für jedes  $n \in \mathbb{N}$  eine Wahrscheinlichkeitsverteilung auf der Menge der Eingaben der Länge  $n$  gegeben sein. Die durchschnittliche Laufzeit des Algorithmus für die Eingabelänge  $n$  ist dann definiert als die erwartete Laufzeit bezüglich der gegebenen Verteilung auf der Menge der Eingaben dieser Länge. Die Schwachstelle dieses Komplexitätsmaßes liegt in der Wahl der Wahrscheinlichkeitsverteilung, denn in praktischen Anwendungen ist in aller Regel nicht klar, welche Eingabe mit welcher Wahrscheinlichkeit ausgewählt wird, und außerdem ändert sich die Verteilung mit dem Anwendungsszenario. Das bedeutet, dass eine Analyse der durchschnittlichen Laufzeit eines Algorithmus stets nur eine Aussage über ein spezielles Anwendungsszenario treffen kann.

Untersucht man die durchschnittliche Komplexität für eine „falsche“ Verteilung, so sind keine Rückschlüsse auf das Verhalten des Algorithmus in der Praxis möglich, denn Eingaben, die in der Praxis vorkommen, haben in der Regel eine spezielle Struktur, die bei den meisten Problemen beispielsweise bei einer uniformen Verteilung der Eingaben nicht vorhanden ist. Deswegen sind Analysen der durchschnittlichen Laufzeit auch nur bedingt dazu geeignet, das Verhalten von Algorithmen in der Praxis zu erklären.

Es ist zwar unmöglich, ein Eingabemodell mathematisch zu beschreiben, das die Situation in praktischen Anwendungen exakt widerspiegelt, ein viel versprechender neuer Ansatz auf dem Weg hin zu realitätsnäheren Eingabemodellen ist jedoch im Jahre 2001 Spielman und Teng gelungen [9], als sie das Komplexitätsmaß der *geglätteten Komplexität* (*smoothed Komplexität*) eingeführt haben. Dieses Maß ist eine Mischung aus worst-case und average-case Komplexität und beruht auf einem zweistufigen *semi-zufälligen Eingabemodell*. Im ersten Schritt darf dabei ein Gegner deterministisch eine beliebige Eingabe vorgeben, die daraufhin im zweiten Schritt einer kleinen zufälligen Änderung (*Perturbation*) unterzogen wird. Der zweite Schritt soll zufällige und willkürliche Faktoren repräsentieren, die bei praktischen Anwendungen stets gegeben sind.

Um die geglättete Laufzeit eines Algorithmus für Eingaben der Länge  $n$  zu bestimmen, wird zunächst für jede Eingabe dieser Länge die erwartete Laufzeit des Algorithmus auf einer zufälligen Perturbation der Eingabe berechnet und anschließend das Maximum über diese Erwartungswerte gebildet. Wie genau die zufälligen Änderungen geschehen, wird durch ein sogenanntes *Perturbationsmodell* festgelegt, dieses sollte so definiert sein, dass eine Perturbation möglichst viele strukturelle Eigenschaften einer Eingabe erhält. Für eine *geglättete Analyse* ist es wichtig, dass das Perturbationsmodell parametrisiert ist und man über einen Parameter festlegen kann, wie stark eine Eingabe, die der Gegner vorgibt, durch eine Perturbation typischerweise verändert wird. Die geglättete Laufzeit wird dann als Funktion dieses Parameters und der Eingabegröße aufgefasst.

Das verdeutlichen wir am Beispiel des *Simplex-Algorithmus*, der von Spielman und Teng in der Arbeit, in der sie das Maß der geglätteten Komplexität eingeführt haben, untersucht wird. In dieser Analyse wird angenommen, dass ein Gegner ein *lineares Programm* vorgibt, bei dem die Koeffizienten der Nebenbedingungen im Intervall  $[-1, 1]$  liegen. Jeder dieser Koeffizienten wird dann unabhängig von den anderen dadurch abgeändert, dass zu ihm eine normalverteilte Zufallsvariable mit Erwartungswert 0 und Standardabweichung  $\sigma$  addiert wird. Spielman und Teng haben gezeigt, dass die geglättete Laufzeit des Simplex-Algorithmus polynomiell in der Anzahl der Variablen und in  $1/\sigma$  ist, und haben somit eine theoretische Begründung dafür gefunden, dass der Simplex-Algorithmus in praktischen Anwendungen effizient ist.

Man kann die geglättete Laufzeit eines Algorithmus auch als ein Maß dafür auffassen, wie



dicht die worst-case Eingaben im Raum aller Eingaben liegen. Gibt es Bereiche, in denen viele für den Algorithmus schlechte Eingaben dicht beisammen liegen, so ist die geglättete Laufzeit groß. Liegen die schlechten Eingaben jedoch isoliert, so ist die geglättete Laufzeit kleiner, dann kann es zwar Eingaben geben, auf denen die Laufzeit des Algorithmus sehr groß ist, man muss jedoch viel Pech haben, um in praktischen Anwendungen tatsächlich eine solche Eingabe zu treffen.

Nach der Arbeit von Spielman und Teng sind zahlreiche weitere Arbeiten erschienen, in denen die geglättete Komplexität diverser Probleme untersucht wurde. (Eine aktuelle Liste aller Arbeiten über geglättete Komplexität ist auf der Homepage von Daniel Spielman zu finden [8].) Beier und Vöcking haben die geglättete Komplexität diskreter Optimierungsprobleme untersucht [2]. Sie haben *binäre Programme* und eine noch allgemeinere Klasse *binärer Optimierungsprobleme* untersucht und einen Zusammenhang zwischen der worst-case Komplexität und der geglätteten Komplexität dieser Probleme herleiten können. Für die betrachtete Klasse binärer Optimierungsprobleme haben sie gezeigt, dass ein Problem aus dieser Klasse genau dann eine polynomielle geglättete Komplexität besitzt, wenn es einen randomisierten Algorithmus für das Problem gibt, der im Worst-Case eine pseudopolynomielle erwartete Laufzeit hat.

In dieser Arbeit werden wir das Resultat von Beier und Vöcking für binäre Optimierungsprobleme auf den Fall *ganzzahliger Optimierungsprobleme* verallgemeinern. Dieses Resultat gilt ebenso wie das von Beier und Vöcking für eine große Klasse von Perturbationsmodellen, allerdings müssen die Absolutwerte der ganzzahligen Variablen und damit auch die Größe des Wertebereiches polynomiell in der Anzahl der Variablen beschränkt sein.

Wir werden im nächsten Abschnitt formal den Begriff eines ganzzahligen Optimierungsproblems einführen und die Klasse derjenigen Probleme beschreiben, auf die sich unsere Analyse anwenden lässt. In dem darauf folgenden Abschnitt werden wir definieren, was wir unter *polynomieller geglätteter Komplexität* verstehen und wie das semi-zufällige Eingabemodell für die untersuchten ganzzahligen Optimierungsprobleme aussieht. Im letzten Abschnitt dieses Kapitels geben wir einen Überblick über die Analyse, die in den darauf folgenden Kapiteln durchgeführt wird.

## 1.2 Ganzzahlige Optimierungsprobleme

**Definition 1.2.1 (Ganzzahliges Optimierungsproblem).** *Ein ganzzahliges Optimierungsproblem  $\Pi$  besteht aus den folgenden Komponenten:*

- eine Menge von Probleminstanzen (Eingaben)  $\mathcal{I}$ ,
- für jede Eingabe  $I \in \mathcal{I}$  eine Menge  $\mathcal{D}(I) \subseteq \mathbb{Z}$ , die den Wertebereich der ganzzahligen Variablen angibt,
- für jede Eingabe  $I \in \mathcal{I}$  eine Zahl  $n(I) \in \mathbb{N}$ , die die Anzahl der ganzzahligen Variablen in der Probleminstanz  $I$  angibt,
- für jede Eingabe  $I \in \mathcal{I}$  eine Menge  $\mathcal{S}(I) \subseteq \mathcal{D}(I)^{n(I)}$  von erlaubten Lösungen,

- für jede Eingabe  $I \in \mathcal{I}$  eine Zielfunktion  $f_I : \mathcal{S}(I) \rightarrow \mathbb{R}$ , die maximiert oder minimiert werden soll.

Eine Probleminstanz mit einer linearen Zielfunktion und einer Menge von erlaubten Lösungen  $\mathcal{S}$ , die ausschließlich durch lineare Nebenbedingungen beschrieben wird, heißt **ganzzahliges Programm**.

Viele ganzzahlige Optimierungsprobleme, die in praktischen Anwendungen auftreten, besitzen eine lineare Zielfunktion, deren Koeffizienten Teil der Eingabe sind. Wir veranschaulichen das an den folgenden Beispielen. Es wird in diesen Beispielen für natürliche Zahlen  $n$  die Notation  $[n]$  verwendet, darunter verstehen wir in dieser Arbeit die Menge  $\{1, \dots, n\}$ .

**Beispiel 1.2.2 (Traveling Salesperson Problem).** Eine Eingabe für das *Traveling Salesperson Problem (TSP)* besteht aus einem vollständigen, gerichteten und gewichteten Graphen  $G = (V, l)$ , wobei  $V = \{1, \dots, n\}$  die Menge der Städte sei und  $l : V \times V \rightarrow \mathbb{R}_{\geq 0}$  die Abstände der Städte angebe. Gesucht ist eine Rundreise minimaler Länge im Graphen  $G$ . Um dieses Problem als ganzzahliges Optimierungsproblem zu formulieren, führen wir für jede Kante eine binäre Variable ein, das heißt, wir setzen  $\mathcal{D} = \{0, 1\}$  und definieren die binären Variablen  $x_{i,j}$  für  $i, j \in [n]$  mit  $i \neq j$ . Durch diese Variablen wird eine Kantenauswahl im Graphen beschrieben, die genau diejenigen Kanten enthält, deren zugehörige Variablen den Wert 1 haben. Die Menge der erlaubten Lösungen  $\mathcal{S} \subseteq \{0, 1\}^{n^2-n}$  besteht in diesem Fall aus genau den Variablenbelegungen, die Rundreisen repräsentieren, und für die Zielfunktion  $f : \mathcal{S} \rightarrow \mathbb{R}$  gilt für jedes  $x \in \{0, 1\}^{n^2-n}$

$$f(x) = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n l(i, j) x_{i,j}.$$

Somit ist die Zielfunktion linear und ihre Koeffizienten sind als Teil der Eingabe gegeben, es sind nämlich die Funktionswerte von  $l$ . ◇

**Beispiel 1.2.3 (Uncapacitated Facility Location Problem).** Ein weiteres Beispiel ist das *Uncapacitated Facility Location Problem (UFL)*. Dabei ist eine Menge von Konsumenten  $N = \{1, \dots, n\}$  und eine Menge von möglichen Fabrikstandorten  $M = \{1, \dots, m\}$  gegeben. Es ist zu entscheiden, an welchen Standorten Fabriken eröffnet werden, unter der Vorgabe, dass die entstehenden Kosten — wie könnte es auch anders sein — minimal sind. Die Kosten setzen sich dabei aus zwei konkurrierenden Summanden zusammen, zum einen kostet das Betreiben einer Fabrik an Standort  $j$  für  $j \in M$  einen Betrag von  $f_j \in \mathbb{R}_{\geq 0}$  vielen Geldeinheiten, zum anderen muss jeder Konsument von einer Fabrik beliefert werden und die Kosten, die dafür anfallen, entsprechen dem Abstand zur Fabrik, von der er beliefert wird. Wird Konsument  $i \in N$  von Fabrik  $j \in M$  beliefert, so fallen dafür  $c_{i,j} \in \mathbb{R}_{\geq 0}$  viele Geldeinheiten als Kosten an. Die Zahlen  $f_j$  und  $c_{i,j}$  sind Teil der Eingabe. Wir werden dieses Problem nun als ganzzahliges Programm formulieren. Dazu führen wir  $nm + m$  viele binäre Variablen ein, für  $i \in N$  und  $j \in M$  sei

$$x_{i,j} = \begin{cases} 1 & \text{falls } \text{Konsument } i \text{ von Fabrik } j \text{ beliefert wird} \\ 0 & \text{sonst} \end{cases}$$

und

$$y_j = \begin{cases} 1 & \text{falls } \text{Fabrik } j \text{ eröffnet wird} \\ 0 & \text{sonst} \end{cases}.$$

Die Zielfunktion  $f$  ist linear, denn es gilt

$$f(x, y) = \sum_{j \in M} y_j f_j + \sum_{i \in N} \sum_{j \in M} x_{i,j} c_{i,j}.$$

Wir können die Menge  $\mathcal{S}$  der erlaubten Lösungen ebenfalls durch lineare Nebenbedingungen beschreiben. Um sicherzustellen, dass jeder Konsument von genau einer Fabrik beliefert wird und dass diese Fabrik wirklich eröffnet wird, genügen die folgenden Nebenbedingungen

$$\begin{aligned} \sum_{j \in M} x_{i,j} &= 1 && \text{für alle } i \in N, \\ \sum_{i \in N} x_{i,j} &\leq ny_j && \text{für alle } j \in M. \end{aligned}$$

Der wesentliche Unterschied zwischen der linearen Zielfunktion und den linearen Nebenbedingungen in diesem Beispiel ist, dass die Koeffizienten der Zielfunktion als Teil der Eingabe gegeben sind, die Koeffizienten der Nebenbedingungen definieren hingegen die Problemstruktur und sind kein expliziter Teil der Eingabe. Im Abschnitt über geglättete Komplexität wird auf diesen Unterschied noch weiter eingegangen.  $\diamond$

Außerdem kann der Fall eintreten, dass ein ganzzahliges Optimierungsproblem lineare Nebenbedingungen besitzt, deren Koeffizienten als Teil der Eingabe gegeben sind. Auch das wollen wir an einem Beispiel veranschaulichen.

**Beispiel 1.2.4 (Constrained Shortest Path Problem).** Wir betrachten das Problem, kürzeste Wege in einem Graphen unter einer zusätzlichen linearen Nebenbedingung zu berechnen, also das sogenannte *Constrained Shortest Path Problem (CSP)*. Sei dazu  $G = (V, E, l)$  ein ungerichteter und gewichteter Graph mit den Knoten  $V$ , den Kanten  $E = \{e_1, \dots, e_m\}$  und den Kantengewichten  $l : E \rightarrow \mathbb{R}_{\geq 0}$ . Ferner seien in  $G$  zwei Knoten  $s \in V$  und  $t \in V$  ausgezeichnet. Wir möchten einen kürzesten Weg von  $s$  nach  $t$  finden, wobei wir unter der Länge eines Weges die Anzahl der benutzten Kanten verstehen wollen, man kann sich beispielsweise ein Routing-Problem vorstellen, bei dem man einen kürzesten Weg von  $s$  nach  $t$  sucht. Die Funktion  $l$  gibt dann für jede Kante an, wie lange es dauert, eine Nachricht über die Kante zu schicken. Man verlangt von dem Pfad von  $s$  nach  $t$ , dass er so kurz wie möglich ist unter der Bedingung, dass das Senden einer Nachricht über den Pfad nicht länger dauert als eine vorgegebene Zeitschranke. Dieses Problem ist NP-hart, wir werden es nun als ganzzahliges Optimierungsproblem formulieren. Dazu führen wir auch hier wieder für jede Kante  $e_i \in E$  eine binäre Variable  $x_i$  ein. Durch diese Variablen wird eine Kantenauswahl beschrieben und die Menge  $\mathcal{S}' \subseteq \{0, 1\}^m$  bezeichne alle Auswahlen, die einen Pfad von  $s$  nach  $t$  darstellen. Die zusätzliche Nebenbedingung ist linear und kann dargestellt werden als

$$x_1 l(e_1) + \dots + x_m l(e_m) \leq T.$$

Diese lineare Nebenbedingung beschreibt einen Halbraum  $\mathcal{B} \subseteq \{0, 1\}^m$  und die Menge der erlaubten Lösungen  $\mathcal{S}$  ergibt sich als Schnitt von  $\mathcal{S}'$  und  $\mathcal{B}$ . Für die Zielfunktion  $f$  gilt nun einfach  $f(x) = x_1 + \dots + x_m$ . Somit ist die Menge der erlaubten Lösungen in diesem Fall teilweise beschrieben durch eine lineare Nebenbedingung, deren Koeffizienten Teil der Eingabe sind. Man könnte das Beispiel noch dahingehend erweitern, dass man mehrere zusätzliche lineare Nebenbedingungen einfügt.  $\diamond$

Durch eine geeignete Kodierung kann jedes diskrete Optimierungsproblem in Form eines ganzzahligen Optimierungsproblems formuliert werden. Uns interessieren in dieser Arbeit nur diejenigen ganzzahligen Optimierungsprobleme, die mindestens eine lineare Komponente enthalten, die Teil der Eingabe ist.

**Definition 1.2.5 (Ganzzahliges lineares Optimierungsproblem).** *Ein ganzzahliges Optimierungsproblem  $\Pi$  heißt **ganzzahliges lineares Optimierungsproblem**, wenn es mindestens eine lineare Komponente enthält, die Teil der Eingabe ist, das heißt, wenn entweder die Zielfunktion linear ist und ihre Koeffizienten Teil der Eingabe sind oder wenn es mindestens eine lineare Nebenbedingung mit dieser Eigenschaft gibt.*

### 1.3 Geglättete Komplexität

Wir beginnen damit, das Perturbationsmodell zu beschreiben, das unserer Analyse zu Grunde liegt, dabei setzen wir die Kenntnis der in Anhang A vorgestellten Begriffe über Wahrscheinlichkeitstheorie voraus. Im Gegensatz zu dem Perturbationsmodell, das bei der Analyse des Simplex-Algorithmus benutzt wird, darf bei diskreten Problemen nicht jede lineare Komponente von stochastischer Natur sein, weil dadurch im Allgemeinen die Problemstruktur zerstört würde. Stattdessen sind in unserem Modell nur solche linearen Komponenten stochastisch, deren Koeffizienten als Teil der Eingabe gegeben sind.

Wir verdeutlichen diesen Sachverhalt an den Beispielen 1.2.3 und 1.2.4. Beim Uncapacitated Facility Location Problem sind zwar sowohl die Zielfunktion als auch die Nebenbedingungen linear, aber nur die Koeffizienten der Zielfunktion sind als Teil der Eingabe gegeben, sie entsprechen den Kosten für das Betreiben einer Fabrik und den Kosten für das Beliefern der Konsumenten. Eine zufällige Änderung dieser Koeffizienten liefert zwar eine andere Instanz des UFL Problems, die Semantik des Problems ändert sich dadurch jedoch nicht. Ändert man hingegen die Koeffizienten der Nebenbedingungen ab, so ändert man die Problemstruktur, die Semantik des so entstehenden Problems hat mit dem UFL Problem im Allgemeinen nichts mehr zu tun. Bei diesem Problem ist es also wichtig, dass nur die Zielfunktion stochastisch ist und der Rest deterministisch bleibt. Beim CSP Problem ist es hingegen nur sinnvoll, die zusätzliche lineare Nebenbedingung zufällig zu ändern, denn bei diesem Problem würde eine Änderung der Zielfunktion oder von  $\mathcal{S}'$  die Semantik ändern.

Das semi-zufällige Eingabemodell, das unserer Analyse der geglätteten Komplexität zu Grunde liegt, sieht vor, dass zunächst diejenigen linearen Komponenten ausgewählt werden, die von stochastischer Natur sein sollen. Dabei ist nur wichtig, dass dies, wie oben erläutert, nur solche linearen Komponenten sein dürfen, deren Koeffizienten Teil der Eingabe sind, und dass mindestens eine Komponente als stochastisch ausgewählt wird. Die Koeffizienten dieser ausgewählten linearen Komponenten werden dann von einem Gegner vorgegeben und anschließend zufällig abgeändert. Dabei gehen wir davon aus, dass der Gegner eine Eingabe vorgibt, bei der die stochastischen Koeffizienten und die Schranken der stochastischen Nebenbedingungen alle im Intervall  $[-1, 1]$  oder im Intervall  $[0, 1]$  liegen, je nachdem, ob sie negative Werte annehmen dürfen oder nicht. Diese Vorgabe ist keine Einschränkung an die Klasse der betrachteten Probleme, denn durch eine entsprechende Skalierung der linearen Komponenten kann sie stets erreicht werden, ohne dabei das Problem zu ändern.

### 1.3. Geglättete Komplexität

---

Wie genau die Perturbation der stochastischen Koeffizienten erfolgt, wird durch eine Wahrscheinlichkeitsdichte spezifiziert. Sei  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  eine Dichte mit  $\sup_{s \in \mathbb{R}} f(s) = 1$ , die eine reelle Zufallsvariable mit endlichem absoluten Erwartungswert beschreibt, für die also  $E = \int_{\mathbb{R}} |s| f(s) ds < \infty$  gilt. Wir definieren für jedes  $\phi \geq 1$  eine Dichte  $f_\phi$  durch Skalieren von  $f$ , dabei gelte  $f_\phi(s) = \phi f(s/\phi)$  für alle  $s \in \mathbb{R}$ . Es ist dann  $\sup_{s \in \mathbb{R}} f_\phi(s) = \phi$  und der absolute Erwartungswert einer Zufallsvariablen, die durch  $f_\phi$  beschrieben wird, ist  $E/\phi$ . Eine  $\phi$ -Perturbation einer Eingabe des Gegners erhält man dadurch, dass zu jedem stochastischen Koeffizienten eine unabhängige Zufallsvariable addiert wird, deren Verteilung durch die Dichte  $f_\phi$  beschrieben wird. Je größer das Supremum  $\phi$  der Dichte ist, desto besser kann ein Gegner worst-case Eingaben approximieren, weil mit steigendem  $\phi$  die erwartete Änderung der Koeffizienten immer kleiner wird.

In dem oben beschriebenen Szenario sind die stochastischen Koeffizienten mit Wahrscheinlichkeit 1 irrationale Zahlen, das heißt, sie besitzen keine endliche Darstellung und können von einem Algorithmus auch nicht komplett gelesen werden. Stattdessen wird jeder stochastische Koeffizient durch ein *Orakel* repräsentiert, das auf Anfrage in konstanter Zeit jeweils ein Bit des Koeffizienten ausgibt, dabei werden die Bits eines Koeffizienten „von links nach rechts“ vom Orakel zurückgegeben. Auch die Schranken der stochastischen Nebenbedingungen, die nicht perturbiert werden, dürfen irrational sein und werden ebenfalls durch Orakel beschrieben. Ein Algorithmus kann dann entscheiden, wie viele Bits er von den Orakeln erfragen möchte, also ab welcher Stelle er die Koeffizienten und die Schranken rundet. Wenn wir die Eingabegröße einer Problem Instanz messen, so werden wir aus diesem Grunde auch die Länge der Darstellungen der Koeffizienten und der Schranken in den stochastischen Komponenten nicht in Betracht ziehen. Stattdessen gehen wir davon aus, dass jeder der stochastischen Koeffizienten und jede Schranke eine „virtuelle Länge“ von 1 hat unabhängig davon, wie lang die Darstellung ist.

Von der Dichte  $f$  fordern wir außer den oben genannten Eigenschaften noch, dass es effizient möglich sein muss, Stichproben gemäß der durch  $f$  beschriebenen Verteilung zu erzeugen. Da es sich bei diesen Stichproben um reelle und im Allgemeinen um irrationale Zahlen handelt, können wir nicht fordern, dass es einen Algorithmus gibt, der eine solche Stichprobe komplett erzeugt. Wir benötigen jedoch die Voraussetzung, dass es eine probabilistische Turing-Maschine gibt, die eines der oben beschriebenen Orakel effizient simulieren kann. Konkret verlangen wir, dass ein Polynom  $p$  existiert und eine probabilistische Turing-Maschine, die die ersten  $k$  Bits rechts vom Komma einer Stichprobe in Zeit  $p(k)$  berechnen kann. Ebenso muss diese Turing-Maschine den ganzzahligen Anteil, also die Bits links vom Komma, in polynomieller Zeit bezogen auf die Anzahl dieser Bits, berechnen können.

Nach dieser Vorrede können wir nun dazu kommen, formal zu definieren, was wir unter ganzzahligen linearen Optimierungsproblemen mit polynomieller geglätteter Komplexität verstehen. Sei  $\Pi$  ein ganzzahliges lineares Optimierungsproblem und sei ein Perturbationsmodell  $\mathcal{H}$ , das heißt eine Dichte  $f$  mit den oben beschriebenen Eigenschaften und eine Auswahl von linearen Komponenten, die von stochastischer Natur sein sollen, gegeben. Dann bezeichne  $\mathcal{I}_N$  alle Eingaben für  $\Pi$  der Länge  $N$  und für eine Eingabe  $I \in \mathcal{I}_N$  bezeichne  $\mathcal{H}(I, \phi)$  die Zufallsvariable, die eine  $\phi$ -Perturbation von  $I$  beschreibt.

**Definition 1.3.1 (Polynomielle geglättete Komplexität).** *Ein ganzzahliges lineares Optimierungsproblem  $\Pi$  besitzt genau dann eine **polynomielle geglättete Komplexität**, wenn es einen randomisierten Algorithmus  $A$  für  $\Pi$  gibt, für den es zwei Konstanten  $\alpha > 0$  und*

$\beta > 0$  gibt, so dass für die Laufzeit  $T_A$  von  $A$  gilt

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \max_{I \in \mathcal{I}_N} \mathbf{E} [T_A(\mathcal{H}(I, \phi))^\alpha] \leq \beta \phi N,$$

wobei der Erwartungswert über die zufällige Eingabe und die zufälligen Entscheidungen des Algorithmus gebildet wird.

Diese Definition von polynomieller geglätteter Komplexität ist analog zur Definition von polynomieller durchschnittlicher Komplexität (siehe z. B. [11]), sie erscheint jedoch auf den ersten Blick wenig natürlich; man könnte vermuten, dass es sinnvoller wäre, den Exponenten  $\alpha$  außerhalb des Erwartungswertes zu platzieren. Dabei entstünde allerdings ein Komplexitätsbegriff, der nicht robust genug bezüglich einer Änderung des Maschinenmodells wäre, wie das folgende Beispiel zeigt.

**Beispiel 1.3.2 (Durchschnittliche Komplexität).** Es sei ein Problem  $\Pi$  gegeben und  $\mathcal{I}_N$  bezeichne die Menge der Probleminstanzen der Länge  $N$ . Es gelte  $|\mathcal{I}_N| = 2^N$  und es sei  $\mathcal{B}_N \subseteq \mathcal{I}_N$  eine Teilmenge der Größe  $|\mathcal{B}_N| = \lfloor 2^N(1 - 2^{-0.1N}) \rfloor$ . Der Algorithmus  $A$  habe auf allen Instanzen aus  $\mathcal{B}_N$  eine polynomielle Laufzeit und auf den Instanzen aus  $\mathcal{I}_N \setminus \mathcal{B}_N$  eine exponentielle Laufzeit von  $2^{0.09N}$ . Legt man für jede Eingabelänge eine uniforme Verteilung der Eingaben zu Grunde, so rechnet man leicht nach, dass die erwartete Laufzeit des Algorithmus polynomiell in der Eingabelänge  $N$  ist.

Wir betrachten nun denselben Algorithmus auf einem Maschinenmodell, auf dem die Ausführungszeit des Algorithmus das Quadrat der Ausführungszeit auf dem ursprünglichen Maschinenmodell ist. Auf diesem Maschinenmodell ist die Laufzeit des Algorithmus auf Eingaben aus  $\mathcal{B}_N$  nach wie vor polynomiell und auf Eingaben aus  $\mathcal{I}_N \setminus \mathcal{B}_N$  exponentiell, nämlich  $2^{0.18N}$ . Legt man hier dieselbe Verteilung der Eingaben zu Grunde, so rechnet man leicht nach, dass die erwartete Laufzeit des Algorithmus exponentiell ist.

Es kann also passieren, dass ein Algorithmus auf einem Maschinenmodell eine polynomielle erwartete Laufzeit hat und auf einem anderen eine exponentielle, auch wenn die Laufzeiten des Algorithmus auf den beiden Maschinenmodellen polynomiell verknüpft sind. Deswegen wäre eine Definition von polynomieller durchschnittlicher Komplexität basierend auf der erwarteten Laufzeit nicht robust genug. Das trifft analog auch auf polynomielle geglättete Komplexität zu.  $\diamond$

Die obige Definition 1.3.1 von polynomieller geglätteter Komplexität ist jedoch robust genug und für Maschinenmodelle, die sich gegenseitig in polynomieller Zeit simulieren können, äquivalent. Um das einzusehen und um die Definition handhabbarer zu machen, werden wir zunächst eine äquivalente Charakterisierung angeben.

**Lemma 1.3.3 (Alternative Charakterisierung polynomieller geglätteter Komplexität).** *Ein ganzzahliges lineares Optimierungsproblem  $\Pi$  besitzt genau dann eine polynomielle geglättete Komplexität, wenn es einen randomisierten Algorithmus  $A$  und ein Polynom  $P$  gibt, so dass für die Laufzeit  $T_A$  des Algorithmus  $A$  gilt*

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \forall \varepsilon \in (0, 1] : \forall I \in \mathcal{I}_N : \Pr \left[ T_A(\mathcal{H}(I, \phi)) \geq P \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \leq \varepsilon,$$

wobei die Wahrscheinlichkeit über die zufällige Eingabe und die zufälligen Entscheidungen des Algorithmus gebildet wird.

### 1.3. Geglättete Komplexität

---

*Beweis.* Um von der Definition 1.3.1 polynomieller geglätteter Komplexität zu dieser alternativen Charakterisierung zu kommen, benötigt man nur die Markoff-Ungleichung. Setzt man für die gegebenen Konstanten  $\alpha > 0$  und  $\beta > 0$

$$P\left(N, \phi, \frac{1}{\varepsilon}\right) = \left(\beta N \phi \frac{1}{\varepsilon}\right)^{1/\alpha},$$

so erhält man für alle  $N \in \mathbb{N}$ , für alle  $\phi \geq 1$ , für alle  $\varepsilon \in (0, 1]$  und für alle Eingaben  $I \in \mathcal{I}_N$

$$\begin{aligned} \Pr\left[T_A(\mathcal{H}(I, \phi)) \geq P\left(N, \phi, \frac{1}{\varepsilon}\right)\right] &= \Pr\left[T_A(\mathcal{H}(I, \phi))^\alpha \geq \beta N \phi \frac{1}{\varepsilon}\right] \\ &\leq \Pr\left[T_A(\mathcal{H}(I, \phi))^\alpha \geq \frac{1}{\varepsilon} \mathbf{E}[T_A(\mathcal{H}(I, \phi))^\alpha]\right] \\ &\leq \varepsilon. \end{aligned}$$

Wir betrachten nun die umgekehrte Richtung, sei dazu ein Polynom  $P$  mit den Eigenschaften, die in Lemma 1.3.3 gefordert sind, gegeben. Wir können o. B. d. A. davon ausgehen, dass es zwei Konstanten  $\alpha' > 0$  und  $\beta' \geq 1$  gibt, so dass gilt

$$P\left(N, \phi, \frac{1}{\varepsilon}\right) = \left(\beta' N \phi \frac{1}{\varepsilon}\right)^{1/\alpha'}.$$

Seien im Folgenden  $N \in \mathbb{N}$ ,  $\phi \geq 1$ ,  $\varepsilon \in (0, 1]$  und  $I \in \mathcal{I}_N$  beliebig. Wir setzen  $\alpha = \alpha'/2$  und erhalten

$$\begin{aligned} \Pr\left[T_A(\mathcal{H}(I, \phi)) \geq P\left(N, \phi, \frac{1}{\varepsilon}\right)\right] &= \Pr\left[T_A(\mathcal{H}(I, \phi)) \geq \left(\beta' N \phi \frac{1}{\varepsilon}\right)^{1/\alpha'}\right] \\ &= \Pr\left[\frac{T_A(\mathcal{H}(I, \phi))^{\alpha'}}{\beta' N \phi} \geq \frac{1}{\varepsilon}\right] \\ &= \Pr\left[\frac{T_A(\mathcal{H}(I, \phi))^\alpha}{\sqrt{\beta' N \phi}} \geq \frac{1}{\sqrt{\varepsilon}}\right]. \end{aligned} \quad (1.1)$$

Wir zeigen nun, dass der Erwartungswert der Zufallsvariablen

$$X = \frac{T_A(\mathcal{H}(I, \phi))^\alpha}{\sqrt{\beta' N \phi}}$$

endlich ist. Aus Gleichung (1.1) und der Voraussetzung des Lemmas folgt für jedes  $\varepsilon \in (0, 1]$

$$\Pr\left[X \geq \frac{1}{\sqrt{\varepsilon}}\right] \leq \varepsilon. \quad (1.2)$$

Wir können Aussage (1.2) umformulieren und erhalten als Ergebnis für alle  $\varepsilon \in [1, \infty)$

$$\Pr[X \geq \varepsilon] \leq \frac{1}{\varepsilon^2}. \quad (1.3)$$

Es sei nun  $Y$  eine reelle Zufallsvariable, für die in (1.3) stets Gleichheit gilt, das heißt, für jedes  $\varepsilon \in [1, \infty)$  gelte  $\Pr[Y \geq \varepsilon] = 1/\varepsilon^2$ . Dann gilt offenbar  $\mathbf{E}[X] \leq \mathbf{E}[Y]$ . Wie man leicht nachrechnet, gilt für die Dichte  $f_Y$  von  $Y$

$$f_Y(y) = \begin{cases} 0 & \text{falls } y < 1 \\ 2y^{-3} & \text{falls } y \geq 1 \end{cases}.$$

Insgesamt ergibt sich also

$$\mathbf{E}[X] \leq \mathbf{E}[Y] = \int_1^\infty 2y^{-2} dy = \left[ -\frac{2}{y} \right]_1^\infty = 2.$$

Mit  $\beta = 2\beta'$  erhält man

$$\mathbf{E}[T_A(\mathcal{H}(I, \phi))^\alpha] \leq 2\sqrt{\beta'\phi N} \leq 2\beta'\phi N = \beta\phi N.$$

Diese Aussage gilt für jede Wahl von  $N$ ,  $\phi$  und  $I$  mit  $N \in \mathbb{N}$ ,  $\phi \geq 1$  und  $I \in \mathcal{I}_N$ . Also erhalten wir, wie gewünscht,

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \max_{I \in \mathcal{I}_N} \mathbf{E}[T(\mathcal{H}(I, \phi))^\alpha] \leq \beta\phi N.$$

□

Mit Hilfe dieser alternativen Charakterisierung von polynomieller geglätteter Komplexität können wir zeigen, dass die Klasse derjenigen Probleme mit polynomieller geglätteter Komplexität robust genug ist, unseren Anforderungen gerecht zu werden.

**Korollar 1.3.4.** *Es sei ein Maschinenmodell  $\mathcal{M}$  und ein Algorithmus  $A$  gegeben, der auf dem Maschinenmodell  $\mathcal{M}$  eine polynomielle geglättete Laufzeit habe. Sei ferner  $\mathcal{M}'$  ein zweites Maschinenmodell, das in der Lage ist,  $\mathcal{M}$  zu simulieren. Dann kann der Algorithmus  $A$  auch auf  $\mathcal{M}'$  ausgeführt werden. Die Anzahl der Rechenschritte, die dafür benötigt werden, sei  $p(t)$  für ein streng monoton wachsendes Polynom  $p$ , wenn die Ausführung von  $A$  auf  $\mathcal{M}$  genau  $t$  viele Rechenschritte benötigt. Dann besitzt der Algorithmus  $A$  auch auf dem Maschinenmodell  $\mathcal{M}'$  eine polynomielle geglättete Laufzeit.*

*Beweis.* Dieses Korollar ergibt sich sofort aus der Charakterisierung von polynomieller geglätteter Komplexität, die in Lemma 1.3.3 gegeben ist. Für eine Eingabe  $I$  bezeichne  $T_A^{\mathcal{M}}(I)$  die Anzahl der Rechenschritte von  $A$  auf der Eingabe  $I$  auf dem Maschinenmodell  $\mathcal{M}$ .  $T_A^{\mathcal{M}'}(I)$  sei entsprechend definiert. Dann gilt für jede Eingabe  $I$  laut Voraussetzung  $T_A^{\mathcal{M}'}(I) = p(T_A^{\mathcal{M}}(I))$ .

Weil  $A$  auf  $\mathcal{M}$  eine polynomielle geglättete Laufzeit besitzt, gibt es ein Polynom  $P$ , so dass gilt

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \forall \varepsilon \in (0, 1] : \forall I \in \mathcal{I}_N : \Pr \left[ T_A^{\mathcal{M}}(\mathcal{H}(I, \phi)) \geq P \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \leq \varepsilon.$$

Setzen wir  $P' = p \circ P$ , so erhalten wir ein Polynom, für das gilt

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \forall \varepsilon \in (0, 1] : \forall I \in \mathcal{I}_N : \Pr \left[ T_A^{\mathcal{M}'}(\mathcal{H}(I, \phi)) \geq P' \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \leq \varepsilon.$$

Damit ist gezeigt, dass der Algorithmus  $A$  auch auf dem Maschinenmodell  $\mathcal{M}'$  eine polynomielle geglättete Laufzeit besitzt. □

Bevor wir das bereits angesprochene Theorem über die geglättete Komplexität ganzzahliger linearer Optimierungsprobleme formulieren, erinnern wir an die Definition der Komplexitätsklasse ZPP.



**Definition 1.3.5 (ZPP).** *Ein Optimierungsproblem  $\Pi$  liegt genau dann in ZPP, wenn es einen effizienten Las-Vegas-Algorithmus für  $\Pi$  gibt, also einen randomisierten Algorithmus, der immer das richtige Ergebnis berechnet und dessen erwartete Laufzeit für jede Eingabe polynomiell ist.*

Alternativ kann man die Klasse ZPP auch wie folgt beschreiben.

**Lemma 1.3.6 (ZPP, alternativ).** *Ein Optimierungsproblem  $\Pi$  liegt genau dann in ZPP, wenn es einen randomisierten Algorithmus für  $\Pi$  gibt, der nie ein falsches Ergebnis ausgibt, der aber für jede Eingabe mit einer Wahrscheinlichkeit von höchstens  $1/2$  kein Ergebnis ausgeben darf und der auf allen Eingaben und für jede Wahl der Zufallsbits eine polynomielle Laufzeit hat.*

Sei ein ganzzahliges lineares Optimierungsproblem  $\Pi$  und ein Perturbationsmodell  $\mathcal{H}$  für  $\Pi$  gegeben, dann bezeichne  $\Pi_u^{\mathcal{H}}$  das ganzzahlige lineare Optimierungsproblem, das aus  $\Pi$  entsteht, wenn die Koeffizienten in den als stochastisch ausgewählten Komponenten unär kodiert werden und die Menge der Eingaben dahingehend eingeschränkt wird, dass diese Koeffizienten ganzzahlig sein müssen. Im Gegensatz zu  $\Pi$  wird bei Eingaben für  $\Pi_u^{\mathcal{H}}$  die Länge der unären Kodierungen der Koeffizienten zur Eingabelänge gezählt.

**Theorem 1.3.7 (Die geglättete Komplexität ganzzahliger linearer Optimierungsprobleme).** *Es sei  $\Pi$  ein ganzzahliges lineares Optimierungsproblem und  $\mathcal{H}$  ein Perturbationsmodell. Für jede Probleminstanz  $I \in \mathcal{I}$  seien die Größe des Wertebereichs der ganzzahligen Variablen  $|\mathcal{D}(I)|$  und der Wert  $m_{\max}(I) = \max\{|a| \mid a \in \mathcal{D}(I)\}$  polynomiell in der Anzahl der Variablen  $n(I)$  beschränkt und  $0^{n(I)}$  gehöre nie zur Menge der erlaubten Lösungen. Dann besitzt  $\Pi$  genau dann eine polynomielle geglättete Komplexität bezüglich des Perturbationsmodells  $\mathcal{H}$ , wenn  $\Pi_u^{\mathcal{H}} \in \text{ZPP}$  gilt.*

Wir beweisen dieses Theorem im dritten Kapitel und greifen dabei auf strukturelle Eigenschaften von ganzzahligen linearen Optimierungsproblemen zurück, die wir im zweiten Kapitel untersuchen. Es bezeichne  $\Pi'$  das Problem, das aus  $\Pi$  hervorgeht, wenn man nur noch solche Eingaben zulässt, in denen die Koeffizienten in den stochastischen Komponenten ganze Zahlen sind. Dann ist die Aussage  $\Pi_u^{\mathcal{H}} \in \text{ZPP}$  gleich bedeutend damit, dass es einen Algorithmus für  $\Pi'$  gibt, dessen erwartete Laufzeit pseudopolynomiell ist bezogen auf die Größe der Zahlen in den stochastischen Komponenten.

## 1.4 Überblick

An dieser Stelle geben wir einen Überblick über die Analyse, die in den folgenden Kapiteln durchgeführt wird. Alle dabei erwähnten Optimierungsprobleme seien o. B. d. A. Maximierungsprobleme, es sei  $\mathcal{D} \subseteq \mathbb{Z}$  der Wertebereich der ganzzahligen Variablen,  $m = |\mathcal{D}|$ ,  $m_{\max} = \max\{|a| \mid a \in \mathcal{D}\}$  und  $k$  bezeichne die Anzahl der stochastischen Komponenten.

Im zweiten Kapitel dieser Arbeit werden wir für ein ganzzahliges lineares Optimierungsproblem die drei Zufallsvariablen *Gewinner-*, *Verlierer* und *Gültigkeits-Gap* definieren und untersuchen. Das Gewinner-Gap  $\Delta$  ist relevant, wenn die Zielfunktion stochastisch ist, denn

es gibt an, um wie viel besser die beste erlaubte Lösung  $x^*$  im Vergleich zur zweitbesten erlaubten Lösung  $x^{**}$  ist, das heißt, es entspricht der Differenz  $c \cdot x^* - c \cdot x^{**}$ , wobei  $c$  den Vektor bezeichne, dessen Komponenten die Koeffizienten der Zielfunktion sind. Wir werden zeigen, dass das Gewinner-Gap, das bei einer  $\phi$ -Perturbation entsteht, mit großer Wahrscheinlichkeit eine Größe hat, die polynomiell in  $(mn\phi)^{-1}$  ist. Dieses Ergebnis ist überraschend, denn ist beispielsweise  $\mathcal{S} = \{0, \dots, m-1\}^n$  und liegt jeder stochastische Koeffizient der linearen Zielfunktion im Intervall  $[0, 1]$ , so hat jede der  $m^n$  vielen Lösungen einen Wert aus dem Intervall  $[0, mn]$ , es werden also exponentiell viele Lösungen auf einen polynomiell großen Bereich verteilt. Deswegen könnte man vermuten, dass das Gewinner-Gap mit großer Wahrscheinlichkeit exponentiell klein wird.

Ein ähnliches Ergebnis für den Fall, dass die Koeffizienten der Zielfunktion gemäß einer diskreten Wahrscheinlichkeitsverteilung gewählt werden, haben bereits Mulmuley, Vazirani und Vazirani erzielt [7]. Sie haben gezeigt, dass die optimale Lösung eines binären Optimierungsproblems mit einer linearen Zielfunktion und  $n$  Variablen mit Wahrscheinlichkeit  $1/2$  eindeutig ist, wenn jeder Koeffizient der Zielfunktion unabhängig von den anderen uniform zufällig aus der Menge  $\{1, 2, \dots, 2n\}$  gewählt wird. Auch dieses Ergebnis entspricht nicht der ersten Intuition, es ist in der Literatur als *Isolating Lemma* bekannt.

Verlierer- und Gültigkeits-Gap werden wir zunächst für den Fall betrachten, dass wir ein ganzzahliges lineares Optimierungsproblem mit genau einer stochastischen Nebenbedingung vorliegen haben. Die Nebenbedingung habe o. B. d. A. die Form  $w \cdot x = w_1x_1 + \dots + w_nx_n \leq t$ , wobei  $w_1, \dots, w_n$  die stochastischen Koeffizienten und  $x_1, \dots, x_n$  die ganzzahligen Variablen seien.  $t$  sei eine fest vorgegebene Schranke, die nicht perturbiert wird. Die Menge  $\mathcal{S}$  der erlaubten Lösungen ergibt sich in diesem Falle als Schnitt einer fest vorgegebenen Menge  $\mathcal{S}' \subseteq \mathcal{D}^n$  mit dem Halbraum  $\mathcal{B}$ , der durch die lineare Nebenbedingung beschrieben wird. Für eine Lösung  $x \in \mathcal{S}'$  bezeichnen wir  $w \cdot x$  auch als *Gewicht* der Lösung. Das Gültigkeits-Gap gibt an, wie weit das Gewicht des *Gewinners*, also das Gewicht der besten erlaubten Lösung, unterhalb der Schranke  $t$  liegt. Eine Lösung aus  $\mathcal{S}'$  heißt *Verlierer*, wenn sie besser als der Gewinner ist, aber wegen der linearen Nebenbedingung nicht zur Menge der erlaubten Lösungen gehört, also nicht in  $\mathcal{B}$  liegt. Der *minimale Verlierer* sei derjenige Verlierer mit dem geringsten Gewicht, also der, der am nächsten an der Schranke  $t$  liegt. Das Verlierer-Gap gibt den Abstand des Gewichtes des minimalen Verlierers von der Schranke  $t$  an. Auch für diese beiden Gaps werden wir eine obere Schranke für die Wahrscheinlichkeit zeigen, dass sie eine bestimmte Größe unterschreiten. Diese Ergebnisse werden dann auf den Fall übertragen, dass es mehr als eine stochastische Nebenbedingung gibt.

Die Ergebnisse über die Gaps werden wir im dritten Kapitel ausnutzen, um Theorem 1.3.7 zu beweisen. Zu zeigen, dass ein Problem mit einer polynomiellen geglätteten Komplexität einen randomisierten pseudopolynomiellen Algorithmus besitzt, ist die einfachere Beweisrichtung. Wenn wir einen pseudopolynomiellen Algorithmus konstruieren wollen, können wir davon ausgehen, dass die Koeffizienten in den stochastischen Komponenten ganze Zahlen sind. Diese werden zunächst einer  $\phi$ -Perturbation unterworfen, wobei  $\phi$  so gewählt wird, dass die optimale Lösung durch die Perturbation mit einer konstanten Wahrscheinlichkeit nicht verändert wird. Um das zu erreichen, muss  $\phi$  linear von der größten vorkommenden Zahl  $M$  in den stochastischen Komponenten abhängen. Dann wird der Algorithmus mit polynomieller geglätteter Laufzeit auf die so erzeugte Eingabe angewendet, er überschreitet eine Rechenzeitschranke, die polynomiell in der Eingabegröße und  $M$  ist, nach Lemma 1.3.3 nur mit einer konstanten

Wahrscheinlichkeit. Ändert die  $\phi$ -Perturbation das Optimum nicht und überschreitet der Algorithmus die Rechenzeitschranke nicht, so wird das Optimum in pseudopolynomieller Zeit berechnet. Wird jedoch die optimale Lösung verändert oder die Rechenzeitschranke überschritten, so wird der Algorithmus abgebrochen und kein Ergebnis ausgegeben.

Bei der anderen Beweisrichtung werden die Ergebnisse über die Gaps benötigt. Wir gehen von der Existenz eines randomisierten pseudopolynomiellen Algorithmus aus und wollen diesen zur Konstruktion eines Algorithmus mit polynomieller geglätteter Laufzeit nutzen. Die wesentliche Idee besteht darin, die stochastischen Koeffizienten geeignet zu runden und dann den pseudopolynomiellen Algorithmus anzuwenden. Das Runden der Koeffizienten wird *adaptiv* geschehen. Zunächst wird nach dem  $b$ -ten Nachkommabit gerundet, wobei wir mit  $b = 1$  beginnen. Nach dem Runden werden die linearen Komponenten dann so skaliert, dass alle darin vorkommenden Zahlen ganz sind, damit der pseudopolynomielle Algorithmus angewendet werden kann. Es muss anschließend getestet werden, ob durch das Runden die optimale Lösung verändert wurde. Ist die Zielfunktion stochastisch, so kann es passieren, dass durch das Runden der Koeffizienten eine suboptimale erlaubte Lösung optimal wird, gibt es stochastische Nebenbedingungen, so kann es passieren, dass eine Lösung optimal wird, die bezüglich der nicht gerundeten Koeffizienten gar nicht erlaubt ist, oder dass die optimale Lösung nach dem Runden nicht mehr gültig ist. Wir werden drei Algorithmen, sogenannte *Zertifizierer*, angeben, die überprüfen, ob das Runden der Koeffizienten der Zielfunktion bzw. der Nebenbedingungen die optimale Lösung verändert hat. Falls die optimale Lösung verändert wurde, wird die Anzahl der gelesenen Bits  $b$  bei jedem Koeffizienten um eins erhöht und das Verfahren mit den genauer gerundeten Koeffizienten wiederholt.

Um zu zeigen, dass bereits für Werte von  $b$ , die in der Größenordnung  $\log(kmn\phi \cdot m_{\max})$  liegen, mit ausreichender Wahrscheinlichkeit keine Änderung der optimalen Lösung eintritt, sind die Eigenschaften der Gaps, die wir im zweiten Kapitel beweisen, essentiell. Wir betrachten zunächst den Fall, dass die Zielfunktion stochastisch ist. Da wir jeden stochastischen Koeffizienten der Zielfunktion nach dem  $b$ -ten Bit runden, ändern wir den Wert jedes Koeffizienten um höchstens  $2^{-b}$ . Es gibt  $n$  viele dieser Koeffizienten und in einer Lösung  $x \in \mathcal{D}^n$  wird jeder Koeffizient höchstens mit  $m_{\max}$  gewichtet. Also wird durch das Runden der Wert jeder Lösung um höchstens  $nm_{\max}2^{-b}$  geändert. Ist das Gewinner-Gap groß genug, das heißt größer als  $2nm_{\max}2^{-b}$ , so kann das Runden der Koeffizienten folgerichtig die optimale Lösung nicht verändern. An dieser Stelle können wir also die Schranke für die Wahrscheinlichkeit, dass das Gewinner-Gap kleiner als  $2nm_{\max}2^{-b}$  wird, die wir im zweiten Kapitel gezeigt haben, ausnutzen.

Auch bei stochastischen Nebenbedingungen können wir analog argumentieren, dass das Gewicht jeder Lösung durch das Runden um höchstens  $nm_{\max}2^{-b}$  und jede Schranke um höchstens  $2^{-b}$  verändert wird. Es kann passieren, dass entweder durch das Runden die optimale Lösung nicht mehr gültig ist oder dass eine bessere Lösung erst durch das Runden gültig wird. Das erste dieser Ereignisse kann nur dann eintreten, wenn das Gültigkeits-Gap kleiner als  $(nm_{\max} + 1)2^{-b}$  ist, und das zweite kann nur dann eintreten, wenn das Verlierer-Gap kleiner als  $(nm_{\max} + 1)2^{-b}$  ist. Also können wir auch an dieser Stelle auf die Ergebnisse aus Kapitel 2 zurückgreifen, um die Wahrscheinlichkeit dieser Ereignisse nach oben zu beschränken.

Nachdem wir die wesentlichen Beweisideen für das Theorem vorgestellt haben, werden wir nun kurz darstellen, welche der in dieser Arbeit präsentierten Ergebnisse neu sind und welche es bereits gab. Das Ergebnis über das Gewinner-Gap, das in Kapitel 2 präsentiert wird,

findet sich bereits in einer unveröffentlichten Version der Arbeit von Beier und Vöcking [2]. Die Resultate über das Verlierer- und das Gültigkeits-Gap sind im Rahmen dieser Diplomarbeit entstanden. Als Vorlage diente auch hier die Analyse des binären Falls, die sich jedoch nicht ohne Weiteres übertragen lässt, weil sie auf speziellen Eigenschaften beruht, die bei ganzzahligen Optimierungsproblemen im Allgemeinen nicht erfüllt sind. Beispielsweise gibt es für eine lineare Funktion der Form  $w_1x_1 + \dots + w_nx_n$  und für zwei verschiedene Lösungen  $x, y \in \{0, 1\}^n$  stets einen Koeffizienten  $w_i$ , von der der Funktionswert genau einer der beiden Lösungen unabhängig ist. Diese Eigenschaft ist nicht erfüllt, wenn der Wertebereich der ganzzahligen Variablen eine beliebige Teilmenge  $\mathcal{D} \subseteq \mathbb{Z}$  ist. Deswegen sind zwar gewisse strukturelle Ähnlichkeiten in der Beweisführung zu erkennen, die Ideen und benutzten Techniken gehen jedoch deutlich über eine kanonische Verallgemeinerung des binären Falls hinaus.

Für den Beweis von Theorem 1.3.7 mussten die Zertifizierer grundlegend verändert werden, denn eine weitere Eigenschaft, die bei der Analyse des binären Falls verwendet wird und die im allgemeinen Fall nicht mehr erfüllt ist, betrifft das Runden der Koeffizienten. Rundet man in einer linearen Nebenbedingung der Form  $x_1w_1 + \dots + x_nw_n \leq t$  alle Koeffizienten  $w_i$  ab, so reduziert man im binären Fall damit das Gewicht  $x_1w_1 + \dots + x_nw_n$  oder verändert es nicht. Sobald  $\mathcal{D}$  auch negative Werte enthält, kann das Gewicht jedoch durch das Abrunden der Koeffizienten auch größer werden.

# Gewinner-, Verlierer- und Gültigkeits-Gap

In diesem Kapitel werden wir für ein ganzzahliges lineares Optimierungsproblem die drei Zufallsvariablen *Gewinner-*, *Verlierer-* und *Gültigkeits-Gap* formal definieren und untersuchen; diese sind von entscheidender Bedeutung bei der Bestimmung der geglätteten Komplexität. Wie bereits in der Einleitung angesprochen, ist es wünschenswert, dass diese Gaps nicht zu klein werden, weil dann die stochastischen Koeffizienten gerundet werden können, ohne dadurch die optimale Lösung zu verändern. In diesem Kapitel werden wir deshalb für jedes der drei Gaps eine obere Schranke für die Wahrscheinlichkeit zeigen, dass es eine bestimmte Größe unterschreitet.

Im Rest dieses Kapitels gehen wir davon aus, dass  $\Pi$  ein ganzzahliges lineares Optimierungsproblem mit den Variablen  $x_1, \dots, x_n$  ist, die Werte aus einer endlichen Menge  $\mathcal{D} \subseteq \mathbb{Z}$  annehmen dürfen. Ferner gehen wir davon aus, dass es sich bei  $\Pi$  um ein Maximierungsproblem handelt. Es sei ein Perturbationsmodell  $\mathcal{H}$  für  $\Pi$  gegeben und es sei  $\mathcal{D} = \{d_1, \dots, d_m\}$  und  $m_{\max} = \max\{|a| \mid a \in \mathcal{D}\}$ .

## 2.1 Vorbereitungen

Bei den stochastischen Koeffizienten der Zielfunktion und der Nebenbedingungen handelt es sich um stetige reelle Zufallsvariablen, deren Verteilungen durch beschränkte Dichten beschrieben werden können. Wie das folgende Lemma zeigt, ist diese Klasse abgeschlossen gegen Bildung von Linearkombinationen.

**Lemma 2.1.1.** *Seien  $X_1, \dots, X_n$  stetige reelle Zufallsvariablen, die durch beschränkte Dichten beschrieben werden können, und seien  $a, a_1, \dots, a_n \in \mathbb{R}$  Konstanten mit  $a_i \neq 0$  für alle  $i \in [n]$ , dann ist auch  $a_1X_1 + \dots + a_nX_n + a$  eine stetige reelle Zufallsvariable, die durch eine beschränkte Dichte beschrieben werden kann.*

*Beweis.* Dieses Lemma ist eine leichte Folgerung aus Lemma A.3.3. □

Die folgenden beiden Lemmata finden in diesem Kapitel an einigen Stellen Anwendung.

**Lemma 2.1.2.** *Es sei  $X$  eine reelle Zufallsvariable, deren Verteilung durch die Dichte  $f_X$  beschrieben wird. Es gelte  $f_X(x) \leq \phi$  für alle  $x \in \mathbb{R}$ . Dann gilt  $\Pr[X \in [a, b]] \leq (b - a) \cdot \phi$  für jedes Intervall  $[a, b]$ . Gilt stets  $X \geq 0$ , so gilt  $\Pr[X < \varepsilon] \leq \varepsilon \cdot \phi$  für jedes  $\varepsilon \geq 0$ . □*

**Lemma 2.1.3.** *Es sei  $(\Omega, \mathcal{A}, P)$  ein beliebiger Wahrscheinlichkeitsraum und  $X, X_1, \dots, X_n$  seien reelle Zufallsvariablen. Ferner sei  $B \in \mathcal{A}$  ein beliebiges Ereignis und für jedes  $\omega \in B$  gelte  $X(\omega) \in \{X_1(\omega), \dots, X_n(\omega)\}$ . Dann gilt für jedes Intervall  $[a, b]$*

$$\Pr[X \in [a, b]|B] \leq \sum_{i=1}^n \Pr[X_i \in [a, b]|B].$$

*Diese Aussage gilt insbesondere im unbedingten Fall, also für  $B = \Omega$ .*

*Beweis.* Es gilt

$$\Pr[X \in [a, b]|B] \leq \Pr[\exists i \in [n] : X_i \in [a, b]|B] \leq \sum_{i=1}^n \Pr[X_i \in [a, b]|B].$$

□

## 2.2 Analyse des Gewinner-Gaps

In diesem Abschnitt gehen wir davon aus, dass die Menge  $\mathcal{S} \subseteq D^n$  der erlaubten Lösungen von  $\Pi$  beliebig mit  $|\mathcal{S}| \geq 2$  festgelegt ist und dass die Zielfunktion  $p : \mathcal{S} \rightarrow \mathbb{R}$  stochastisch und linear ist. Für  $x \in \mathcal{S}$  gelte  $p(x) = c \cdot x$ , wobei  $c = (c_1, \dots, c_n)$  den Vektor bezeichne, dessen Komponenten die stochastischen Koeffizienten der Zielfunktion sind. Es seien  $f_1, \dots, f_n$  beschränkte Dichten der Koeffizienten  $c_1, \dots, c_n$ , dabei sei  $f_i$  durch  $\phi_i$  beschränkt und es gelte  $\phi = \max\{\phi_i | i \in [n]\}$ .

Wir werden unter diesen Voraussetzungen eine obere Schranke für die Wahrscheinlichkeit herleiten, dass das *Gewinner-Gap*  $\Delta$  eine bestimmte Schranke  $\varepsilon \geq 0$  unterschreitet. Da wir diese Aussage für jede feste Wahl von  $\mathcal{S}$  mit  $|\mathcal{S}| \geq 2$  zeigen, gilt sie insbesondere für jede zufällige Wahl mit dieser Eigenschaft.

**Definition 2.2.1 (Gewinner-Gap).** *Es sei  $x^* = \operatorname{argmax}\{c \cdot x | x \in \mathcal{S}\}$  die beste erlaubte Lösung, diese nennen wir auch den **Gewinner**. Weiter sei  $x^{**} = \operatorname{argmax}\{c \cdot x | x \in \mathcal{S} \setminus \{x^*\}\}$  die zweitbeste erlaubte Lösung. Das **Gewinner-Gap**  $\Delta$  gibt an, um wie viel besser der Gewinner gegenüber der zweitbesten Lösung ist, also  $\Delta = c \cdot x^* - c \cdot x^{**}$ .*

Wir werden nun zeigen, dass der Gewinner mit hoher Wahrscheinlichkeit einen Abstand zur zweitbesten Lösung hat, der polynomiell in  $(mn\phi)^{-1}$  ist.

**Theorem 2.2.2 (Größe des Gewinner-Gaps).** *Für jedes  $\varepsilon \geq 0$  gilt*

$$\Pr[\Delta < \varepsilon] \leq \varepsilon \cdot (3m - 4) \sum_{i \in [n]} \phi_i \leq \varepsilon \cdot (3m - 4)n\phi.$$

*Beweis.* Falls es eine Stelle  $i \in [n]$  gibt, in der alle erlaubten Lösungen den gleichen Wert annehmen, so beeinflusst die Zufallsvariable  $c_i$  das Gewinner-Gap nicht und kann aus dem Problem entfernt werden, weswegen wir o. B. d. A. davon ausgehen, dass es für jedes  $i \in [n]$  mindestens zwei erlaubte Lösungen gibt, die sich in der  $i$ -ten Stelle unterscheiden.

## 2.2. Analyse des Gewinner-Gaps

---

Für jedes  $i \in [n]$  definieren wir das Gewinner-Gap  $\Delta_i$  bezüglich Position  $i$  als den Abstand, den der Gewinner von der besten Lösung aus  $\mathcal{S}$  hat, die sich in der  $i$ -ten Stelle von ihm unterscheidet, also  $\Delta_i = c \cdot x^* - x \cdot y$ , wobei  $y = \operatorname{argmax}\{c \cdot x \mid x \in \mathcal{S}, x_i \neq x_i^*\}$  ist.

Wir möchten Lemma 2.1.3 anwenden, dafür müssen wir zeigen, dass das Gewinner-Gap  $\Delta$  stets einen der Werte der Zufallsvariablen  $\Delta_i$  annimmt und dass die  $\Delta_i$  mit einer entsprechenden Wahrscheinlichkeit nicht zu klein werden. Die erste dieser beiden Aussagen ist leicht einzusehen; ist nämlich  $x^*$  der Gewinner und  $x^{**}$  die zweitbeste Lösung, so gibt es eine Position  $i \in [n]$ , in der sich  $x^*$  und  $x^{**}$  unterscheiden. Für diese Position  $i \in [n]$  gilt dann offensichtlich  $\Delta = \Delta_i$ , denn es ist  $x^{**} = \operatorname{argmax}\{c \cdot x \mid x \in \mathcal{S}, x_i \neq x_i^*\}$ .

Wir wählen nun ein  $i \in [n]$  fest und betrachten die Zufallsvariable  $\Delta_i$ . Dazu partitionieren wir die Menge der erlaubten Lösungen  $\mathcal{S}$  zunächst in  $m$  disjunkte Klassen  $\mathcal{S}_1, \dots, \mathcal{S}_m$ , wobei für  $j \in [m]$  die Klasse  $\mathcal{S}_j$  durch  $\mathcal{S}_j = \{x \in \mathcal{S} \mid x_i = d_j\}$  definiert ist. Wir gehen davon aus, dass alle  $c_j$  mit  $j \neq i$  bereits beliebig festgelegt wurden. Unter diesen Voraussetzungen ändert der Wert von  $c_i$  die Reihenfolge der Elemente innerhalb einer Klasse nicht mehr, denn  $c_i$  geht in alle Lösungen aus derselben Klasse in der gleichen Weise ein. Wir können also unabhängig von  $c_i$  in jeder Klasse  $\mathcal{S}_j$  eine beste Lösung  $x^{(j)}$  bestimmen.

Nach dieser Vorarbeit können wir das Gewinner-Gap  $\Delta_i$  bezüglich Position  $i$  auch schreiben als  $\Delta_i = c \cdot x^{(j^*)} - c \cdot x^{(k^*)}$ , wobei  $j^*$  eine Zufallsvariable ist, die den Index der Teilmenge  $\mathcal{S}_{j^*}$  angibt, die den Gewinner enthält, und  $k^*$  eine Zufallsvariable ist, die den Index der Teilmenge  $\mathcal{S}_{k^*}$  angibt, die die beste Lösung enthält, die sich von  $x^*$  in der  $i$ -ten Stelle unterscheidet. Es gilt also  $j^* = \operatorname{argmax}\{c \cdot x^{(j)} \mid j \in [m]\}$  und  $k^* = \operatorname{argmax}\{c \cdot x^{(k)} \mid k \in [m] \setminus \{j^*\}\}$ .

Die Werte aller Zufallsvariablen  $c_j$  mit  $j \neq i$  stehen bereits fest, das bedeutet, dass die Zufallsvariablen  $j^*$  und  $k^*$  nur noch von  $c_i$  abhängen. Steht der Wert von  $c_i$  fest, so steht also auch fest, aus welcher Teilmenge  $\mathcal{S}_{j^*}$  der Gewinner kommt und aus welcher Teilmenge  $\mathcal{S}_{k^*}$  die beste Lösung stammt, die sich vom Gewinner in der  $i$ -ten Stelle unterscheidet. Wir untersuchen nun die Frage, wie viele verschiedene Werte die Zufallsvariable  $(j^*, k^*)$  annehmen kann.

Eine triviale obere Schranke für diese Anzahl ist  $m(m-1)$ , diese können wir jedoch essentiell verbessern. Für jedes  $j \in [m]$  bezeichne  $s_j$  den Wert der Lösung  $x^{(j)}$  unter der Bedingung, dass  $c_i = 0$  gilt, also  $s_j = \sum_{k \in [m] \setminus \{i\}} x_k^{(j)} c_k$ . Es gilt dann

$$p(x^{(j)}) = c \cdot x^{(j)} = s_j + x_i^{(j)} c_i = s_j + d_j c_i,$$

wobei  $s_j \in \mathbb{R}$  und  $d_j \in \mathbb{Z}$  ist. Also kann für jedes  $j \in [m]$  der Wert  $p(x^{(j)})$  als lineare Funktion in  $c_i$  aufgefasst werden. Das folgende Lemma zeigt, dass unter diesen Voraussetzungen die Zufallsvariable  $(j^*, k^*)$  nur höchstens  $3m-4$  viele verschiedene Werte annehmen kann.

**Lemma 2.2.3.** *Gegeben seien  $m$  Objekte, die sich im eindimensionalen Raum  $\mathbb{R}$  bewegen. Das  $i$ -te Objekt habe eine konstante Geschwindigkeit  $d_i \in \mathbb{Z}$  und befinde sich zum Zeitpunkt 0 an der Position  $s_i \in \mathbb{R}$ . Zum Zeitpunkt  $t \in \mathbb{R}$  befindet sich Objekt  $i$  also im Punkt  $s_i + d_i t$ . Dann gibt es höchstens  $3m-4$  viele Paare  $(i, j) \in [m]^2$ , so dass zu einem Zeitpunkt  $t \in \mathbb{R}$  das  $i$ -te Objekt erster ist, sich also an der größten Position befindet, und das  $j$ -te Objekt zweiter ist.*

*Beweis.* Zum Beweis benutzen wir ein Argument, das auf einer Potenzialfunktion beruht. Dazu sortieren wir die Objekte zunächst aufsteigend nach ihren Geschwindigkeiten, es soll

also o. B. d. A.  $d_1 \leq \dots \leq d_m$  gelten. Sei  $j$  der Index des führenden Objektes und  $k$  der Index des zweitbesten, so sei das Potenzial definiert als  $2j + k$ . Auf diese Weise ist das Potenzial mindestens  $2 \cdot 1 + 2 = 4$  und höchstens  $2m + m - 1 = 3m - 1$ . Jedes Mal, wenn ein Führungswechsel stattfindet, wenn sich also das geordnete Paar der führenden Objekte verändert, erhöht sich das Potenzial um mindestens 1, weil nur schnellere Objekte langsamere überholen können. Also gibt es höchstens  $3m - 4$  viele verschiedene Paare, die zu einem Zeitpunkt  $t \in \mathbb{R}$  die führenden Objekte repräsentieren.  $\square$

*Fortsetzung des Beweises von Theorem 2.2.2.* Wir können nun das Ergebnis von Lemma 2.2.3 benutzen, um den Beweis abzuschließen. Wir wissen, dass es höchstens  $3m - 4$  viele Paare  $(j, k)$  von Indizes gibt, für die  $(j, k) = (j^*, k^*)$  gelten kann. Sei  $M \subseteq [m]^2$  die Menge dieser Paare von Indizes, dann gibt es stets ein Paar  $(j, k) \in M$ , so dass

$$\Delta_i = c \cdot x^{(j)} - c \cdot x^{(k)} = s_j + d_j \cdot c_i - (s_k + d_k \cdot c_i) = (s_j - s_k) + (d_j - d_k) \cdot c_i$$

gilt. Dabei ist  $(s_j - s_k) \in \mathbb{R}$  unabhängig von  $c_i$  und bereits fest gewählt. Ferner ist  $(d_j - d_k) \in \mathbb{Z} \setminus \{0\}$ . Also ist der Term  $(s_j - s_k) + (d_j - d_k) \cdot c_i$  eine Zufallsvariable mit einer Dichte von höchstens  $\phi_i / |d_j - d_k| \leq \phi_i$ .  $\Delta_i$  nimmt den Wert eines dieser höchstens  $3m - 4$  vielen Terme an. Mit Lemma 2.1.2 und Lemma 2.1.3 folgt somit

$$\begin{aligned} \Pr [\Delta_i \leq \varepsilon] &= \Pr [\Delta_i \in [0, \varepsilon]] \\ &\leq \sum_{(j,k) \in M} \Pr [(s_j - s_k) + (d_j - d_k) \cdot c_i \in [0, \varepsilon]] \\ &\leq \sum_{(j,k) \in M} \varepsilon \cdot \phi_i \\ &\leq \varepsilon \cdot (3m - 4) \phi_i. \end{aligned}$$

Eine weitere Anwendung von Lemma 2.1.3 liefert nun das gewünschte Ergebnis

$$\begin{aligned} \Pr [\Delta \leq \varepsilon] &\leq \sum_{i \in [n]} \Pr [\Delta_i \leq \varepsilon] \\ &\leq \varepsilon \cdot (3m - 4) \sum_{i \in [n]} \phi_i \\ &\leq \varepsilon \cdot (3m - 4) n \phi. \end{aligned}$$

$\square$

## 2.3 Verlierer- und Gültigkeits-Gap im Falle einer stochastischen Nebenbedingung

Wir betrachten das ganzzahlige lineare Optimierungsproblem  $\Pi$  und gehen davon aus, dass es genau eine stochastische Nebenbedingung gibt. Die Menge der gültigen Lösungen ergibt sich dann als Schnitt von einer fest vorgegebenen Menge  $\mathcal{S} \subseteq \mathcal{D}^n$  mit einem Halbraum  $\mathcal{B}$ , der durch die stochastische lineare Nebenbedingung, die o. B. d. A. die Form  $w \cdot x \leq t$  habe, beschrieben wird. Die Koeffizienten  $w_1, \dots, w_n$  seien unabhängige stetige Zufallsvariablen, die



gemäß den Verteilungen  $F_1, \dots, F_n$  mit den beschränkten Dichten  $f_1, \dots, f_n$  verteilt seien. Die absoluten Erwartungswerte dieser Zufallsvariablen seien endlich. Für  $i \in [n]$  sei  $f_i$  durch  $\phi_i$  beschränkt und es sei  $\phi = \max\{\phi_i | i \in [n]\}$ . Die Schranke  $t \in \mathbb{R}$  kann beliebig fest gewählt werden.

Von der Zielfunktion verlangen wir lediglich, dass sie auf den Lösungen aus  $\mathcal{S}$  eine eindeutige Reihenfolge induziert, ansonsten kann sie beliebig sein. Ist die Zielfunktion stochastisch und linear, also von der Form  $c \cdot x$  mit stetigen Zufallsvariablen  $c_1, \dots, c_n$  als Koeffizienten, so ist diese Bedingung mit Wahrscheinlichkeit 1 erfüllt. Um das einzusehen, stellt man zunächst fest, dass zwei Lösungen  $x \in \mathcal{S}$  und  $x' \in \mathcal{S}$  durch  $c \cdot x = c \cdot x'$  eine Hyperebene im Raum  $\mathbb{R}^n$  definieren. Die Reihenfolge, die durch die Zielfunktion induziert wird, ist genau dann nicht eindeutig, wenn es zwei Lösungen  $x \in \mathcal{S}$  und  $x' \in \mathcal{S}$  gibt, für die  $c \cdot x = c \cdot x'$  gilt, für die also der Vektor  $c = (c_1, \dots, c_n)$  in der Hyperebene liegt, die durch  $x$  und  $x'$  definiert wird. Hyperebenen sind *Nullmengen*, die Wahrscheinlichkeit, dass  $c$  in einer gegebenen Hyperebene liegt, ist also Null, wenn jeder Koeffizient  $c_i$  gemäß einer kontinuierlichen Verteilung gewählt wird. Es gibt endlich viele solcher Hyperebenen und jede ist eine Nullmenge, deswegen ist die Wahrscheinlichkeit, dass  $c$  in einer dieser Hyperebenen liegt, Null.

**Definition 2.3.1 (Gültigkeits-Gap).** *Der Gewinner  $x^*$  sei die Lösung mit dem höchsten Rang aus  $\mathcal{S} \cap \mathcal{B}$ . Das **Gültigkeits-Gap**  $\Gamma$  beschreibt den Abstand, den der Gewinner von der Schranke  $t$  hat, also*

$$\Gamma = \begin{cases} t - w \cdot x^* & \text{falls } \mathcal{S} \cap \mathcal{B} \neq \emptyset \\ \perp & \text{sonst} \end{cases} . \quad (2.1)$$

Dabei ist zu beachten, dass durchaus der Fall eintreten kann, dass es keine erlaubte Lösung gibt. In diesem Fall ist das Gültigkeits-Gap undefiniert. Wir werden darauf im nächsten Kapitel zurückkommen, für die Ergebnisse dieses Kapitels ist dieser Fall jedoch nicht von Interesse.

**Definition 2.3.2 (Verlierer-Gap).** *Wir nennen eine Lösung aus  $\mathcal{S}$  einen **Verlierer**, wenn sie einen höheren Rang als der Gewinner  $x^*$  hat, aber wegen der stochastischen Nebenbedingung nicht zur Menge der erlaubten Lösungen zählt. Die Menge der Verlierer wird mit  $\mathcal{L}$  bezeichnet. Falls es wegen  $\mathcal{S} \cap \mathcal{B} = \emptyset$  keinen Gewinner gibt, so definieren wir  $\mathcal{L} = \mathcal{S}$ . Der Verlierer  $x^{\min}$  mit dem kleinsten Gewicht heißt **minimaler Verlierer**, es gilt also*

$$x^{\min} = \begin{cases} \operatorname{argmin}\{w \cdot x | x \in \mathcal{L}\} & \text{falls } \mathcal{L} \neq \emptyset \\ \perp & \text{sonst} \end{cases} .$$

*Gibt es mehrere minimale Verlierer mit dem gleichen Gewicht  $w \cdot x$ , so sei  $x^{\min}$  derjenige mit dem höchsten Rang. Das **Verlierer-Gap**  $\Lambda$  beschreibt den Abstand des minimalen Verlierers von der Schranke  $t$ , also*

$$\Lambda = \begin{cases} w \cdot x^{\min} - t & \text{falls } \mathcal{L} \neq \emptyset \\ \perp & \text{sonst} \end{cases} . \quad (2.2)$$

Auch hier kann der Fall eintreten, dass das Verlierer-Gap undefiniert ist, weil die Menge der Verlierer leer ist, aber auch auf dieses Ereignis wird erst im nächsten Kapitel eingegangen. Ist im Folgenden von  $\Pr[\Lambda < \varepsilon]$  die Rede, so ist damit  $\Pr[\Lambda \neq \perp \wedge \Lambda \in (-\infty, \varepsilon)]$  gemeint.

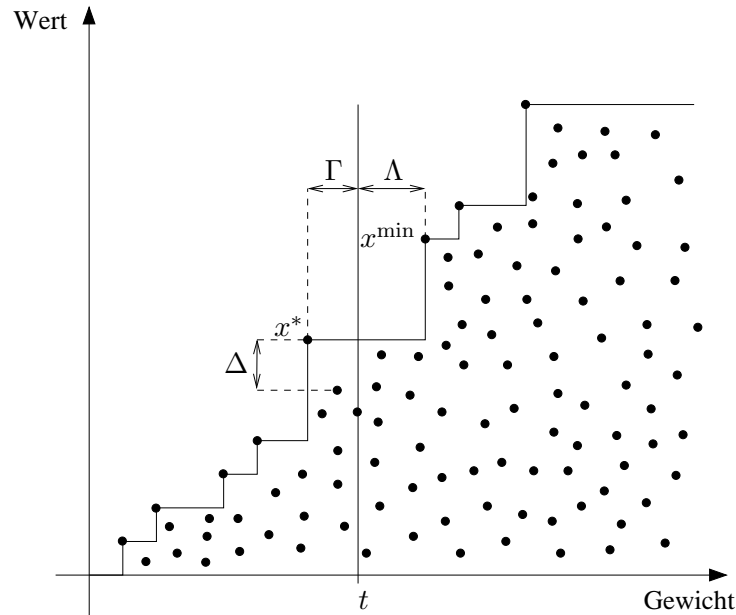


Abbildung 2.1: In dieser Abbildung werden die Begriffe Gewinner, minimaler Verlierer, Gewinner-, Verlierer- und Gültigkeits-Gap veranschaulicht. Die Punkte entsprechen den Lösungen aus  $\mathcal{S}$ . Die Punkte auf der „Treppe“ sind genau die pareto-optimalen Lösungen, sie dominieren alle Punkte unterhalb der „Treppe“.

Gleiches gilt auch für das Gültigkeits-Gap  $\Gamma$ . Die vorangegangenen Definitionen werden in Abbildung 2.1 noch einmal graphisch veranschaulicht.

Die Aufgabe ist es nun, zu zeigen, dass das Verlierer- und das Gültigkeits-Gap mit einer hinreichend großen Wahrscheinlichkeit nicht zu klein werden. Wir müssen im Fall  $0 \in \mathcal{D}$  jedoch zunächst bei der Analyse ausschließen, dass  $0^n \in \mathcal{S}$  gilt, denn diese Lösung unterscheidet sich wesentlich von den anderen, weil ihre Gültigkeit unabhängig von den Koeffizienten  $w_1, \dots, w_n$  ist. Nehmen wir an,  $0^n$  habe den höchsten Rang von allen Lösungen aus  $\mathcal{S}$ . Dann können wir durch  $t = 0$  erreichen, dass  $\Gamma = 0$  gilt. Ebenso können wir durch  $t \rightarrow 0$  und  $t < 0$  auch  $\Lambda \rightarrow 0$  erreichen.

Eine obere Schranke für die Wahrscheinlichkeit, dass das Verlierer- oder das Gültigkeits-Gap zu klein werden, werden wir in mehreren Schritten herleiten. Im ersten Schritt werden wir annehmen, dass die Träger der Dichten  $f_1, \dots, f_n$  beschränkt sind, dass es also ein  $s \in \mathbb{R}_{\geq 0}$  gibt, so dass  $f_i(x) = 0$  für alle  $x \notin [-s, s]$  und alle  $i \in [n]$  gilt. Wir werden die behandelte Situation zunächst noch weiter dadurch einschränken, dass wir davon ausgehen, dass je zwei Elemente aus  $\mathcal{S}$  *linear unabhängig* sind, wobei wir die folgende Definition von linear unabhängig zu Grunde legen.

**Definition 2.3.3 (Linear unabhängig).** Zwei Vektoren  $x, y \in \mathcal{S} \subseteq \mathcal{D}^n$  heißen *linear abhängig*, falls es ein  $\lambda \in \mathbb{R}$  gibt, so dass  $x_i = \lambda y_i$  für alle  $i \in [n]$  gilt, oder falls es ein  $\lambda \in \mathbb{R}$  gibt, so dass  $y_i = \lambda x_i$  für alle  $i \in [n]$  gilt. Sind  $x$  und  $y$  nicht linear abhängig, so nennen wir sie *linear unabhängig*.

In diesem Spezialfall werden wir zeigen, dass das Verlierer-Gap mit großer Wahrscheinlichkeit

nicht zu klein wird. Wir werden dann zeigen, dass sich Schranken, die für das Verlierer-Gap gelten, auf das Gültigkeits-Gap übertragen lassen und umgekehrt. Beim Gültigkeits-Gap werden wir die Einschränkung auf Mengen  $\mathcal{S}$  ohne linear abhängige Elemente auf Kosten einer etwas schlechteren Schranke fallen lassen können. Diese Schranke überträgt sich dann wiederum auf das Verlierer-Gap. Im letzten Schritt werden wir dann die Schranken auf den Fall allgemeiner Dichten mit nicht mehr zwangsweise beschränkten Trägern verallgemeinern.

In den folgenden Abschnitten wird die Kenntnis der Begriffe, Lemmata und Sätze über kontinuierliche Wahrscheinlichkeitstheorie, die in Anhang A vorgestellt werden, vorausgesetzt.

### 2.3.1 Verlierer- und Gültigkeits-Gap für Dichten mit beschränkten Trägern

Wir werden in diesem Abschnitt eine obere Schranke für die Wahrscheinlichkeit zeigen, dass das Verlierer-Gap kleiner als  $\varepsilon$  wird. Dabei gehen wir davon aus, dass die Träger der Dichten  $f_1, \dots, f_n$  beschränkt sind und dass  $\mathcal{S}$  keine linear abhängigen Elemente enthält. Sei im Folgenden  $s \in \mathbb{R}_{\geq 0}$  so gewählt, dass  $f_i(x) = 0$  für alle  $i \in [n]$  und für alle  $x \notin [-s, s]$  gilt. Um die Eigenschaft, dass  $\mathcal{S}$  keine linear abhängigen Elemente enthält, besser nutzen zu können, werden wir von folgendem Lemma Gebrauch machen.

**Lemma 2.3.4.** *Es seien  $x, y \in \mathcal{D}^n$  zwei linear unabhängige Elemente, dann gibt es zwei Indizes  $i, j \in [n]$ , so dass die Vektoren  $(x_i, x_j) \in \mathcal{D}^2$  und  $(y_i, y_j) \in \mathcal{D}^2$  linear unabhängig sind.*

*Beweis.* Wir nehmen an, dass  $x$  und  $y$  linear unabhängig sind, dass aber trotzdem für alle Indizes  $i, j \in [n]$  die Vektoren  $(x_i, x_j)$  und  $(y_i, y_j)$  linear abhängig sind. Wäre  $x = 0^n$  oder  $y = 0^n$ , so wären  $x$  und  $y$  linear abhängig, deswegen gilt unter unserer Annahme  $x \neq 0^n$  und  $y \neq 0^n$ . Wir können also o. B. d. A davon ausgehen, dass  $x_1 \neq 0$  ist. Für das Indexpaar  $(i, j) = (1, 2)$  gibt es laut Voraussetzung ein  $\lambda \in \mathbb{R}$ , so dass  $x_1 = \lambda y_1$  und  $x_2 = \lambda y_2$  oder  $y_1 = \lambda x_1$  und  $y_2 = \lambda x_2$  gilt.

Im ersten dieser beiden Fälle muss wegen  $x_1 \neq 0$  auch  $\lambda \neq 0$  und  $y_1 \neq 0$  gelten, dann ist  $y_1 = x_1/\lambda$  und  $y_2 = x_2/\lambda$  und  $\lambda$  ist eindeutig bestimmt. Betrachtet man nun die Indexpaare  $(1, 3), (1, 4), \dots, (1, n)$ , so stellt man fest, dass  $y_i = x_i/\lambda$  für alle  $i \in [n]$  gilt. Die Vektoren  $x$  und  $y$  sind also im Widerspruch zur Voraussetzung linear abhängig. Im zweiten Fall fallen die Einschränkungen  $\lambda \neq 0$  und  $y_1 \neq 0$  weg, einen Widerspruch zur Annahme, dass  $x$  und  $y$  linear unabhängig sind, erhält man jedoch analog zum ersten Fall, denn  $\lambda$  ist auch in diesem Fall eindeutig bestimmt.  $\square$

Wir wissen laut Voraussetzung an  $\mathcal{S}$ , dass der Gewinner  $x^*$  und der minimale Verlierer  $x^{\min}$  linear unabhängig sind. Nach Lemma 2.3.4 können wir also stets zwei Indizes  $i, j \in [n]$  mit  $i < j$  finden, so dass die Vektoren  $(x_i^*, x_j^*)$  und  $(x_i^{\min}, x_j^{\min})$  linear unabhängig sind.

Um die gewünschte obere Schranke für  $\mathbf{Pr}[\Lambda < \varepsilon]$  zu zeigen, werden wir von Lemma 2.1.3 Gebrauch machen. Für jede Wahl von  $i, j \in [n]$  mit  $i < j$  und  $\bar{m} = (m_1, m_2, m_3, m_4) \in \mathcal{D}^4$  mit linear unabhängigen Vektoren  $(m_1, m_2)$  und  $(m_3, m_4)$  definieren wir eine Zufallsvariable  $\Lambda_{i,j}^{\bar{m}}$ , die das Verlierer-Gap unter der Bedingung beschreibt, dass  $(m_3, m_4) = (x_i^*, x_j^*)$  und  $(m_1, m_2) = (x_i^{\min}, x_j^{\min})$  gilt. Das Verlierer-Gap nimmt dann stets einen der Werte der  $\Lambda_{i,j}^{\bar{m}}$  an.

Wir werden die Dichten dieser Zufallsvariablen nach oben beschränken, um eine obere Schranke für die Wahrscheinlichkeit dafür zu erhalten, dass eine der Zufallsvariablen  $\Lambda_{i,j}^{\bar{m}}$  zu klein wird, dies gelingt uns jedoch nicht ohne weitere Annahmen. Unsere Analyse versagt allerdings nur dann, wenn ein bestimmtes, unwahrscheinliches Ereignis  $E(p)$  eintritt. Dieses Ereignis ist mit einem  $p \in \mathbb{R}$  mit  $p \geq 2$  parametrisiert. Je größer man  $p$  wählt, desto unwahrscheinlicher wird es, dass das Ereignis  $E(p)$  eintritt, allerdings wächst mit  $p$  auch die Schranke, die wir für die Wahrscheinlichkeit  $\Pr[\Lambda < \varepsilon | \neg E(p)]$  erhalten. Man muss also zwischen diesen beiden konkurrierenden Eigenschaften abwägen und für  $p$  den richtigen Kompromiss finden.

Wir beginnen nun damit, die oben beschriebenen Zufallsvariablen formal zu definieren.

**Definition 2.3.5.** Für jede Kombination aus  $i, j \in [n]$  und  $\bar{m} = (m_1, m_2, m_3, m_4) \in \mathcal{D}^4$  mit  $i < j$ , für die  $(m_1, m_2)$  und  $(m_3, m_4)$  linear unabhängig sind, definieren wir eine Zufallsvariable  $\Lambda_{i,j}^{\bar{m}}$ . Dazu sei zunächst  $x_{i,j}^{*,m_3,m_4}$  definiert als das Element mit dem höchsten Rang aus der Menge  $\{x \in \mathcal{S} | x_i = m_3, x_j = m_4\} \cap \mathcal{B}$ . Anschaulich gesprochen ist  $x_{i,j}^{*,m_3,m_4}$  also der **Gewinner** der Elemente, für die  $x_i = m_3$  und  $x_j = m_4$  gilt.

Nun können wir auf  $x_{i,j}^{*,m_3,m_4}$  aufbauend **Verlierermengen**  $\mathcal{L}_{i,j}^{\bar{m}}$  und **minimale Verlierer**  $x_{i,j}^{\min,\bar{m}}$  definieren. Es sei dazu

$$\mathcal{L}_{i,j}^{\bar{m}} = \{x \in \mathcal{S} \mid x_i = m_1, x_j = m_2, x \text{ hat höheren Rang als } x_{i,j}^{*,m_3,m_4}\}$$

die Menge der Verlierer bezüglich  $x_{i,j}^{*,m_3,m_4}$ , für die  $x_i = m_1$  und  $x_j = m_2$  gilt. Die minimalen Verlierer seien definiert als

$$x_{i,j}^{\min,\bar{m}} = \begin{cases} \operatorname{argmin}\{w \cdot x \mid x \in \mathcal{L}_{i,j}^{\bar{m}}\} & \text{falls } \mathcal{L}_{i,j}^{\bar{m}} \neq \emptyset \\ \perp & \text{sonst} \end{cases}.$$

Nach dieser Vorarbeit können wir die Zufallsvariable  $\Lambda_{i,j}^{\bar{m}}$  definieren als

$$\Lambda_{i,j}^{\bar{m}} = \begin{cases} w \cdot x_{i,j}^{\min,\bar{m}} - t & \text{falls } \mathcal{L}_{i,j}^{\bar{m}} \neq \emptyset \\ \perp & \text{sonst} \end{cases}.$$

Man beachte, dass bei dieser Definition auch  $\Lambda_{i,j}^{\bar{m}} < 0$  gelten kann. Wir werden jedoch im Folgenden sehen, dass die gewünschten Eigenschaften gewährleistet sind, zum einen nimmt  $\Lambda$  stets einen der Werte der  $\Lambda_{i,j}^{\bar{m}}$  an und zum anderen können wir die Dichten der  $\Lambda_{i,j}^{\bar{m}}$  unter der Bedingung  $\neg E(p)$ , die noch zu definieren ist, beschränken.

**Lemma 2.3.6.** Je zwei Elemente der Menge  $\mathcal{S} \subseteq \mathcal{D}^n$  mit  $0^n \notin \mathcal{S}$  seien linear unabhängig. Dann gibt es stets eine Kombination von  $i, j \in [n]$  und  $\bar{m} \in \mathcal{D}^4$ , so dass die Zufallsvariable  $\Lambda$  den Wert der Zufallsvariablen  $\Lambda_{i,j}^{\bar{m}}$  annimmt.

*Beweis.* Es sei  $x^*$  der Gewinner und  $x^{\min}$  der minimale Verlierer. Auf Grund der Voraussetzung an  $\mathcal{S}$  sind  $x^*$  und  $x^{\min}$  linear unabhängig und damit gibt es nach Lemma 2.3.4 zwei Indizes  $i, j \in [n]$  mit  $i < j$ , für die die Vektoren  $(m_3, m_4) := (x_i^*, x_j^*)$  und  $(m_1, m_2) := (x_i^{\min}, x_j^{\min})$  linear unabhängig sind. Setzen wir  $\bar{m} = (m_1, m_2, m_3, m_4)$ , so gilt in diesem Fall  $\Lambda = \Lambda_{i,j}^{\bar{m}}$ , denn offensichtlich ist  $x^* = x_{i,j}^{*,m_3,m_4}$  und damit gilt  $\mathcal{L}_{i,j}^{\bar{m}} = \mathcal{L} \cap \{x \in \mathcal{S} | x_i = m_1, x_j = m_2\}$ . Also ist  $x^{\min} \in \mathcal{L}_{i,j}^{\bar{m}}$  und damit  $x^{\min} = x_{i,j}^{\min,\bar{m}}$ .

### 2.3. Verlierer- und Gültigkeits-Gap im Falle einer stochastischen Nebenbedingung

Der Fall  $\Lambda = \perp$  tritt nur dann ein, wenn  $\mathcal{L} = \emptyset$  gilt, wenn es also keine Lösung gibt, die einen höheren Rang als der Gewinner hat. Es ist  $x^* \neq 0^n$ , deswegen gehen wir o. B. d. A. davon aus, dass  $(m_3, m_4) := (x_1^*, x_2^*) \neq (0, 0)$  gilt. Sei  $(m_1, m_2) \in \mathcal{D}^2$  beliebig, aber linear unabhängig von  $(m_3, m_4)$ , dann gilt für  $\bar{m} = (m_1, m_2, m_3, m_4)$  auch  $\mathcal{L}_{1,2}^{\bar{m}} = \emptyset$ , weil es keine Lösung mit einem höheren Rang als  $x_{1,2}^{*,m_3,m_4} = x^*$  gibt. Also gilt für solche  $\bar{m}$  ebenfalls  $\Lambda_{1,2}^{\bar{m}} = \perp$ .  $\square$

Nun bleibt nur noch die Frage zu klären, wie und unter welcher Bedingung wir die Dichte der Zufallsvariablen  $\Lambda_{i,j}^{\bar{m}}$  beschränken können. Dafür definieren wir zunächst als Abkürzung  $A_{i,j}^{\alpha,\beta} = \alpha w_i + \beta w_j$ . Wir gehen davon aus, dass ein Gegner alle  $w_k$  mit  $k \neq i$  und  $k \neq j$  beliebig vorgibt und dass er ebenfalls  $A_{i,j}^{m_3,m_4} = b$  beliebig vorgibt. Können wir die Dichte der  $\Lambda_{i,j}^{\bar{m}}$  für jede beliebige Wahl des Gegners beschränken, so gilt die Schranke, die wir erhalten, ebenfalls für jede zufällige Wahl der  $w_k$  und von  $b$ .

Durch die Wahlen des Gegners ist festgelegt, welche Elemente der Menge  $\{x \in \mathcal{S} | x_i = m_3, x_j = m_4\}$  gültig sind und welche ungültig sind, das heißt, welche in  $\mathcal{B}$  liegen und welche nicht. Da die Zielfunktion fest ist, steht somit der Gewinner  $x_{i,j}^{*,m_3,m_4}$  fest. Damit steht die Verlierermenge  $\mathcal{L}_{i,j}^{\bar{m}}$  und damit auch der minimale Verlierer  $x_{i,j}^{\min,\bar{m}}$  fest, weil  $w_i$  und  $w_j$  in der gleichen Weise in alle Elemente aus  $\mathcal{L}_{i,j}^{\bar{m}}$  eingehen. Falls  $\mathcal{L}_{i,j}^{\bar{m}} \neq \emptyset$  gilt, kann man das Verlierer-Gap  $\Lambda_{i,j}^{\bar{m}}$  unter diesen Vorgaben also schreiben als

$$\Lambda_{i,j}^{\bar{m}} = w \cdot x_{i,j}^{\min,\bar{m}} - t = \kappa + m_1 w_i + m_2 w_j = \kappa + A_{i,j}^{m_1,m_2}$$

für eine Konstante  $\kappa$ , die von den Wahlen des Gegners abhängt. Somit stimmen die Dichten von  $\Lambda_{i,j}^{\min,\bar{m}}$  und  $A_{i,j}^{m_1,m_2}$  unter den Vorgaben des Gegners bis auf eine konstante Verschiebung überein. Insbesondere übertragen sich obere Schranken der Dichte von  $A_{i,j}^{m_1,m_2}$  auf die Dichte von  $\Lambda_{i,j}^{\min,\bar{m}}$ . Wir müssen jedoch beachten, dass wir stets die Bedingung  $A_{i,j}^{m_3,m_4} = b$  zu Grunde liegen haben, und werden sehen, dass wir durch eine ungünstige Wahl von  $b$  die bedingte Dichte von  $A_{i,j}^{m_1,m_2}$  im Allgemeinen beliebig groß machen können.

**Lemma 2.3.7.** *Es seien Indizes  $i, j \in [n]$  mit  $i < j$  und linear unabhängige Paare  $(m_1, m_2), (m_3, m_4) \in \mathcal{D}^2$  beliebig vorgegeben. Dann gilt mit der Dichte  $f_{A_{i,j}^{m_3,m_4}}$  von  $A_{i,j}^{m_3,m_4}$  für die bedingte Dichte von  $A_{i,j}^{m_1,m_2}$*

$$f_{A_{i,j}^{m_1,m_2} | A_{i,j}^{m_3,m_4} = b}(a) \leq \frac{\phi^2}{f_{A_{i,j}^{m_3,m_4}}(b)},$$

falls  $f_{A_{i,j}^{m_3,m_4}}(b) > 0$  ist.

*Beweis.* Zunächst bestimmen wir formal die Dichte von  $A_{i,j}^{m_1,m_2}$  unter der Bedingung  $A_{i,j}^{m_3,m_4} = b$ . Mit der gemeinsamen Dichte  $f_{i,j}^{\bar{m}}$  von  $A_{i,j}^{m_1,m_2}$  und  $A_{i,j}^{m_3,m_4}$  gilt nach Definition und Satz A.5.1

$$f_{A_{i,j}^{m_1,m_2} | A_{i,j}^{m_3,m_4} = b}(a) = \frac{f_{i,j}^{\bar{m}}(a, b)}{\int_{-\infty}^{\infty} f_{i,j}^{\bar{m}}(x, b) dx} = \frac{f_{i,j}^{\bar{m}}(a, b)}{f_{A_{i,j}^{m_3,m_4}}(b)},$$

falls  $f_{A_{i,j}^{m_3,m_4}}(b) > 0$  ist. Es gilt  $(A_{i,j}^{m_1,m_2}, A_{i,j}^{m_3,m_4}) = \Phi(w_i, w_j)$  für die Transformation  $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  mit

$$\Phi : (x, y) \mapsto (m_1 x + m_2 y, m_3 x + m_4 y).$$

Da die Vektoren  $(m_1, m_2)$  und  $(m_3, m_4)$  laut Voraussetzung linear unabhängig sind, ist  $\Phi$  bijektiv, und die Umkehrabbildung  $\Phi^{-1}$  kann mit der Cramerschen Regel berechnet werden. Man erhält

$$\Phi^{-1} : (a, b) \mapsto \left( \frac{m_4 a - m_2 b}{m_1 m_4 - m_2 m_3}, \frac{m_1 b - m_3 a}{m_1 m_4 - m_2 m_3} \right).$$

Wir berechnen nun die Determinante der Jacobimatrix von  $\Phi^{-1}$ . Mit  $d = m_1 m_4 - m_2 m_3$  gilt

$$\det \begin{pmatrix} \frac{m_4}{d} & -\frac{m_2}{d} \\ -\frac{m_3}{d} & \frac{m_1}{d} \end{pmatrix} = \frac{m_1 m_4 - m_2 m_3}{d^2} = \frac{1}{d}.$$

Somit ergibt sich die gemeinsame Dichte  $f_{i,j}^{\overline{m}}$  von  $A_{i,j}^{m_1, m_2}$  und  $A_{i,j}^{m_3, m_4}$  nach Satz A.4.4 wegen der Unabhängigkeit von  $w_i$  und  $w_j$  zu

$$f_{i,j}^{\overline{m}}(a, b) = \frac{1}{|d|} \cdot f_i \left( \frac{m_4 a - m_2 b}{d} \right) f_j \left( \frac{m_1 b - m_3 a}{d} \right) \leq \frac{1}{|d|} \phi^2 \leq \phi^2. \quad (2.3)$$

Insgesamt gilt somit, wie gewünscht,

$$f_{A_{i,j}^{m_1, m_2} | A_{i,j}^{m_3, m_4} = b}(a) \leq \frac{\phi^2}{f_{A_{i,j}^{m_3, m_4}}(b)}.$$

□

Das folgende Beispiel zeigt, dass der Ausdruck  $f_{A_{i,j}^{m_3, m_4}}(b)$  im Allgemeinen beliebig klein werden kann, wenn  $b$  entsprechend gewählt wird und die Dichten  $f_i$  und  $f_j$  es zulassen.

**Beispiel 2.3.8.** Man muss kein kompliziertes Beispiel konstruieren, um zu sehen, dass die Dichte von  $A_{i,j}^{m_1, m_2}$  unter der Bedingung  $A_{i,j}^{m_3, m_4} = b$  beliebig groß werden kann. Sei  $i = 1$  und  $j = 2$  und seien  $w_1$  und  $w_2$  unabhängig und uniform zufällig im Intervall  $[0, 1]$  verteilt. Ferner gelte  $m_1 = 2$  und  $m_2 = m_3 = m_4 = 1$ . Wir möchten also die Dichte von  $2w_1 + w_2$  unter der Bedingung  $w_1 + w_2 = b$  berechnen. Aus Gleichung (2.3) erhält man durch eine leichte Rechnung die gemeinsame Dichte  $f$  von  $2w_1 + w_2$  und  $w_1 + w_2$  als

$$f(a, b) = \begin{cases} 1 & \text{falls } 0 \leq b \leq 1 \wedge b \leq a \leq 2b \\ 1 & \text{falls } 1 \leq b \leq 2 \wedge 2b - 1 \leq a \leq 1 + b \\ 0 & \text{sonst} \end{cases}.$$

Für die Dichte  $f_{w_1 + w_2}$  von  $w_1 + w_2$  gilt nach Lemma A.3.3

$$f_{w_1 + w_2}(b) = \begin{cases} b & \text{falls } b \in [0, 1] \\ 2 - b & \text{falls } b \in [1, 2] \\ 0 & \text{sonst} \end{cases}.$$

Wir betrachten, was passiert, wenn sich  $b$  von oben der Stelle 0 annähert. Für  $b > 0$  und  $b \rightarrow 0$  gilt  $f(b, b) \rightarrow 1$  und  $f_{w_1 + w_2}(b) \rightarrow 0$ . Also gilt

$$f_{2w_1 + w_2 | w_1 + w_2 = b}(b) \rightarrow \infty.$$

Somit kann die bedingte Dichte von  $2w_1 + w_2$  beliebig groß werden. Dieses Ergebnis wird in Abbildung 2.2 noch einmal anschaulich erläutert. ◇

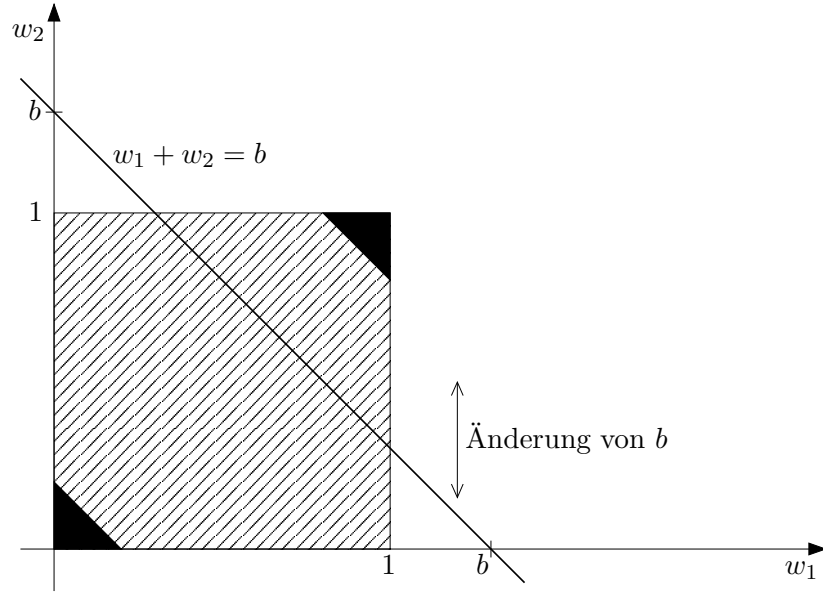


Abbildung 2.2: In dieser Abbildung kann man die gemeinsame Dichte von  $w_1$  und  $w_2$  ablesen. Sie ist innerhalb des schraffierten und der schwarz ausgefüllten Bereiche 1 und an allen anderen Stellen 0. Die Vorgabe  $w_1 + w_2 = b$  schränkt die Menge der möglichen Werte von  $w_1$  und  $w_2$  auf die eingezeichnete Gerade ein. Je kleiner die Menge der Punkte mit Dichte 1 wird, die auf der Gerade liegen, desto genauer sind die Werte von  $w_1$  und  $w_2$  bestimmt und um so größer wird die bedingte Dichte von  $2w_1 + w_2$  unter der Bedingung  $w_1 + w_2 = b$ . Die Extremfälle sind  $w_1 + w_2 = 0$  und  $w_1 + w_2 = 2$ , in diesen Fällen sind  $w_1$  und  $w_2$  komplett bestimmt und damit auch  $2w_1 + w_2$ . Die schwarz ausgefüllten Bereiche sind diejenigen, durch die die Gerade nicht laufen darf, weil sonst die bedingte Dichte zu groß wird.

Deswegen werden wir jetzt nicht mehr einem Gegner uneingeschränkt die Wahl von  $b$  überlassen, sondern alle Wahlen von  $b$  verbieten, für die  $f_{A_{i,j}^{m_3,m_4}}(b)$  zu klein wird.

**Definition 2.3.9.** Sei im Folgenden  $p \in \mathbb{R}$  mit  $p \geq 2$  beliebig. Für  $i, j \in [n]$  mit  $i < j$  und  $(m_3, m_4) \in \mathcal{D}^2$  mit  $(m_3, m_4) \neq (0, 0)$  sei

$$M_{i,j}^{m_3,m_4}(p) = \left\{ b \in \mathbb{R} \mid 0 \leq f_{A_{i,j}^{m_3,m_4}}(b) \leq \frac{1}{4n^2 m^2 m_{\max} s p} \right\}$$

und  $E_{i,j}^{m_3,m_4}(p)$  bezeichne das Ereignis, dass  $A_{i,j}^{m_3,m_4} \in M_{i,j}^{m_3,m_4}(p)$  gilt. Ferner bezeichne  $E(p)$  das Ereignis, dass für mindestens eine Kombination aus  $i, j$  und  $(m_3, m_4)$  das Ereignis  $E_{i,j}^{m_3,m_4}(p)$  gilt, also

$$E(p) = \bigcup_{i,j,m_3,m_4} E_{i,j}^{m_3,m_4}(p).$$

Das Ereignis  $E(p)$  stellt einen Fehler dar. Falls es gilt, lassen sich für mindestens eine Kombination aus  $i, j$  und  $(m_3, m_4)$  die bedingten Dichten  $f_{A_{i,j}^{m_1,m_2} | A_{i,j}^{m_3,m_4} = b}(a)$  nicht beschränken. Wir werden nun die Fehlerwahrscheinlichkeit abschätzen.

**Lemma 2.3.10.** Für jede Kombination aus  $i, j \in [n]$  mit  $i < j$  und  $(m_3, m_4) \in \mathcal{D}^2$  mit  $(m_3, m_4) \neq (0, 0)$  gilt

$$\Pr \left[ E_{i,j}^{m_3, m_4}(p) \right] \leq \frac{1}{n^2 m^2 p} \quad \text{und} \quad \Pr [E(p)] \leq \frac{1}{2p}.$$

*Beweis.* Wir definieren

$$M_{i,j}^{*, m_3, m_4}(p) = \left\{ b \in M_{i,j}^{m_3, m_4}(p) \mid f_{A_{i,j}^{m_3, m_4}}(b) > 0 \right\}$$

und erhalten als Abschätzung für  $\Pr \left[ E_{i,j}^{m_3, m_4}(p) \right]$

$$\begin{aligned} \Pr \left[ E_{i,j}^{m_3, m_4}(p) \right] &= \int_{M_{i,j}^{m_3, m_4}(p)} f_{A_{i,j}^{m_3, m_4}}(b) \, db \\ &\leq \frac{1}{4n^2 m^2 m_{\max} s p} \int_{M_{i,j}^{*, m_3, m_4}(p)} 1 \, db. \end{aligned}$$

Wir müssen nun noch das Integral  $\int_{M_{i,j}^{*, m_3, m_4}(p)} 1 \, db$  abschätzen. Da diese Abschätzung nicht kompliziert, jedoch lang ist, wird sie nur in Anhang B dargestellt. Als Ergebnis erhält man  $\int_{M_{i,j}^{*, m_3, m_4}(p)} 1 \, db \leq 4m_{\max} s$ . Als Konsequenz für die Wahrscheinlichkeit, dass  $E_{i,j}^{m_3, m_4}(p)$  gilt, bedeutet das

$$\Pr \left[ E_{i,j}^{m_3, m_4}(p) \right] \leq \frac{1}{n^2 m^2 p}.$$

Für die Wahrscheinlichkeit, dass der Fehler  $E(p)$  eintritt, gilt

$$\begin{aligned} \Pr [E(p)] &= \Pr \left[ \bigcup_{i,j, m_3, m_4} E_{i,j}^{m_3, m_4}(p) \right] \\ &\leq \sum_{i,j, m_3, m_4} \Pr \left[ E_{i,j}^{m_3, m_4}(p) \right] \\ &\leq \binom{n}{2} m^2 \cdot \frac{1}{n^2 m^2 p} \\ &\leq \frac{1}{2p}. \end{aligned}$$

□

**Lemma 2.3.11.** Die Menge  $\mathcal{S} \subseteq \mathcal{D}^n$  enthalte keine linear abhängigen Elemente und es gelte  $0^n \notin \mathcal{S}$ . Dann gilt für jedes  $p \in \mathbb{R}$  mit  $p \geq 2$  und für jedes  $\varepsilon \in \mathbb{R}_{\geq 0}$

$$\Pr [\Lambda < \varepsilon \mid \neg E(p)] \leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 s p.$$

*Beweis.* Wir werden nun die oben bereits begonnene Analyse fortsetzen. Um die Dichte der Zufallsvariablen  $\Lambda_{i,j}^{\overline{m}}$  zu beschränken, gehen wir also davon aus, dass ein Gegner die  $w_k$  mit  $k \neq i$  und  $k \neq j$  beliebig vorgibt. Ferner gibt er  $A_{i,j}^{m_3, m_4} = m_3 w_i + m_4 w_j = b$  vor. Da wir nun allerdings in der Situation  $\neg E(p)$  sind, muss er  $b$  so wählen, dass

$$f_{A_{i,j}^{m_3, m_4}}(b) > \frac{1}{4n^2 m^2 m_{\max} s p}$$



### 2.3. Verlierer- und Gültigkeits-Gap im Falle einer stochastischen Nebenbedingung

---

gilt. Für die bedingte Dichte von  $A_{i,j}^{m_1,m_2}$  gilt für jedes  $b \notin M_{i,j}^{m_3,m_4}(p)$  nach Lemma 2.3.7

$$f_{A_{i,j}^{m_1,m_2}|A_{i,j}^{m_3,m_4}=b}(a) < 4n^2 m^2 m_{\max} \phi^2 sp.$$

Wir wissen, dass  $\Lambda_{i,j}^{\bar{m}}$  unter der Bedingung  $A_{i,j}^{m_3,m_4} = b$  und für festgelegte  $w_k$  mit  $k \neq i$  und  $k \neq j$  eine Zufallsvariable ist, die im Vergleich zu  $A_{i,j}^{m_1,m_2}$  nur verschoben ist. Für  $\bar{w} \in \mathbb{R}^{n-2}$  bezeichne  $B(\bar{w})$  das Ereignis, dass die Zufallsvariablen  $w_k$  mit  $k \neq i$  und  $k \neq j$  die Werte  $\bar{w}$  annehmen, also

$$(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_{j-1}, w_{j+1}, \dots, w_n) = \bar{w}.$$

Damit erhalten wir folgendes Ergebnis

$$\forall b \notin M_{i,j}^{m_3,m_4}(p) : \forall \bar{w} \in \mathbb{R}^{n-2} : \Pr \left[ \Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | A_{i,j}^{m_3,m_4} = b \wedge B(\bar{w}) \right] \leq \varepsilon \cdot 4n^2 m^2 m_{\max} \phi^2 sp.$$

Der Fall, dass die  $w_k$  mit  $k \neq i$  und  $k \neq j$  und  $A_{i,j}^{m_3,m_4}$  so festgelegt sind, dass  $\mathcal{L}_{i,j}^{\bar{m}} = \emptyset$ , also  $\Lambda_{i,j}^{\bar{m}} = \perp$  gilt, ist durch diese Abschätzung ebenfalls abgedeckt, denn in diesem Fall gilt  $\Pr \left[ \Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] \right] = 0$ . Es bezeichne  $f^* : \mathbb{R}^{n-2} \rightarrow \mathbb{R}$  die gemeinsame Dichte der  $w_k$  mit  $k \neq i$  und  $k \neq j$ . Da die Koeffizienten unabhängige Zufallsvariablen sind, gilt nach Satz A.5.4 für jedes  $b \in \mathbb{R} \setminus M_{i,j}^{m_3,m_4}(p)$

$$\begin{aligned} & \Pr \left[ \Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | A_{i,j}^{m_3,m_4} = b \right] \\ &= \int_{\mathbb{R}^{n-2}} \Pr \left[ \Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | A_{i,j}^{m_3,m_4} = b \wedge B(\bar{w}) \right] f^*(\bar{w}) d\bar{w} \\ &\leq \varepsilon \cdot 4n^2 m^2 m_{\max} \phi^2 sp \cdot \int_{\mathbb{R}^{n-2}} f^*(\bar{w}) d\bar{w} \\ &= \varepsilon \cdot 4n^2 m^2 m_{\max} \phi^2 sp. \end{aligned}$$

Mit Korollar A.5.3 erhalten wir

$$\begin{aligned} & \Pr \left[ \Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | \neg E_{i,j}^{m_3,m_4}(p) \right] \\ &= \Pr \left[ \Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | A_{i,j}^{m_3,m_4} \in \mathbb{R} \setminus M_{i,j}^{m_3,m_4}(p) \right] \\ &= \frac{1}{\Pr \left[ A_{i,j}^{m_3,m_4} \in \mathbb{R} \setminus M_{i,j}^{m_3,m_4}(p) \right]} \cdot \int_{\mathbb{R} \setminus M_{i,j}^{m_3,m_4}(p)} \Pr \left[ \Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | A_{i,j}^{m_3,m_4} = b \right] f_{A_{i,j}^{m_3,m_4}}(b) db \\ &\leq \frac{1}{\Pr \left[ A_{i,j}^{m_3,m_4} \in \mathbb{R} \setminus M_{i,j}^{m_3,m_4}(p) \right]} \cdot \varepsilon \cdot 4n^2 m^2 m_{\max} \phi^2 sp \int_{\mathbb{R} \setminus M_{i,j}^{m_3,m_4}(p)} f_{A_{i,j}^{m_3,m_4}}(b) db \\ &\leq \varepsilon \cdot 4n^2 m^2 m_{\max} \phi^2 sp. \end{aligned}$$

Es gilt somit

$$\begin{aligned}
 \Pr [\Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | \neg E(p)] &= \frac{1}{\Pr [\neg E(p)]} \cdot \Pr [\Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] \wedge \neg E(p)] \\
 &\leq \frac{2p}{2p-1} \cdot \Pr [\Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] \wedge \neg E(p)] \\
 &\leq 2 \cdot \Pr [\Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] \wedge \neg E_{i,j}^{m_3, m_4}(p)] \\
 &\leq 2 \cdot \Pr [\Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | \neg E_{i,j}^{m_3, m_4}(p)] \\
 &\leq \varepsilon \cdot 8n^2 m^2 m_{\max} \phi^2 sp.
 \end{aligned}$$

Nun können wir Lemma 2.1.3 anwenden, um die Wahrscheinlichkeit, dass  $\Lambda$  kleiner als  $\varepsilon$  wird, unter der Bedingung  $\neg E(p)$  zu beschränken. Es gilt nämlich

$$\begin{aligned}
 \Pr [\Lambda < \varepsilon | \neg E(p)] &= \Pr [\Lambda \in [0, \varepsilon] | \neg E(p)] \\
 &\leq \sum_{i,j,\bar{m}} \Pr [\Lambda_{i,j}^{\bar{m}} \in [0, \varepsilon] | \neg E(p)] \\
 &\leq \binom{n}{2} m^4 \cdot \varepsilon \cdot 8n^2 m^2 m_{\max} \phi^2 sp \\
 &\leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp.
 \end{aligned}$$

□

Das nächste Zwischenziel ist es, das soeben für das Verlierer-Gap gezeigte Resultat auf das Gültigkeits-Gap zu übertragen. Dazu verallgemeinern wir zunächst die Definitionen von Verlierer- und Gültigkeits-Gap ein wenig.

**Definition 2.3.12.** Für eine gegebene Schranke  $t \in \mathbb{R}$  bezeichne  $\Gamma(t)$  die Zufallsvariable, die die Größe des Gültigkeits-Gaps bezüglich der Nebenbedingung  $w \cdot x \leq t$  angibt, und  $\Lambda(t)$  bezeichne die Zufallsvariable, die die Größe des Verlierer-Gaps bezüglich dieser Nebenbedingung angibt.

**Lemma 2.3.13.** Es sei  $B$  ein beliebiges Ereignis, so dass für alle  $t \in \mathbb{R}$  und  $\varepsilon \in \mathbb{R}_{\geq 0}$  gelte

$$\Pr [\Lambda(t) = \varepsilon | B] = \Pr [\Gamma(t) = \varepsilon | B] = 0.$$

Dann gilt für alle  $t \in \mathbb{R}$  und für jedes  $\varepsilon \in \mathbb{R}_{\geq 0}$

$$\Pr [\Lambda(t) < \varepsilon | B] = \Pr [\Gamma(t + \varepsilon) < \varepsilon | B].$$

*Beweis.* Wir fassen das zu lösende Optimierungsproblem als *bikriterielles Problem* auf. Es wird eine Lösung  $x \in \mathcal{S}$  mit einem möglichst hohen Rang und einem möglichst kleinen Gewicht  $w \cdot x$  gesucht. Eine Lösung  $x \in \mathcal{S}$  heißt in diesem Kontext *pareto-optimal*, falls es keine Lösung  $y \in \mathcal{S}$  gibt, die einen höheren Rang als  $x$  und trotzdem kein größeres Gewicht als  $x$  hat.

Wir zeigen nun, dass der Gewinner und der minimale Verlierer des ursprünglichen Optimierungsproblems in dem so definierten bikriteriellen Problem *pareto-optimal* sind. Betrachten wir zunächst den Gewinner. Offensichtlich kann es keine Lösung mit einem höheren Rang und

### 2.3. Verlierer- und Gültigkeits-Gap im Falle einer stochastischen Nebenbedingung

---

einem höchstens genauso großen Gewicht geben, weil sonst diese andere Lösung der Gewinner wäre. Somit ist der Gewinner eine pareto-optimale Lösung.

Auch der minimale Verlierer  $x^{\min}$  ist stets eine pareto-optimale Lösung. Nehmen wir an, es gäbe eine Lösung  $x'$  mit einem höheren Rang und einem höchstens genauso großen Gewicht, dann können zwei Fälle auftreten, entweder ist  $w \cdot x' \leq t$  oder  $w \cdot x' > t$ . Im ersten dieser beiden Fälle wäre  $x'$  oder eine Lösung mit einem noch größeren Rang der Gewinner und  $x^{\min}$  würde per Definition und im Widerspruch zur Annahme gar nicht zur Menge der Verlierer gehören. Im Falle  $w \cdot x' > t$  wäre  $x'$  ein Verlierer, der näher an der Schranke  $t$  liegt als  $x^{\min}$ , dann wäre  $x^{\min}$  im Widerspruch zur Voraussetzung offenbar nicht der minimale Verlierer. Es kann auch der Fall  $w \cdot x^{\min} = w \cdot x'$  eintreten, aber auch dann wäre  $x^{\min}$  nicht der minimale Verlierer, weil  $x'$  einen höheren Rang als  $x^{\min}$  hat.

Andererseits kann man aber auch für jede pareto-optimale Lösung durch eine geeignete Wahl von  $t$  erreichen, dass sie der Gewinner oder der minimale Verlierer wird. Sei  $x$  eine beliebige pareto-optimale Lösung. Setzen wir  $t = w \cdot x$ , so ist  $x$  der Gewinner. Setzen wir  $t = w \cdot x - \varepsilon'$  für ein geeignetes  $\varepsilon' > 0$ , so ist  $x$  der minimale Verlierer.

Es bezeichne  $\mathcal{P} \subseteq \mathcal{S}$  die Menge der pareto-optimalen Lösungen. Dann können wir  $\Gamma(t)$  und  $\Lambda(t)$  auch wie folgt mit Hilfe der Menge  $\mathcal{P}$  beschreiben

$$\Gamma(t) = \min\{t - w \cdot x \mid x \in \mathcal{P}, w \cdot x \leq t\},$$

$$\Lambda(t) = \min\{w \cdot x - t \mid x \in \mathcal{P}, w \cdot x > t\}.$$

Um eine bessere Anschauung zu erhalten, kann man sich vorstellen, dass alle pareto-optimalen Lösungen  $x \in \mathcal{P}$  auf eine Linie gemäß ihrem Gewicht  $w \cdot x$  abgebildet werden. Das Verlierer-Gap  $\Lambda(t)$  entspricht dann dem Abstand des Punktes  $t$  auf dieser Linie zur nächsten pareto-optimalen Lösung, die rechts von  $t$  liegt. Das Gültigkeits-Gap  $\Gamma(t)$  entspricht dem Abstand des Punktes  $t$  auf dieser Linie zur nächsten pareto-optimalen Lösung, die links von  $t$  oder auf  $t$  liegt.

Es gilt also

$$\begin{aligned} \Pr[\Lambda(t) < \varepsilon \mid B] &= \Pr[\exists x \in \mathcal{P} : w \cdot x \in (t, t + \varepsilon) \mid B] \\ &= \Pr[\exists x \in \mathcal{P} : w \cdot x \in (t, t + \varepsilon] \mid B] - \Pr[\exists x \in \mathcal{P} : w \cdot x = t + \varepsilon \mid B] \\ &= \Pr[\exists x \in \mathcal{P} : w \cdot x \in (t, t + \varepsilon] \mid B] \\ &= \Pr[\Gamma(t + \varepsilon) < \varepsilon \mid B]. \end{aligned}$$

□

**Korollar 2.3.14.** *Die Menge  $\mathcal{S} \subseteq \mathcal{D}^n$  enthalte keine linear abhängigen Elemente und es gelte  $0^n \notin \mathcal{S}$ . Dann gilt für jedes  $p \in \mathbb{R}$  mit  $p \geq 2$  und für jedes  $\varepsilon \in \mathbb{R}_{\geq 0}$*

$$\Pr[\Gamma < \varepsilon \mid \neg E(p)] \leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 s p.$$

*Beweis.* Dieses Korollar erhält man als direkte Folgerung aus Lemma 2.3.11 und Lemma 2.3.13. Nach Lemma 2.3.11 gilt für jedes  $t' \in \mathbb{R}$

$$\Pr[\Lambda(t') < \varepsilon \mid \neg E(p)] \leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 s p.$$

Man kann Lemma 2.3.13 anwenden, denn es gilt

$$\Pr[\Lambda(t) = \varepsilon | \neg E(p)] = \frac{\Pr[\Lambda(t) = \varepsilon \wedge \neg E(p)]}{\Pr[\neg E(p)]} \leq \frac{\Pr[\Lambda(t) = \varepsilon]}{\Pr[\neg E(p)]} = 0.$$

und analog  $\Pr[\Gamma(t) = \varepsilon | \neg E(p)] = 0$ . Sind nun  $t$  und  $\varepsilon$  beliebig vorgegeben, so erhält man mit Lemma 2.3.13

$$\begin{aligned} \Pr[\Gamma(t) < \varepsilon | \neg E(p)] &= \Pr[\Lambda(t - \varepsilon) < \varepsilon | \neg E(p)] \\ &\leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp. \end{aligned}$$

□

Wir werden diese Ergebnisse nun auf beliebige Wahlen von  $\mathcal{S}$  mit  $0^n \notin \mathcal{S}$  verallgemeinern.

**Lemma 2.3.15.** *Es sei  $0^n \notin \mathcal{S}$ . Dann gilt für jedes  $p \in \mathbb{R}$  mit  $p \geq 2$  und für jedes  $\varepsilon \in \mathbb{R}_{\geq 0}$*

$$\Pr[\Gamma < \varepsilon | \neg E(p)] \leq \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 sp$$

und

$$\Pr[\Lambda < \varepsilon | \neg E(p)] \leq \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 sp.$$

*Beweis.* Wir zeigen zunächst die Aussage über das Gültigkeits-Gap. Mit Lemma 2.3.13 überträgt sich diese Aussage dann analog zu Korollar 2.3.14 auf das Verlierer-Gap.

Wir werden  $\mathcal{S}$  dafür in  $m$  disjunkte Klassen  $\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(m)}$  partitionieren, die jeweils keine linear abhängigen Elemente enthalten. Bevor wir jedoch die Konstruktion dieser Partition darstellen, werden wir zeigen, wie daraus die Aussage dieses Lemmas folgt. Wir definieren für  $k \in [m]$  das Gültigkeits-Gap  $\Gamma^{(k)}$  der Menge  $\mathcal{S}^{(k)}$ . Dazu definieren wir zunächst den Gewinner  $x^{*,(k)}$  der Menge  $\mathcal{S}^{(k)}$  als das Element mit dem höchsten Rang aus  $\mathcal{S}^{(k)} \cap \mathcal{B}$ . Das Gültigkeits-Gap  $\Gamma^{(k)}$  sei nun definiert als

$$\Gamma^{(k)} = \begin{cases} t - w \cdot x^{*,(k)} & \text{falls } \mathcal{S}^{(k)} \cap \mathcal{B} \neq \emptyset \\ \perp & \text{sonst} \end{cases}.$$

$\Gamma$  nimmt stets einen der Werte der  $\Gamma^{(k)}$  an, denn für den Gewinner  $x^*$  gibt es stets ein  $k \in [m]$ , so dass  $x^* \in \mathcal{S}^{(k)}$  gilt.  $x^*$  ist dann offenbar auch das Element mit dem höchsten Rang aus  $\mathcal{S}^{(k)} \cap \mathcal{B}$  und damit gilt  $\Gamma = \Gamma^{(k)}$ . Ist  $\Gamma = \perp$ , so gilt  $\mathcal{S} \cap \mathcal{B} = \emptyset$  und somit gilt auch  $\Gamma^{(k)} = \perp$  für alle  $k \in [m]$ . Wenn wir also die Wahrscheinlichkeiten, dass die Zufallsvariablen  $\Gamma^{(k)}$  kleiner als  $\varepsilon$  werden, beschränken können, so liefert Lemma 2.1.3 auch eine entsprechende Schranke für die Wahrscheinlichkeit, dass  $\Gamma$  kleiner als  $\varepsilon$  wird. Die erwähnten Wahrscheinlichkeiten sind dabei stets unter der Bedingung  $\neg E(p)$  zu verstehen.

Wir können auf Korollar 2.3.14 zurückgreifen, denn die Menge  $\mathcal{S}^{(k)}$  erfüllt die dort verlangten Voraussetzung, deswegen folgt für jedes  $k \in [m]$

$$\Pr[\Gamma^{(k)} < \varepsilon | \neg E(p)] \leq \varepsilon \cdot 4n^4 m^6 m_{\max} \phi^2 sp.$$

Mit Lemma 2.1.3 folgt nun wie gewünscht

$$\Pr[\Gamma < \varepsilon | \neg E(p)] \leq \sum_{k \in [m]} \Pr[\Gamma^{(k)} < \varepsilon | \neg E(p)] \leq \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 sp.$$

### 2.3. Verlierer- und Gültigkeits-Gap im Falle einer stochastischen Nebenbedingung

---

Wir müssen nun angeben, wie man ein beliebiges  $\mathcal{S} \subseteq D^n \setminus \{0^n\}$  so in  $m$  Klassen  $\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(m)}$  partitionieren kann, dass je zwei Elemente aus derselben Klasse linear unabhängig sind.

Falls  $0 \notin D = \{d_1, \dots, d_m\}$  gilt, so kann man eine gewünschte Partition einfach erzeugen, indem man  $\mathcal{S}^{(k)} = \{x \in \mathcal{S} \mid x_1 = d_k\}$  setzt. Ist jedoch  $0 \in D$ , so erfüllt diese Konstruktion nicht die gewünschten Eigenschaften. Sei o. B. d. A.  $d_m = 0$ , dann definieren wir stattdessen für  $k \in [m-1]$

$$\mathcal{S}^{(k)} = \{x \in \mathcal{S} \mid \exists i \in [n] : x_1 = \dots = x_{i-1} = 0 \text{ und } x_i = d_k\}.$$

Seien nun  $x, y \in \mathcal{S}^{(k)}$  mit  $x \neq y$  und  $k \in [m-1]$  beliebig, wir werden zeigen, dass es kein  $\alpha \in \mathbb{R}$  mit  $\alpha x = y$  oder  $\alpha y = x$  gibt. Nehmen wir an, es gebe ein solches  $\alpha$ . Für den Beweis unterscheiden wir zwei Fälle; stimmt die Anzahl der führenden Nullen von  $x$  und  $y$  überein, so stimmt die erste Stelle hinter den führenden Nullen von  $x$  und  $y$  überein. Es kommt also nur  $\alpha = 1$  in Frage, was im Widerspruch zur Voraussetzung  $x \neq y$  implizieren würde.

Wir betrachten nun den Fall, dass die Anzahl der führenden Nullen von  $x$  und  $y$  nicht überein stimmt. Habe o. B. d. A.  $y$  mehr führende Nullen, dann gibt es eine Stelle  $i$  mit  $y_i = 0$  und  $x_i \neq 0$ . In diesem Fall kann nicht  $\alpha y = x$  gelten, es kommt also nur  $\alpha x = y$  mit  $\alpha = 0$  in Frage, was  $y = 0^n$  im Widerspruch zur Voraussetzung an  $\mathcal{S}$  bedeuten würde.  $\square$

Fassen wir die Ergebnisse von Lemma 2.3.10 und von Lemma 2.3.15 zusammen, so erhalten wir das folgende Korollar.

**Korollar 2.3.16.** *Es sei  $\mathcal{S} \subseteq D^n$  mit  $0^n \notin \mathcal{S}$  beliebig. Dann gilt für alle  $p \in \mathbb{R}$  mit  $p \geq 2$  und für jedes  $\varepsilon \in \mathbb{R}_{\geq 0}$*

$$\Pr[\Gamma < \varepsilon] \leq \frac{1}{2p} + \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 s p$$

und

$$\Pr[\Lambda < \varepsilon] \leq \frac{1}{2p} + \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 s p.$$

*Beweis.* Es gilt

$$\begin{aligned} \Pr[\Lambda < \varepsilon] &= \Pr[E(p)] \cdot \Pr[\Lambda < \varepsilon | E(p)] + \Pr[\neg E(p)] \cdot \Pr[\Lambda < \varepsilon | \neg E(p)] \\ &\leq \Pr[E(p)] + \Pr[\Lambda < \varepsilon | \neg E(p)] \\ &\leq \frac{1}{2p} + \varepsilon \cdot 4n^4 m^7 m_{\max} \phi^2 s p. \end{aligned}$$

Die Aussage für das Gültigkeits-Gap  $\Gamma$  folgt analog.  $\square$

#### 2.3.2 Verlierer- und Gültigkeits-Gap für allgemeine Dichten

Es steht nun noch aus, die Ergebnisse des letzten Abschnittes auf den Fall zu verallgemeinern, dass die Träger der Dichten nicht mehr zwangsweise beschränkt sind. Wir werden diesen Fall auf den speziellen Fall beschränkter Träger zurückführen. Die wesentliche Idee dazu ist es, ein  $s$  so zu wählen, dass die Wahrscheinlichkeit, dass einer der Koeffizienten  $w_i$  außerhalb des Intervalls  $[-s, s]$  liegt, höchstens  $1/(2p)$  beträgt. Dieses Ereignis stellt ebenso wie  $E(p)$  einen

Fehler dar, wir werden es im Folgenden mit  $D(p)$  bezeichnen. Wir können die Wahrscheinlichkeit, dass  $\Lambda$  oder  $\Gamma$  zu klein werden, nur unter der Bedingung  $F(p) = \neg(D(p) \cup E(p))$  beschränken.

Es stellt sich die Frage, wie groß wir  $s$  wählen müssen, damit das Ereignis  $D(p)$  höchstens eine Wahrscheinlichkeit von  $1/(2p)$  hat. Das hängt natürlich essentiell von den gewählten Verteilungen der  $w_i$  ab, aber mit der Markoff-Ungleichung lässt sich  $s$  für jede reelle Zufallsvariable mit endlichem absoluten Erwartungswert beschränken.

**Theorem 2.3.17 (Größe von Verlierer- und Gültigkeits-Gap).** *Es gelte  $0^n \notin \mathcal{S}$  und  $c = \max_{i \in [n]} \mathbf{E}[|w_i|]$ . Dann gilt für alle  $p \in \mathbb{R}$  mit  $p \geq 2$  und für jedes  $\varepsilon \in \mathbb{R}_{\geq 0}$*

$$\Pr[\Gamma < \varepsilon] \leq \frac{1}{p} + \varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2 p^2$$

und

$$\Pr[\Lambda < \varepsilon] \leq \frac{1}{p} + \varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2 p^2.$$

*Beweis.*  $D_i(p)$  bezeichne das Ereignis, dass  $w_i \notin [-2npc, 2npc]$  gilt. Für die Wahrscheinlichkeit dieses Ereignisses gilt

$$\Pr[D_i(p)] = \Pr[w_i \notin [-2npc, 2npc]] = \Pr[|w_i| > 2npc] \leq \Pr[|w_i| > 2np \mathbf{E}[|w_i|]] \leq \frac{1}{2np},$$

wobei die letzte Ungleichung aus der Markoff-Ungleichung folgt. Ferner sei  $D(p) = \bigcup_{i \in [n]} D_i(p)$ . Dann gilt

$$\Pr[D(p)] = \Pr\left[\bigcup_{i \in [n]} D_i(p)\right] \leq \sum_{i \in [n]} \Pr[D_i(p)] \leq n \cdot \frac{1}{2np} = \frac{1}{2p}.$$

Tritt das Ereignis  $\neg D(p)$  ein, so nehmen die Koeffizienten  $w_i$  nur Werte aus  $[-2npc, 2npc]$  an. Wir berechnen nun die Dichte der Zufallsvariablen  $w_1, \dots, w_n$  unter der Bedingung  $\neg D(p)$ . Sei  $s = 2npc$ , dann erhält man für  $i \in [n]$  eine Dichte von  $w_i$  unter der Bedingung  $\neg D(p)$  durch

$$\begin{aligned} f_{i|\neg D(p)}(x) &= \begin{cases} 0 & \text{falls } x \notin [-s, s] \\ \frac{f_i(x)}{F_i(s) - F_i(-s)} & \text{sonst} \end{cases} \\ &\leq \begin{cases} 0 & \text{falls } x \notin [-s, s] \\ \frac{2np}{2np-1} f_i(x) & \text{sonst} \end{cases} \\ &\leq \begin{cases} 0 & \text{falls } x \notin [-s, s] \\ 2f_i(x) & \text{sonst} \end{cases}. \end{aligned}$$

Unter der Bedingung  $\neg D(p)$  sind die Träger der bedingten Dichten der  $w_i$  beschränkt und wir können in diesem Falle auf Lemma 2.3.15 zurückgreifen, um die Wahrscheinlichkeit, dass das Verlierer- oder das Gültigkeits-Gap kleiner als  $\varepsilon$  werden, unter der Bedingung  $F(p) = \neg D(p) \cap \neg E(p)$  zu beschränken. Wir müssen dabei jedoch beachten, dass die maximale Dichte durch die Bedingung  $\neg D(p)$  um den Faktor 2 größer werden kann. Es gilt

$$\Pr[\neg F(p)] = \Pr[D(p) \cup E(p)] \leq \Pr[D(p)] + \Pr[E(p)] \leq \frac{1}{p}$$

## 2.4. Verlierer- und Gültigkeits-Gap im Falle mehrerer stochastischer Nebenbedingungen

und

$$\Pr[\Lambda < \varepsilon | F(p)] \leq \varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2 p^2.$$

Damit ergibt sich die Behauptung des Satzes wie folgt:

$$\begin{aligned} \Pr[\Lambda < \varepsilon] &= \Pr[\neg F(p)] \cdot \Pr[\Lambda < \varepsilon | \neg F(p)] + \Pr[F(p)] \cdot \Pr[\Lambda < \varepsilon | F(p)] \\ &\leq \Pr[\neg F(p)] + \Pr[\Lambda < \varepsilon | F(p)] \\ &\leq \frac{1}{p} + \varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2 p^2. \end{aligned}$$

Die Aussage über das Gültigkeits-Gap  $\Gamma$  folgt analog.  $\square$

**Korollar 2.3.18.** *Es gelte  $0^n \notin \mathcal{S}$  und  $c = \max_{i \in [n]} \mathbf{E}[|w_i|]$ . Dann gilt für alle  $n \in \mathbb{N}$  und für jedes  $\varepsilon \in \mathbb{R}_{\geq 0}$  mit  $\varepsilon \leq (256n^5 m^7 m_{\max} \phi^2)^{-1}$*

$$\Pr[\Gamma < \varepsilon] \leq 2(\varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2)^{1/3}$$

und

$$\Pr[\Lambda < \varepsilon] \leq 2(\varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2)^{1/3}.$$

*Beweis.* Das Korollar folgt direkt aus Theorem 2.3.17, indem man  $p = (\varepsilon \cdot 32n^5 m^7 m_{\max} \phi^2)^{-1/3}$  setzt. Die oberer Schranke für  $\varepsilon$  folgt aus der Bedingung  $p \geq 2$ .  $\square$

## 2.4 Verlierer- und Gültigkeits-Gap im Falle mehrerer stochastischer Nebenbedingungen

In diesem Abschnitt werden wir die Resultate, die wir für das Verlierer- und das Gültigkeits-Gap im Falle einer stochastischen Nebenbedingung erhalten haben, auf den Fall verallgemeinern, dass es  $k \geq 1$  viele stochastische Nebenbedingungen gibt. Die  $j$ -te Nebenbedingung mit  $j \in [k]$  habe o. B. d. A. die Form  $w^{(j)} \cdot x \leq t_j$  für ein  $w^{(j)} \in \mathbb{R}^n$  und die Menge der Lösungen aus  $\mathcal{D}^n$ , die diese Nebenbedingung erfüllen, werde mit  $\mathcal{B}_j$  bezeichnet. Ebenso wie im Falle einer stochastischen Nebenbedingung gehen wir auch in diesem Fall davon aus, dass eine Menge  $\mathcal{S} \subseteq \mathcal{D}^n$  und eine Zielfunktion beliebig vorgegeben sind. Auch hier fordern wir von der Zielfunktion lediglich, dass sie auf den Lösungen aus  $\mathcal{S}$  eine eindeutige Reihenfolge induziert.

Wir müssen zunächst die Definitionen der Gaps an den Fall mehrerer stochastischer Nebenbedingungen anpassen.

**Definition 2.4.1 (Gültigkeits-Gap).** *Der Gewinner  $x^*$  sei die Lösung mit dem höchsten Rang aus  $\mathcal{S} \cap \mathcal{B}_1 \cap \dots \cap \mathcal{B}_k$ . Das Gültigkeits-Gap  $\Gamma$  sei der kleinste Abstand, den der Gewinner von einer der Schranken  $t_j$  hat. Es gilt also*

$$\Gamma = \begin{cases} \min\{t_j - w^{(j)} \cdot x^* | j \in [k]\} & \text{falls } \mathcal{S} \cap \mathcal{B}_1 \cap \dots \cap \mathcal{B}_k \neq \emptyset \\ \perp & \text{sonst} \end{cases}.$$

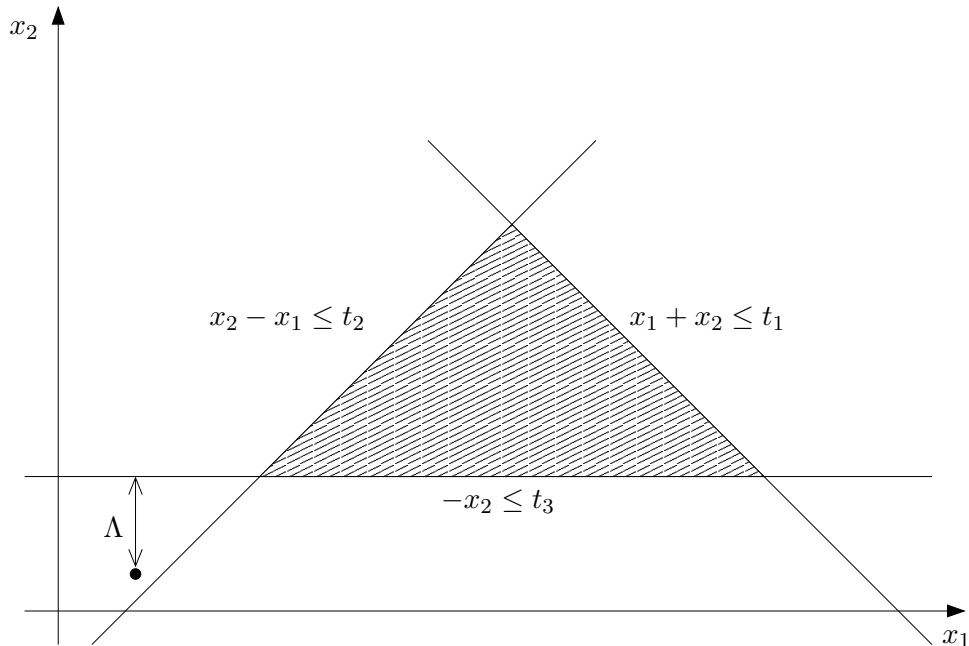


Abbildung 2.3: In dieser Abbildung wird das Verlierer-Gap für mehrere Nebenbedingungen erläutert. Der eingezeichnete Punkt sei der einzige Verlierer und der schraffierte Bereich sei der erlaubte Bereich, der durch die drei Nebenbedingungen beschrieben wird. Zwei dieser Nebenbedingungen sind für den eingezeichneten Suchpunkt nicht erfüllt. Das Verlierer-Gap entspricht dem Abstand zu der Nebenbedingung, die am deutlichsten nicht erfüllt ist.

Eine Lösung ist ein Verlierer, wenn sie einen höheren Rang als der Gewinner  $x^*$  hat, jedoch wegen mindestens einer der stochastischen Nebenbedingungen nicht zur Menge der erlaubten Lösungen zählt. Im Allgemeinen kann der Fall eintreten, dass ein Verlierer die  $j$ -te Nebenbedingung nicht erfüllt, aber trotzdem einen sehr kleinen Abstand zu der Schranke  $t_j$  hat. Für die Anwendung des Verlierer-Gaps im nächsten Kapitel ist es jedoch nur wichtig, dass es für jeden Verlierer eine Nebenbedingung gibt, die er deutlich nicht erfüllt. Das rechtfertigt die folgende Definition des Verlierer-Gaps, die in Abbildung 2.3 veranschaulicht wird.

**Definition 2.4.2 (Verlierer-Gap).** *Es bezeichne  $\mathcal{L}$  die Menge der **Verlierer**, also derjenigen Elemente aus  $\mathcal{S}$ , die einen höheren Rang als der Gewinner  $x^*$  haben, jedoch in mindestens einer der Mengen  $\mathcal{B}_1, \dots, \mathcal{B}_k$  nicht enthalten sind. Das **Verlierer-Gap**  $\Lambda$  ist nun definiert als*

$$\Lambda = \begin{cases} \min_{x \in \mathcal{L}} \max_{j \in [k]} \{w^{(j)} \cdot x - t_j\} & \text{falls } \mathcal{L} \neq \emptyset \\ \perp & \text{sonst} \end{cases}. \quad (2.4)$$

Der **minimale Verlierer**  $x^{\min}$  sei diejenige Lösung  $x \in \mathcal{L}$ , für die in Gleichung (2.4) das Minimum angenommen wird.

Aufbauend auf den Ergebnissen, die wir für eine stochastische Nebenbedingung erhalten haben, können wir das folgende Theorem beweisen.



## 2.4. Verlierer- und Gültigkeits-Gap im Falle mehrerer stochastischer Nebenbedingungen

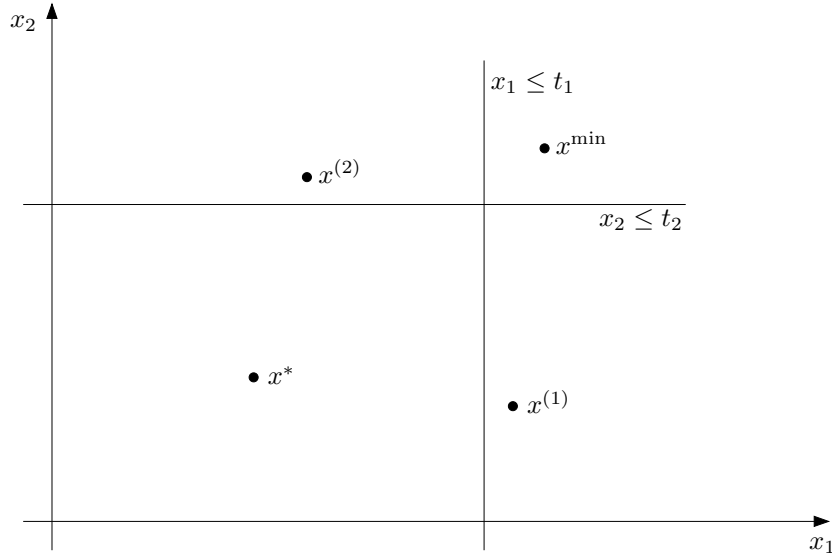


Abbildung 2.4: In dieser Abbildung ist ein Beispiel dafür eingezeichnet, dass der minimale Verlierer bezüglich der Nebenbedingungen kein minimaler Verlierer bezüglich einer der Nebenbedingungen sein muss.  $x^*$  sei der Gewinner und die Elemente  $x^{\min}$ ,  $x^{(1)}$  und  $x^{(2)}$  haben einen höheren Rang als  $x^*$ .  $x^{\min}$  ist der minimale Verlierer bezüglich der beiden Nebenbedingungen.  $x^{(1)}$  ist der minimale Verlierer bezüglich der Nebenbedingung  $x_1 \leq t_1$  und  $x^{(2)}$  ist der minimale Verlierer bezüglich der Nebenbedingung  $x_2 \leq t_2$ .

**Theorem 2.4.3 (Größe von Verlierer- und Gültigkeits-Gap).** *Es gelte  $0^n \notin \mathcal{S}$  und  $c = \max_{j \in [k]} \max_{i \in [n]} \mathbf{E} \left[ |w_i^{(j)}| \right]$ . Dann gilt für alle  $\varepsilon \in \mathbb{R}_{\geq 0}$  mit  $\varepsilon \leq (256n^5 m^7 m_{\max} \phi^2)^{-1}$*

$$\Pr [\Gamma < \varepsilon] \leq 2k(\varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2)^{1/3}$$

und

$$\Pr [\Lambda < \varepsilon] \leq 2k(\varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2)^{1/3}.$$

*Beweis.* Wir zeigen zunächst die Aussage über das Gültigkeits-Gap. Gilt  $\Gamma < \varepsilon$ , so gibt es ein  $j \in [k]$ , für das  $t_j - w^{(j)} \cdot x^* < \varepsilon$  gilt. Somit ist

$$\Pr [\Gamma < \varepsilon] \leq \Pr \left[ \exists j \in [k] : t_j - w^{(j)} \cdot x^* < \varepsilon \right] \leq \sum_{j \in [k]} \Pr \left[ t_j - w^{(j)} \cdot x^* < \varepsilon \right].$$

Da die Koeffizienten der Nebenbedingungen unabhängig sind, können wir beim Betrachten der  $j$ -ten Nebenbedingung davon ausgehen, dass die Koeffizienten der anderen Nebenbedingungen bereits beliebig gewählt wurden. Wir sind dann in einer Situation, in der wir Korollar 2.3.18

anwenden können. Wie gewünscht ergibt sich insgesamt

$$\begin{aligned}
 \Pr [\Gamma < \varepsilon] &\leq \sum_{j \in [k]} \Pr \left[ t_j - w^{(j)} \cdot x^* < \varepsilon \right] \\
 &\leq \sum_{j \in [k]} 2(\varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2)^{1/3} \\
 &\leq 2k(\varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2)^{1/3}.
 \end{aligned}$$

Wir wenden uns nun dem Verlierer-Gap zu und stellen fest, dass eine Übertragung der Ergebnisse des Falles mit einer stochastischen Nebenbedingung wie beim Gültigkeits-Gap nicht ohne weiteres möglich ist, weil der minimale Verlierer bezüglich der stochastischen Nebenbedingungen im Allgemeinen kein minimaler Verlierer bezüglich einer der Nebenbedingungen ist. Ein Beispiel dafür ist in Abbildung 2.4 dargestellt. Wir können jedoch auf das Ergebnis für das Gültigkeits-Gap zurückgreifen.

Nehmen wir an, dass  $\Lambda < \varepsilon$  gilt. Dann gibt es einen Verlierer  $x$ , so dass für alle Nebenbedingungen  $j \in [k]$  gilt  $w^{(j)} \cdot x - t_j < \varepsilon$ . Es sei  $x_L$  der Verlierer mit dem höchsten Rang, der diese Ungleichungen erfüllt. Wir ändern das gegebene Optimierungsproblem nun dadurch, dass wir jede Schranke  $t_j$  mit  $j \in [k]$  um  $\varepsilon$  erhöhen. Dadurch erhalten wir ein neues ganzzahliges lineares Optimierungsproblem  $\Pi'$ , in dem  $x_L$  alle Nebenbedingungen erfüllt und somit kein Verlierer ist.  $x_L$  ist sogar der Gewinner des neuen Problems  $\Pi'$ , denn  $x_L$  ist nach Definition die Lösung mit dem höchsten Rang, die alle Nebenbedingungen erfüllt. Da  $x_L$  im ursprünglichen Problem ungültig ist, gibt es ein  $j \in [k]$ , für das  $t_j < w^{(j)} \cdot x_L < t_j + \varepsilon$  gilt, das Gültigkeits-Gap  $\Gamma'$  in dem neuen Problem  $\Pi'$  ist also kleiner als  $\varepsilon$ . Die Wahrscheinlichkeit dafür ist nach obigem Ergebnis höchstens  $2k(\varepsilon \cdot 32cn^5 m^7 m_{\max} \phi^2)^{1/3}$ .  $\square$

# Die geglättete Komplexität ganzzahliger Optimierungsprobleme

## 3.1 Beweis von Theorem 1.3.7

Wir werden in diesem Abschnitt Theorem 1.3.7 beweisen. Es seien ein ganzzahliges lineares Optimierungsproblem  $\Pi$  und ein Perturbationsmodell  $\mathcal{H}$  für  $\Pi$  beliebig gewählt. Als erstes werden wir zeigen, wie man einen Algorithmus mit polynomieller erwarteter Laufzeit für  $\Pi_u^{\mathcal{H}}$  nutzen kann, um einen Algorithmus mit polynomieller geglätteter Laufzeit für  $\Pi$  zu konstruieren. Dazu stellen wir zunächst die Zertifizierer vor, also die Algorithmen, die entscheiden, ob durch das Runden der Koeffizienten die optimale Lösung verändert wird.

Es sei  $I$  eine Eingabe für  $\Pi$ , die mit dem Perturbationsmodell  $\mathcal{H}$  erzeugt wurde. Wir gehen davon aus, dass  $b$  viele Nachkommabits von jedem stochastischen Koeffizienten in der Eingabe  $I$  gelesen werden und dass die Koeffizienten nach dem  $b$ -ten Nachkommabit abgerundet werden. Es sei ein Algorithmus  $A$  für  $\Pi$  gegeben, der im Erwartungswert eine pseudopolynomielle Laufzeit hat und der die optimale Lösung von Probleminstanzen berechnen kann, bei denen jeder Koeffizient in den als stochastisch ausgewählten Komponenten eine ganze Zahl ist. Sind die Koeffizienten in den stochastischen Komponenten rational, so können sie so skaliert werden, dass sie ganz werden, dann kann Algorithmus  $A$  benutzt werden, um das Optimum zu berechnen. Algorithmus  $A$  kann also insbesondere eingesetzt werden, um die optimale Lösung  $x'$  bezüglich der abgerundeten Koeffizienten zu berechnen.

Ist eines der Gaps zu klein, so kann nicht bestätigt werden, dass  $x'$  auch das Optimum der Probleminstanz  $I$  ist, bei der die Koeffizienten nicht gerundet sind. Je größer die Rechengenauigkeit  $b$  wird, desto kleiner wird die untere Schranke für die Gaps, ab der  $x'$  als Optimum zertifiziert werden kann.

Wir werden im Folgenden drei verschiedene Zertifizierer vorstellen, dabei ist durch das Perturbationsmodell bestimmt, welcher Zertifizierer benutzt werden muss. Es werden die drei Fälle unterschieden, dass nur die Zielfunktion stochastisch ist, dass die Zielfunktion feststeht und mindestens eine Nebenbedingung stochastisch ist oder dass sowohl die Zielfunktion als auch mindestens eine Nebenbedingung von stochastischer Natur sind.

Für  $a \in \mathbb{R}$  bezeichne  $\lfloor a \rfloor_b$  die Zahl, die man erhält, wenn man  $a$  nach dem  $b$ -ten Nachkommabit abrundet. Ist  $B \in \mathbb{R}^{j \times l}$  eine Matrix, so bezeichne  $\lfloor B \rfloor_b$  die Matrix, die aus  $B$  hervorgeht, indem jede Komponente nach dem  $b$ -ten Nachkommabit abgerundet wird, und  $B + a$  sei die Matrix, die aus  $B$  hervorgeht, indem zu jeder Komponente  $a$  addiert wird. Diese Konventionen gelten

insbesondere für Vektoren  $B \in \mathbb{R}^j$ . Ferner gelten analoge Konventionen für  $\lceil \cdot \rceil_b$ . Für zwei Vektoren  $B_1, B_2 \in \mathbb{R}^j$  sei  $B_1 \leq B_2$  komponentenweise zu verstehen.

### 3.1.1 Ein Zertifizierer für stochastische Zielfunktionen im binären Fall

Wir betrachten zunächst den binären Fall. Die Menge der gültigen Lösungen  $\mathcal{S} \subseteq \{0, 1\}^n$  stehe beliebig fest und die lineare Zielfunktion  $p : \mathcal{S} \rightarrow \mathbb{R}$  sei von stochastischer Natur. Sie habe also die Form  $p(x) = c_1x_1 + \dots + c_nx_n = c \cdot x$ , wobei  $c = (c_1, \dots, c_n)$  die stochastischen Koeffizienten und  $x = (x_1, \dots, x_n)^T$  die binären Variablen bezeichne.

Da die Koeffizienten  $c$  mit Wahrscheinlichkeit 1 irrational sind, können wir Algorithmus  $A$  nicht anwenden, um das Optimum der Problem Instanz  $I$  direkt zu berechnen. Wir können jedoch das Optimum bezüglich einer beliebigen Funktion  $p(x) = a \cdot x$  mit  $a \in \mathbb{Q}^n$  bestimmen, denn diese Funktion kann so skaliert werden, dass alle Koeffizienten ganze Zahlen sind und wir Algorithmus  $A$  zur Bestimmung des Optimums einsetzen können.

Zunächst nutzen wir Algorithmus  $A$ , um das Optimum  $x'$  bezüglich der abgerundeten Koeffizienten  $\lfloor c \rfloor_b$  zu berechnen. Um zu überprüfen, ob  $x'$  auch optimal bezüglich der nicht gerundeten Koeffizienten ist, genügt ein weiterer Aufruf von Algorithmus  $A$  und zwar mit den Koeffizienten  $\bar{c}$ , die wie folgt definiert sind

$$\bar{c}_i = \begin{cases} \lceil c_i \rceil_b = \lfloor c_i \rfloor_b + 2^{-b} & \text{falls } x'_i = 0 \\ \lfloor c_i \rfloor_b & \text{falls } x'_i = 1 \end{cases} .$$

Ist  $x'$  auch das Optimum bezüglich der Koeffizienten  $\bar{c}$ , so ist  $x'$  ebenfalls das Optimum bezüglich der nicht gerundeten Koeffizienten. Um das zu zeigen, betrachten wir die Funktion  $\delta(x) = c \cdot x - \bar{c} \cdot x$ . Diese wird durch  $x'$  maximiert. Ist  $x'$  auch das Optimum bezüglich der Koeffizienten  $\bar{c}$ , so maximiert  $x'$  die Funktionen  $\delta(x)$  und  $\bar{c} \cdot x$  und somit auch ihre Summe  $\delta(x) + \bar{c} \cdot x = c \cdot x$ .

Liefert der Aufruf von  $A$  bezüglich der Koeffizienten  $\bar{c}$  eine Lösung zurück, die ungleich  $x'$  ist, so kann kein zertifiziertes Optimum berechnet werden. Dieser Fall kann nur dann eintreten, wenn das Gewinner-Gap nicht größer als  $n2^{-b}$  ist, gilt nämlich  $c \cdot x' - c \cdot y > n2^{-b}$  für alle erlaubten Lösungen  $y \in \mathcal{S}$  mit  $y \neq x'$ , so gilt für jedes solche  $y$  auch nach dem Runden noch  $\bar{c} \cdot x' > \bar{c} \cdot y$  und  $\lfloor c \rfloor_b \cdot x' > \lfloor c \rfloor_b \cdot y$ .

Eine direkte Übertragung dieses Zertifizierers auf den allgemeinen Fall ganzzahliger Optimierungsprobleme ist nicht möglich. Wir vermuten sogar, dass es keinen Zertifizierer gibt, der nur Aufrufe von  $A$  durchführt, bei denen jeder Koeffizient entweder nach dem  $b$ -ten Bit auf- oder abgerundet ist, und der im Allgemeinen nach polynomiell vielen Aufrufen von  $A$  das Optimum zertifizieren kann.

Dazu untersuchen wir, welche Informationen ein Aufruf von  $A$  liefern kann, bei dem jeder Koeffizient entweder nach dem  $b$ -ten Bit auf- oder abgerundet ist. Für  $a \in \{0, 1\}^n$  bezeichne  $c^a$  den Vektor bestehend aus Koeffizienten

$$c_i^a = \begin{cases} \lfloor c_i \rfloor_b & \text{falls } a_i = 0 \\ \lceil c_i \rceil_b & \text{falls } a_i = 1 \end{cases} .$$

Wir gehen davon aus, dass wir bereits einen Aufruf von  $A$  mit den Koeffizienten  $c^a$  für ein beliebiges  $a \in \{0, 1\}^n$  durchgeführt haben, der uns die Lösung  $x'$  zurückgeliefert hat,

### 3.1. Beweis von Theorem 1.3.7

---

und untersuchen nun, welche Information uns ein Aufruf mit den Koeffizienten  $c^{a'}$  für ein  $a' \in \{0, 1\}^n$  mit  $a' \neq a$  liefert. Erhalten wir eine andere Lösung als  $x'$  zurück, so können wir  $x'$  nicht als Optimum zertifizieren, erhalten wir wieder  $x'$  als Lösung, so können wir — wie eine leichte Rechnung zeigt — genau die Elemente der folgenden Menge  $M(a')$  als optimale Lösung bezüglich der nicht gerundeten Koeffizienten ausschließen

$$M(a') = \left\{ x \in \mathcal{S} \setminus \{x'\} \mid \forall i \in [n] : \begin{array}{l} a'_i = 0 \Rightarrow x_i \leq x'_i \\ a'_i = 1 \Rightarrow x_i \geq x'_i \end{array} \right\}.$$

Der Wertebereich der ganzzahligen Variablen sei  $\mathcal{D} = \{d_1, \dots, d_m\}$  mit  $d_1 < \dots < d_m$ . Falls für jede Komponente  $i \in [n]$  des Vektors  $x'$  gilt  $x'_i \in \{d_1, d_m\}$ , so gilt für  $a' \in \{0, 1\}^n$  mit

$$a'_i = \begin{cases} 0 & \text{falls } x'_i = d_m \\ 1 & \text{falls } x'_i = d_1 \end{cases}$$

$M(a') = \mathcal{S} \setminus \{x'\}$ . Wir können dann also  $x'$  durch einen weiteren Aufruf von  $A$  als Optimum zertifizieren. Diese Eigenschaft ist im binären Fall immer erfüllt, weswegen dort stets zwei Aufrufe von  $A$  genügen. Die Koeffizienten  $\bar{c}$ , mit denen Algorithmus  $A$  beim binären Zertifizierer aufgerufen wird, entsprechen gerade den Koeffizienten  $c^{a'}$ .

Sei nun  $m > 2$  angenommen. Die wenigsten Lösungen kann man ausschließen, wenn für jede Komponente  $x'_i = d_{\lceil m/2 \rceil}$  gilt. Man kann dann durch einen Aufruf von  $A$  nur höchstens  $(\lceil m/2 \rceil + 1)^n$  viele Lösungen ausschließen, was nur ein exponentiell kleiner Anteil an der Menge aller möglichen Lösungen ist, falls  $\mathcal{S} = \mathcal{D}^n$  gilt. Deswegen vermuten wir, dass es im Allgemeinen nicht möglich ist, mit polynomiell vielen Aufrufen von  $A$  ein Element als garantiertes Optimum zu bestimmen, wenn bei jedem Aufruf jeder Koeffizient entweder nach dem  $b$ -ten Bit auf- oder abgerundet wird. Wir werden jedoch im nächsten Abschnitt zeigen, dass man doch mit polynomiell vielen Aufrufen von  $A$  auskommt, wenn man auch andere Koeffizienten als nur auf- oder abgerundete zulässt.

#### 3.1.2 Ein Zertifizierer für stochastische Zielfunktionen im allgemeinen Fall

Auch in diesem Fall betrachten wir die Situation, dass die Menge der gültigen Lösungen  $\mathcal{S} \subseteq \mathcal{D}^n$  beliebig feststeht und dass die lineare Zielfunktion  $p : \mathcal{S} \rightarrow \mathbb{R}$  von stochastischer Natur ist. Im Pseudocode werden wir das Optimum bezüglich der Zielfunktion  $a \cdot x$  für  $a \in \mathbb{Q}^n$  mit  $A(a)$  bezeichnen, da wir Algorithmus  $A$  aufrufen, um es zu bestimmen.

Zunächst werden wir mit Algorithmus  $A$  das Optimum  $x'$  bezüglich der abgerundeten Koeffizienten berechnen und dann überprüfen, ob dieses Optimum auch optimal bezüglich der nicht gerundeten Koeffizienten ist. Da das Runden nach dem  $b$ -ten Nachkommabit jeden Koeffizienten um höchstens  $2^{-b}$  ändert und es  $n$  Koeffizienten gibt, die jeweils höchstens mit  $m_{\max}$  gewichtet sind, wird der Wert jeder Lösung aus  $\mathcal{S}$  durch das Runden um höchstens  $nm_{\max}2^{-b}$  geändert. Falls also der Abstand des Wertes von  $x'$  von dem Wert jeder anderen erlaubten Lösung größer als  $2nm_{\max}2^{-b}$  ist, so wissen wir, dass  $x'$  auch vor dem Runden optimal gewesen sein muss. Diese Beobachtung wird in Algorithmus 1 ausgenutzt, dessen Eigenschaften in Lemma 3.1.1 zusammengefasst sind. Die Rückgabe des Algorithmus ist das Optimum von  $I$ , also das Optimum bezüglich der nicht gerundeten Koeffizienten, oder  $\perp$ , falls dieses Optimum mit der gegebenen Rechengenauigkeit  $b$  nicht eindeutig bestimmt werden kann.

---

**Algorithmus 1** ZertifiziererZF (nur die Zielfunktion ist stochastisch)

---

```

 $x' = A(\lfloor c \rfloor_b)$ 
for  $i = 1$  to  $n$  do
     $\bar{c} = \lfloor c \rfloor_b$ 
     $\bar{c}_i = \lfloor c_i \rfloor_b + (nm_{\max} + 1)2^{-b+1}$ 
     $x = A(\bar{c})$ 
    if  $x \neq x'$  then
        return  $\perp$ 
    end if
     $\bar{c} = \lfloor c \rfloor_b$ 
     $\bar{c}_i = \lfloor c_i \rfloor_b - (nm_{\max} + 1)2^{-b+1}$ 
     $x = A(\bar{c})$ 
    if  $x \neq x'$  then
        return  $\perp$ 
    end if
end for
return  $x'$ 

```

---

**Lemma 3.1.1.** *Gibt Algorithmus 1 eine Lösung zurück, so ist diese optimal bezüglich der nicht gerundeten Koeffizienten. Wenn er keine Lösung zurückgibt, so ist das Gewinner-Gap nicht größer als  $nm_{\max}^2 2^{-b+4}$ .*

*Beweis.* Wir betrachten zunächst den Fall, dass ein Optimum  $x'$  zurückgegeben wird. Wir möchten zeigen, dass  $x'$  die optimale Lösung der nicht gerundeten Problem Instanz  $I$  ist. Im Widerspruch dazu nehmen wir an, es gäbe ein  $y \in \mathcal{S}$  mit  $y \neq x'$ , für das  $c \cdot y \geq c \cdot x'$  gilt. Durch das Runden wird der Wert jeder Lösung aus  $\mathcal{S}$  um höchstens  $nm_{\max} 2^{-b}$  geändert. Weil  $x'$  das Optimum bezüglich der abgerundeten Koeffizienten ist, gilt  $\lfloor c \rfloor_b \cdot x' \geq \lfloor c \rfloor_b \cdot y$ , somit ergibt sich insgesamt  $0 \leq \lfloor c \rfloor_b \cdot x' - \lfloor c \rfloor_b \cdot y \leq nm_{\max} 2^{-b+1}$ .

Nach Voraussetzung existiert eine Stelle  $i \in [n]$ , in der sich  $y$  und  $x'$  unterscheiden. Wir behandeln zunächst den Fall, dass  $y_i > x'_i$  gilt. Algorithmus 1 führt einen Aufruf von  $A$  durch, bei dem alle Koeffizienten abgerundet werden und der  $i$ -te Koeffizient zusätzlich um  $(nm_{\max} + 1)2^{-b+1}$  erhöht wird. Es bezeichne  $\bar{c}$  diese Koeffizienten. Dann gilt

$$\begin{aligned}
 \bar{c} \cdot y - \bar{c} \cdot x' &= \lfloor c \rfloor_b \cdot y - \lfloor c \rfloor_b \cdot x' + (y_i - x'_i)(nm_{\max} + 1)2^{-b+1} \\
 &\geq \lfloor c \rfloor_b \cdot y - \lfloor c \rfloor_b \cdot x' + (nm_{\max} + 1)2^{-b+1} \\
 &\geq -nm_{\max} 2^{-b+1} + (nm_{\max} + 1)2^{-b+1} \\
 &= 2^{-b+1} > 0.
 \end{aligned}$$

Somit wäre beim Aufruf von  $A$  mit den Koeffizienten  $\bar{c}$  nicht  $x'$  als Optimum herausgekommen und die Rückgabe des Zertifizierers wäre  $\perp$  gewesen. Der Fall  $y_i < x'_i$  kann analog behandelt werden, wenn man den Aufruf von  $A$  betrachtet, bei dem der Koeffizient  $c_i$  nach dem Abrunden noch um  $(nm_{\max} + 1)2^{-b+1}$  verkleinert wird. Somit ist der erste Teil des Lemmas gezeigt, denn die Annahme, dass  $x'$  nicht optimal ist, wenn es ausgegeben wird, führt in jedem Falle zu einem Widerspruch.

Wir müssen nun noch zeigen, dass höchstens dann kein Optimum zurückgegeben wird, wenn das Gewinner-Gap nicht größer als  $nm_{\max}^2 2^{-b+4}$  ist. Sei  $\Delta > nm_{\max}^2 2^{-b+4}$  angenommen und

### 3.1. Beweis von Theorem 1.3.7

---

$x^*$  das Optimum der Problem Instanz  $I$ . Da der Wert jeder Lösung durch das Runden um höchstens  $nm_{\max}2^{-b}$  verändert wird, ist  $x^*$  auch das Optimum bezüglich der abgerundeten Koeffizienten, es gilt also  $x' = x^*$ .

Nach dem Aufruf mit den abgerundeten Koeffizienten wird Algorithmus  $A$  noch  $2n$  mal aufgerufen. Wir betrachten einen beliebigen solchen Aufruf, dabei bezeichne  $\bar{c}$  den dazugehörigen Koeffizientenvektor und  $y \in \mathcal{S}$  sei eine beliebige Lösung. Für alle  $x \in \mathcal{S}$  gilt

$$|c \cdot x - \bar{c} \cdot x| \leq nm_{\max}2^{-b} + m_{\max}(nm_{\max} + 1)2^{-b+1},$$

wobei der erste Summand daraus resultiert, dass jeder Koeffizient durch das Runden um höchstens  $2^{-b}$  verändert wird, und der zweite daraus, dass ein Koeffizient, der höchstens mit  $m_{\max}$  gewichtet ist, noch zusätzlich um  $(nm_{\max} + 1)2^{-b+1}$  verändert wird. Durch eine einfache Abschätzung erhält man für jedes  $x \in \mathcal{S}$  als Ergebnis

$$|c \cdot x - \bar{c} \cdot x| < nm_{\max}^2 2^{-b+3}.$$

Weil  $\Delta > nm_{\max}^2 2^{-b+4}$  ist, gilt  $c \cdot x' - c \cdot y > nm_{\max}^2 2^{-b+4}$ . Somit erhält man insgesamt

$$\begin{aligned} \bar{c} \cdot x' - \bar{c} \cdot y &= (c \cdot x' - c \cdot y) + (\bar{c} \cdot x' - c \cdot x') - (\bar{c} \cdot y - c \cdot y) \\ &> nm_{\max}^2 2^{-b+4} + (\bar{c} \cdot x' - c \cdot x') - (\bar{c} \cdot y - c \cdot y) \\ &> nm_{\max}^2 2^{-b+4} - 2 \cdot nm_{\max}^2 2^{-b+3} = 0. \end{aligned}$$

Falls das Gewinner-Gap größer als  $nm_{\max}^2 2^{-b+4}$  ist, wird also  $x'$  bei jedem Aufruf von  $A$  als Optimum zurückgeliefert und somit wird  $x'$  auch von Algorithmus 1 als zertifiziertes Optimum ausgegeben.  $\square$

#### 3.1.3 Ein Zertifizierer für stochastische Nebenbedingungen

Wir betrachten nun die Situation, dass die Zielfunktion feststeht und dass es  $k' \geq 1$  viele stochastische Nebenbedingungen gibt. Diese können o. B. d. A. in der Form  $Bx \leq t$  formuliert werden, wobei  $B \in \mathbb{R}^{k' \times n}$  eine Matrix ist, die die stochastischen Koeffizienten enthält, und  $t \in \mathbb{R}^{k'}$  ein Vektor, der die Schranken enthält. Die Menge der erlaubten Lösungen ergibt sich in diesem Fall als Schnitt einer beliebig vorgegebenen Menge  $\mathcal{S} \subseteq \mathcal{D}^n$  mit den Halbräumen  $\mathcal{B}_1, \dots, \mathcal{B}_{k'}$ , die durch die Nebenbedingungen beschrieben werden.

An dieser Stelle erinnern wir noch einmal daran, dass das Verlierer-Gap undefiniert ist, wenn es keinen Verlierer gibt, und dass das Gültigkeits-Gap undefiniert ist, wenn keine gültige Lösung existiert. Die Nichtexistenz eines Verlierers beeinträchtigt den Zertifizierer nicht und auch der Fall, dass keine Lösung aus  $\mathcal{S}$  existiert, die alle Nebenbedingungen erfüllt, ist unproblematisch. Wir behandeln die Antwort des Algorithmus  $A$ , dass keine gültige Lösung existiert, wie jede andere Ausgabe. Insbesondere kann also auch zertifiziert werden, dass bezüglich der nicht gerundeten Koeffizienten keine Lösung gültig ist.

Auch in dieser Situation können wir das Optimum der Problem Instanz  $I$  nicht direkt mit Algorithmus  $A$  berechnen, weil die stochastischen Koeffizienten mit Wahrscheinlichkeit 1 irrationale Zahlen sind. Ist  $C \in \mathbb{Q}^{k' \times n}$  und  $t' \in \mathbb{Q}^{k'}$ , so kann das Optimum bezüglich der Nebenbedingungen  $Cx \leq t'$  mit Algorithmus  $A$  bestimmt werden, indem die Koeffizienten durch eine entsprechende Skalierung ganzzahlig gemacht werden. Einen Aufruf von Algorithmus  $A$  mit diesen Nebenbedingungen werden wir im Pseudocode mit  $A(C, t')$  bezeichnen.

---

**Algorithmus 2** ZertifiziererNB (nur Nebenbedingungen sind stochastisch)

---

```

 $x = A(\lfloor B \rfloor_b, \lfloor t \rfloor_b + (nm_{\max} + 1)2^{-b})$ 
if  $\lfloor B \rfloor_b x \leq \lfloor t \rfloor_b - (nm_{\max} + 1)2^{-b}$  then
    return  $x$ 
else
    return  $\perp$ 
end if

```

---

**Lemma 3.1.2.** *Gibt Algorithmus 2 eine Lösung zurück, so ist diese optimal bezüglich der nicht gerundeten Koeffizienten. Wenn er keine Lösung zurückgibt, so ist entweder das Verlierer-Gap oder das Gültigkeits-Gap nicht größer als  $(nm_{\max} + 1)2^{-b+1}$ .*

*Beweis.* Wir betrachten zunächst den Fall, dass Algorithmus 2 ein Optimum  $x$  zurückgibt, und möchten zeigen, dass  $x$  dann wirklich das Optimum bezüglich der nicht gerundeten Koeffizienten ist. Es können prinzipiell zwei Fehler eintreten, entweder ist das wirkliche Optimum durch das Runden ungültig geworden oder eine Lösung mit einem höheren Rang als das Optimum ist durch das Runden gültig geworden. Ersteres kann nicht eintreten, denn ist eine Lösung  $y \in \mathcal{S}$  gültig bezüglich der nicht gerundeten Koeffizienten, so gilt  $By \leq t$ . Durch das Runden wird das Gewicht in jeder Nebenbedingung um höchstens  $nm_{\max}2^{-b}$  und der Wert jeder Schranke um höchstens  $2^{-b}$  verändert, also gilt

$$\lfloor B \rfloor_b y \leq By + nm_{\max}2^{-b} \leq t + nm_{\max}2^{-b} \leq \lfloor t \rfloor_b + (nm_{\max} + 1)2^{-b}.$$

Wir müssen nun noch ausschließen, dass  $x$  erst durch das Runden gültig geworden ist. In diesem Fall gibt es eine Nebenbedingung  $j \in [k']$ , für die  $(Bx)_j > t_j$  gilt. Nach dem Runden gilt dann  $(\lfloor Bx \rfloor_b)_j > \lfloor t_j \rfloor_b - (nm_{\max} + 1)2^{-b}$ , es würde also  $x$  nicht als Optimum zurückgegeben werden. Somit ist der erste Teil des Lemmas gezeigt.

Wir betrachten nun den Fall, dass kein Optimum ausgegeben wird. Das passiert genau dann, wenn es ein  $j \in [k']$  mit  $(\lfloor B \rfloor_b x)_j > \lfloor t_j \rfloor_b - (nm_{\max} + 1)2^{-b}$  gibt. Wir unterscheiden die Fälle, dass  $x$  das tatsächliche Optimum ist oder eine Lösung, die bezüglich der nicht gerundeten Koeffizienten ungültig ist. Im ersten Fall muss das Gültigkeits-Gap kleiner als  $(nm_{\max} + 1)2^{-b+1}$  sein, denn für ein  $j \in [k']$  mit  $(\lfloor B \rfloor_b x)_j > \lfloor t_j \rfloor_b - (nm_{\max} + 1)2^{-b}$  gilt

$$t_j - (Bx)_j \leq \lfloor t_j \rfloor_b - (\lfloor B \rfloor_b x)_j + (nm_{\max} + 1)2^{-b} < (nm_{\max} + 1)2^{-b+1}.$$

Im zweiten Fall ist  $x$  erst durch das Runden und die Verschiebung der Schranken gültig geworden. Das bedeutet, dass das Verlierer-Gap nicht größer als  $(nm_{\max} + 1)2^{-b+1}$  sein kann, denn gäbe es ein  $j \in [k']$ , für das  $(Bx)_j > t_j + (nm_{\max} + 1)2^{-b+1}$  gilt, so wäre nach dem Runden

$$\begin{aligned}
 (\lfloor B \rfloor_b x)_j &\geq (Bx)_j - nm_{\max}2^{-b} \\
 &> t_j + (nm_{\max} + 1)2^{-b+1} - nm_{\max}2^{-b} \\
 &\geq \lfloor t_j \rfloor_b + (nm_{\max} + 1)2^{-b+1} - nm_{\max}2^{-b} - 2^{-b} \\
 &= \lfloor t_j \rfloor_b + (nm_{\max} + 1)2^{-b}.
 \end{aligned}$$

Somit wäre  $x$  im Widerspruch zur Voraussetzung auch nach dem Runden und dem Verschieben der Schranken noch ungültig.  $\square$



### 3.1.4 Ein Zertifizierer für stochastische Zielfunktionen und stochastische Nebenbedingungen

Wir betrachten nun den Fall, dass sowohl die Zielfunktion als auch mindestens eine Nebenbedingung von stochastischer Natur sind. In diesem Fall können wir auf die oben beschriebenen Zertifizierer zurückgreifen und müssen sie ineinander verschachteln. Das heißt konkret, dass wir Algorithmus 2 ausführen und den Aufruf von  $A$  gegen einen Aufruf von Algorithmus 1 ersetzen. Bei der Beschreibung von Algorithmus 1 sind wir davon ausgegangen, dass die Menge der erlaubten Lösungen beliebig feststeht. Da es aber in diesem Fall auch stochastische Nebenbedingungen gibt, deren Koeffizienten im Allgemeinen irrationale Zahlen sind, können wir Algorithmus 1 nicht ohne weiteres anwenden, sondern müssen vorher die Koeffizienten in den stochastischen Nebenbedingungen runden. Das machen wir genau wie in Algorithmus 2 und übergeben diese gerundeten Nebenbedingungen und die verschobenen Schranken an Algorithmus 1. Dieses Verfahren ist in Algorithmus 3 in Pseudocode dargestellt.

---

**Algorithmus 3** Zertifizierer (Zielfunktion und einige Nebenbedingungen sind stochastisch)

---

```

 $x = \text{ZertifizierterZF}(\lfloor B \rfloor_b, \lfloor t \rfloor_b + (nm_{\max} + 1)2^{-b})$ 
if  $x = \perp$  then
    return  $\perp$ 
end if
if  $\lfloor B \rfloor_b x \leq \lfloor t \rfloor_b - (nm_{\max} + 1)2^{-b}$  then
    return  $x$ 
else
    return  $\perp$ 
end if

```

---

In diesem Szenario ergibt sich die Menge der erlaubten Lösungen wieder als Schnitt einer fest vorgegebenen Menge  $\mathcal{S} \subseteq \mathcal{D}^n$  mit den Halbräumen  $\mathcal{B}_1, \dots, \mathcal{B}_{k'}$ , die durch die linearen Nebenbedingungen beschrieben werden. Da wir Algorithmus 1 mit gerundeten Koeffizienten und verschobenen Schranken aufrufen, ist für uns das Gewinner-Gap bezüglich dieser veränderten Nebenbedingungen interessant. Mit  $\Delta'$  bezeichnen wir im Folgenden das Gewinner-Gap, das sich ergibt, wenn die stochastischen Nebenbedingungen die Form  $\lfloor B \rfloor_b x \leq \lfloor t \rfloor_b + (nm_{\max} + 1)2^{-b}$  haben.

**Lemma 3.1.3.** *Gibt Algorithmus 3 eine Lösung zurück, so ist diese optimal bezüglich der nicht gerundeten Koeffizienten. Wenn er keine Lösung zurückgibt, so ist entweder  $\Delta' \leq nm_{\max}^2 2^{-b+4}$  oder das Verlierer-Gap oder das Gültigkeits-Gap sind nicht größer als  $(nm_{\max} + 1)2^{-b+1}$ .*

*Beweis.* Wir betrachten zunächst den Fall, dass eine Lösung  $x$  zurückgegeben wird. Der Aufruf von Algorithmus 1 sichert, dass  $x$  das Optimum bezüglich der nicht gerundeten Zielfunktion und der Nebenbedingungen  $\lfloor B \rfloor_b x \leq \lfloor t \rfloor_b + (nm_{\max} + 1)2^{-b}$  ist. Prinzipiell können zwei Fehler eintreten, entweder ist das wirkliche Optimum durch das Runden ungültig geworden oder eine Lösung mit einem höheren Rang als das Optimum ist durch das Runden gültig geworden. Analog zu Algorithmus 2 kann man argumentieren, dass der erste Fall nicht eintreten kann, weil die Menge der erlaubten Lösungen bezüglich der veränderten Nebenbedingungen

höchstens größer wird, und dass im zweiten Falle  $x$  verworfen worden wäre, weil dann nicht  $\lfloor B \rfloor_b x \leq \lfloor t \rfloor_b - (nm_{\max} + 1)2^{-b}$  gelten kann.

Auch der Beweis des zweiten Teils des Lemmas verläuft im Wesentlichen analog zum Beweis von Lemma 3.1.2. Wird kein Optimum zurückgegeben, so kann die Ursache darin liegen, dass Algorithmus 1 kein Optimum bezüglich der gerundeten Nebenbedingungen und der nicht gerundeten Zielfunktion berechnen kann. Aus Lemma 3.1.1 folgt, dass dieser Fall nur eintreten kann, wenn  $\Delta' \leq nm_{\max}^2 2^{-b+4}$  gilt. Ist das Gewinner-Gap nicht die Ursache dafür, dass kein Optimum zurückgegeben wird, so muss analog zu Lemma 3.1.2 entweder  $\Gamma \leq (nm_{\max} + 1)2^{-b+1}$  oder  $\Lambda \leq (nm_{\max} + 1)2^{-b+1}$  gelten.  $\square$

### 3.1.5 Vom pseudopolynomiellen Algorithmus zur polynomiellen geglätteten Komplexität

Nach dieser Vorarbeit sind wir nun dazu in der Lage, Theorem 1.3.7 zu beweisen. Als erstes werden wir zeigen, wie man einen Algorithmus mit pseudopolynomieller erwarteter Laufzeit dazu nutzen kann, um einen Algorithmus mit polynomieller geglätteter Laufzeit zu konstruieren.

*Beweis von Theorem 1.3.7, „ $\Leftarrow$ “.* Es sei ein Algorithmus  $A$  für  $\Pi_u^{\mathcal{H}}$  mit polynomieller erwarteter Laufzeit gegeben. Die erwartete Laufzeit von  $A$  ist also polynomiell in  $N$  und  $W$ , wobei  $N$  die Eingabegröße angebe und  $W$  den größten Absolutwert bezeichne, der von einem der stochastischen Koeffizienten angenommen wird. Es sei  $n$  die Anzahl der ganzzahligen Variablen und  $k$  die Anzahl der stochastischen Komponenten. Es gilt stets  $N \geq kn$ , weil die stochastischen Koeffizienten eine virtuelle Länge von 1 haben. Ferner bezeichne  $k'$  die Anzahl der stochastischen Nebenbedingungen. Es gilt  $k' \in \{k-1, k\}$ , je nachdem ob die Zielfunktion stochastisch ist oder nicht.  $E$  sei der absolute Erwartungswert der Verteilung, die durch die Dichte  $f$  beschrieben wird, die durch das Perturbationsmodell  $\mathcal{H}$  vorgegeben ist.

Auf Algorithmus  $A$  aufbauend konstruieren wir nun einen Algorithmus mit polynomieller geglätteter Laufzeit. Die wesentliche Idee dazu ist das *adaptive Runden*, das bereits in der Einleitung angesprochen wurde und in Algorithmus 4 noch einmal in Pseudocode dargestellt wird. Wir betrachten im Folgenden nur den Fall, dass sowohl die Zielfunktion als auch mindestens eine Nebenbedingung stochastisch sind. Die anderen beiden Fälle können analog behandelt werden. Für die Zielfunktion  $p$  gelte  $p(x) = c \cdot x$  mit  $c \in \mathbb{R}^n$  und die Nebenbedingungen haben die Form  $Bx \leq t$  mit  $B \in \mathbb{R}^{k' \times n}$  und  $t \in \mathbb{R}^{k'}$ .

---

#### Algorithmus 4 Adaptives Runden

---

$b = 1$

**repeat**

$c' = \lfloor c \rfloor_b, B' = \lfloor B \rfloor_b, t' = \lfloor t \rfloor_b + nm_{\max} 2^{-b}$

Multipliziere  $c', B'$  und  $t'$  mit  $2^b$ , um sie ganzzahlig zu machen.

Rufe  $A$  auf, um das Optimum  $x$  bezüglich  $c', B'$  und  $t'$  zu berechnen.

Überprüfe mit Algorithmus 3, ob  $x$  auch das Optimum bezüglich  $c, B$  und  $t$  ist.

$b = b + 1$

**until** Zertifizierung von  $x$  erfolgreich

---

### 3.1. Beweis von Theorem 1.3.7

---

Wir möchten nun zeigen, dass Algorithmus 4 eine polynomielle geglättete Laufzeit besitzt, und bedienen uns dabei der alternativen Charakterisierung von polynomieller geglätteter Komplexität, die in Lemma 1.3.3 gegeben ist. Wir werden also zeigen, dass ein Polynom  $P$  existiert, für das gilt

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \forall \varepsilon \in (0, 1] : \forall I \in \mathcal{I}_N : \Pr \left[ T_{\text{AR}}(\mathcal{H}(I, \phi)) \geq P \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \leq \varepsilon.$$

Dabei bezeichne  $T_{\text{AR}}$  eine Zufallsvariable, die die Laufzeit von Algorithmus 4 auf der gegebenen Eingabe angibt. Die Wahrscheinlichkeit wird also sowohl über die zufällige Eingabe als auch über die zufälligen Entscheidungen des Algorithmus gebildet.

Da die erwartete Laufzeit von Algorithmus  $A$  polynomiell in  $N$  und  $W$  ist, gibt es zwei Konstanten  $c_1, c_2 \in \mathbb{R}$  mit  $c_2 \geq 1$ , so dass für jede Eingabe  $I \in \mathcal{I}_N$ , bei der der größte vorkommende Absolutwert eines Koeffizienten  $W$  ist, gilt

$$\mathbf{E}[T_A(I)] \leq (c_1 N W)^{c_2},$$

wobei  $T_A$  eine Zufallsvariable ist, die die Laufzeit von Algorithmus  $A$  auf der gegebenen Eingabe angibt.

Mit Wahrscheinlichkeit 1 ist jedes der drei Gaps größer als 0. Also gibt es mit Wahrscheinlichkeit 1 ein  $b_0 < \infty$ , so dass  $b_0$  viele gelesene Bits ausreichen, um das Optimum bezüglich der nicht gerundeten Koeffizienten eindeutig zu bestimmen. Als Vorbereitung der Analyse berechnen wir zunächst für ein beliebiges  $b_0 \in \mathbb{N}$ , wie groß die erwartete Laufzeit  $\mathbf{E}[T_{\text{AR}}]$  des adaptiven Rundens ist, wenn die Schleife nach dem  $b_0$ -ten Durchlauf verlassen wird.

Algorithmus  $A$  wird in diesem Fall  $b_0$  mal aufgerufen, dabei ändert sich die Eingabelänge  $N$  nicht, denn das Runden und Skalieren betrifft nur die stochastischen Koeffizienten, deren Beitrag zur Eingabelänge fest ist. Die größte vorkommende Zahl  $W = W_1 W_2$  ändert sich hingegen von Aufruf zu Aufruf. Sie setzt sich dabei aus zwei Faktoren zusammen. Der erste Faktor ist  $W_1 = 2^b$ , er entsteht durch das Runden nach dem  $b$ -ten Bit und das anschließende Skalieren der Koeffizienten. Der zweite Faktor ist  $W_2$ , dieser gibt den absolut größten vorkommenden ganzzahligen Anteil eines Koeffizienten vor dem Skalieren an.  $W_2$  ist konstant,  $W_1$  ändert sich von Aufruf zu Aufruf.

$\mathcal{H}(I, \phi)$  ist eine Zufallsvariable, die die zufällige Perturbation der Eingabe  $I$  angibt. Mit  $[\mathcal{H}(I, \phi)]_b$  bezeichnen wir im Folgenden die zufällige Eingabe, die entsteht, wenn wir die stochastischen Koeffizienten in der Eingabe  $\mathcal{H}(I, \phi)$  nach dem  $b$ -ten Bit abrunden und anschließend so skalieren, dass sie ganze Zahlen sind.

In einem Schleifendurchlauf fallen nicht nur die Kosten für den Aufruf von Algorithmus  $A$  an, sondern auch Kosten für das Runden und Skalieren der Koeffizienten, diese werden wir pro Schleifendurchlauf mit  $cnk$  für eine geeignete Konstante  $c$  ansetzen. Für die erwartete

Laufzeit erhält man dann

$$\begin{aligned}
 \mathbf{E}[T_{\text{AR}}(I)] &= \sum_{b=1}^{b_0} (\mathbf{E}[T_A([\mathcal{H}(I, \phi)]_b)] + cnk) \\
 &= b_0 cnk + \sum_{b=1}^{b_0} \mathbf{E}[T_A([\mathcal{H}(I, \phi)]_b)] \\
 &\leq cb_0 N + \sum_{b=0}^{b_0} (c_1 N 2^b W_2)^{c_2} \\
 &= cb_0 N + (c_1 N W_2)^{c_2} \sum_{b=0}^{b_0} (2^{c_2})^b \\
 &\leq cb_0 N + (c_1 N 2^{b_0+1} W_2)^{c_2}. \tag{3.1}
 \end{aligned}$$

Um eine Aussage über die erwartete Laufzeit zu erhalten, müssen wir also untersuchen, wie groß  $b_0$  und  $W_2$  typischerweise werden. Wir betrachten zunächst  $W_2$ . Es sei  $r_{j,i}$  die Zufallsvariable, die zum  $i$ -ten Koeffizienten der  $j$ -ten stochastischen Komponente beim Perturbieren addiert wird. Weil jeder stochastische Koeffizient im ersten Schritt des semi-zufälligen Eingabemodells im Intervall  $[-1, 1]$  vorgegeben wird, gilt  $W_2 \leq \max_{i \in [n], j \in [k]} |r_{i,j}| + 1$ . Die absoluten Erwartungswerte von  $r_{j,i}$  sind laut Voraussetzung an das Perturbationsmodell  $\mathcal{H}$  durch eine Konstante  $E$  beschränkt. Mit der Markoff-Ungleichung erhalten wir für jedes  $\alpha \geq 1$

$$\begin{aligned}
 \Pr[W_2 > \alpha E + 1] &\leq \Pr\left[\max_{i \in [n], j \in [k]} |r_{j,i}| + 1 > \alpha E + 1\right] \\
 &\leq \Pr[\exists i \in [n], j \in [k] : |r_{j,i}| > \alpha E] \leq \frac{nk}{\alpha}.
 \end{aligned}$$

Setzt man  $W_2(N, \varepsilon) = 4NE/\varepsilon + 1$ , so erhält man wegen  $N \geq nk$  für alle  $\varepsilon \in (0, 1]$

$$\Pr[W_2 > W_2(N, \varepsilon)] \leq \frac{\varepsilon}{4}.$$

Wir untersuchen nun die Frage, wie groß man  $b$  wählen muss, damit die Wahrscheinlichkeit, dass das Optimum nach dem Lesen von  $b$  vielen Nachkommabits noch nicht eindeutig bestimmt ist, höchstens  $\varepsilon/2$  beträgt. Wir gehen davon aus, dass  $b$  viele Nachkommabits jedes stochastischen Koeffizienten gelesen werden. Diese Genauigkeit reicht nach Lemma 3.1.3 nur dann nicht aus, um die optimale Lösung eindeutig zu bestimmen, wenn eines der folgenden Ereignisse eintritt:

- Ereignis  $\mathcal{G}_1$ :  $\Delta' \leq nm_{\max}^2 2^{-b+4}$
- Ereignis  $\mathcal{G}_2$ :  $\Lambda \leq (nm_{\max} + 1)2^{-b+1}$
- Ereignis  $\mathcal{G}_3$ :  $\Gamma \leq (nm_{\max} + 1)2^{-b+1}$ .

Damit gilt für die Wahrscheinlichkeit, dass nach dem Aufdecken von  $b$  vielen Nachkommabits pro Koeffizient das Optimum noch nicht eindeutig feststeht,

$$\Pr[b_0 > b] \leq \Pr[\mathcal{G}_1] + \Pr[\mathcal{G}_2] + \Pr[\mathcal{G}_3].$$

### 3.1. Beweis von Theorem 1.3.7

---

Die Wahrscheinlichkeiten dieser Ereignisse lassen sich mit den Theoremen 2.2.2 und 2.4.3 abschätzen. Bei der Formulierung von Theorem 2.2.2 haben wir nur vorausgesetzt, dass die Menge der gültigen Lösungen beliebig feststeht, deswegen gilt seine Aussage auch für das Gewinner-Gap  $\Delta'$  bezüglich der Nebenbedingungen  $\lfloor B \rfloor_{bx} \leq \lfloor t \rfloor_b + (nm_{\max} + 1)2^{-b}$ .

Im Falle  $(nm_{\max} + 1)2^{-b+1} \leq (256n^5m^7m_{\max}\phi^2)^{-1}$  gilt

$$\begin{aligned} \Pr \left[ \Delta' \leq nm_{\max}^2 2^{-b+4} \right] &\leq n^2 m_{\max}^2 2^{-b+4} (3m - 4)\phi, \\ \Pr \left[ \Lambda \leq (nm_{\max} + 1)2^{-b+1} \right] &\leq 2k'(En^6m^7m_{\max}^2\phi^2 2^{-b+7})^{1/3}, \\ \Pr \left[ \Gamma \leq (nm_{\max} + 1)2^{-b+1} \right] &\leq 2k'(En^6m^7m_{\max}^2\phi^2 2^{-b+7})^{1/3}. \end{aligned}$$

Die Wahrscheinlichkeit, dass das Optimum nach  $b$  gelesenen Nachkommabits noch nicht eindeutig bestimmt ist, soll höchstens  $\varepsilon/2$  betragen. Um das zu erreichen, soll  $\Pr[\mathcal{G}_1] \leq \varepsilon/4$ ,  $\Pr[\mathcal{G}_2] \leq \varepsilon/8$  und  $\Pr[\mathcal{G}_3] \leq \varepsilon/8$  gelten, das führt auf die Ungleichungen

$$\begin{aligned} n^2 m_{\max}^2 2^{-b+4} (3m - 4)\phi &\leq \varepsilon/4, \\ 2k'(En^6m^7m_{\max}^2\phi^2 2^{-b+7})^{1/3} &\leq \varepsilon/8, \\ (nm_{\max} + 1)2^{-b+1} &\leq (256n^5m^7m_{\max}\phi^2)^{-1}. \end{aligned}$$

Formt man diese um, so erhält man

$$b \geq \log \left( \frac{2^{12} E k'^3 n^6 m^7 m_{\max}^2 \phi^2}{\varepsilon^3} \right) + 7 =: b(N, \phi, \varepsilon).$$

Zusammengefasst gilt also

$$\Pr[b_0 > \lceil b(N, \phi, \varepsilon) \rceil] \leq \varepsilon/2.$$

Wir definieren

$$W_1(N, \phi, \varepsilon) := \frac{2^{21} (E + 1) k'^3 n^6 m^7 m_{\max}^2 \phi^2}{\varepsilon^3} \geq 2^{\lceil b(N, \phi, \varepsilon) \rceil + 1}.$$

Unter der Voraussetzung, dass  $m$  und  $m_{\max}$  polynomiell beschränkt in  $n$  sind, ist  $W_1(N, \phi, \varepsilon)$  polynomiell in  $N$ ,  $\phi$  und  $1/\varepsilon$  beschränkt. Auch  $\lceil b(N, \phi, \varepsilon) \rceil$  lässt sich durch ein Polynom  $q(N, \phi, 1/\varepsilon)$  nach oben abschätzen.

Wir ersetzen nun in der Formel (3.1)  $b_0$ ,  $2^{b_0+1}$  und  $W_2$  durch die polynomiellen oberen Schranken  $q(N, \phi, 1/\varepsilon)$ ,  $W_1(N, \phi, \varepsilon)$  und  $W_2(N, \varepsilon)$  und multiplizieren das Ergebnis anschließend mit  $4/\varepsilon$ . Das Ergebnis ist ein Polynom  $P$ , das wie folgt aussieht

$$P \left( N, \phi, \frac{1}{\varepsilon} \right) = \frac{4}{\varepsilon} \left( cq \left( N, \phi, \frac{1}{\varepsilon} \right) N + (c_1 N W_1(N, \phi, \varepsilon) W_2(N, \varepsilon))^{c_2} \right).$$

Dieses Polynom besitzt die gewünschten Eigenschaften, denn übersteigt die Laufzeit des adaptiven Rundens die Schranke, die durch dieses Polynom gegeben ist, so ist entweder  $b_0 > \lceil b(N, \phi, \varepsilon) \rceil$  oder  $W_2 > W_2(N, \varepsilon)$  oder die Laufzeit von Algorithmus 4 ist mindestens

$(4/\varepsilon)$  mal so groß wie seine erwartete Laufzeit. Es gilt also für alle  $N \in \mathbb{N}$ , für alle  $\phi \geq 1$ , für alle  $\varepsilon \in (0, 1]$  und für alle  $I \in \mathcal{I}_N$

$$\begin{aligned} & \Pr \left[ T_{\text{AR}}(\mathcal{H}(I, \phi)) \geq P \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \\ & \leq \Pr [b_0 > \lceil b(N, \phi, \varepsilon) \rceil] + \Pr [W_2 > W_2(N, \varepsilon)] + \Pr \left[ T_{\text{AR}}(\mathcal{H}(I, \phi)) > \frac{4}{\varepsilon} \mathbf{E} [T_{\text{AR}}(\mathcal{H}(I, \phi))] \right] \\ & \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon. \end{aligned}$$

Somit ist gezeigt, dass Algorithmus 4 eine polynomielle geglättete Laufzeit hat.  $\square$

### 3.1.6 Von polynomieller geglätteter Komplexität zum pseudopolynomiellen Algorithmus

*Beweis von Theorem 1.3.7, „ $\Rightarrow$ “.* Wir müssen nun noch zeigen, wie man einen Algorithmus  $A$  mit polynomieller geglätteter Laufzeit für ein ganzzahliges lineares Optimierungsproblem  $\Pi$  nutzen kann, um einen randomisierten Algorithmus für das Problem  $\Pi_u^{\mathcal{H}}$  zu konstruieren, der eine polynomielle Laufzeit hat, der aber mit einer Wahrscheinlichkeit von höchstens  $1/2$  keine Lösung ausgibt.

Die Menge der Eingaben von  $\Pi_u^{\mathcal{H}}$  ist dahingehend eingeschränkt, dass die Koeffizienten in den als stochastisch ausgewählten Komponenten ganze Zahlen sein müssen. Wir werden eine kleine Perturbation dieser Koeffizienten durchführen und anschließend prüfen, ob durch die Perturbation das Optimum geändert wurde. Falls das Optimum geändert wurde, so brechen wir den Algorithmus ab und geben keine Lösung zurück, ansonsten nutzen wir Algorithmus  $A$ , um das Optimum zu berechnen.

Es sei eine beliebige Eingabe  $I$  für  $\Pi_u^{\mathcal{H}}$  gegeben. Bei der Definition der geglätteten Komplexität ganzzahliger linearer Optimierungsprobleme haben wir gefordert, dass alle Koeffizienten und Schranken in den stochastischen Komponenten vor dem Perturbieren im Intervall  $[-1, 1]$  liegen. Um das zu erreichen, skalieren wir alle stochastischen Komponenten in der Eingabe  $I$  zunächst mit  $M^{-1}$ , wobei  $M$  den größten vorkommenden Absolutwert eines der Koeffizienten oder einer der Schranken bezeichne. Nach dem Skalieren liegen alle Koeffizienten und Schranken in den stochastischen Komponenten im Intervall  $[-1, 1]$  und wir können eine Perturbation gemäß dem Perturbationsmodell  $\mathcal{H}$  durchführen. Diese Perturbation soll die Menge der erlaubten Lösungen und die optimale Lösung nicht verändern. Um das zu erreichen, erhöhen wir jede der Schranken in den stochastischen Nebenbedingungen um den Wert  $1/(2M)$ , das entspricht einer Erhöhung der Schranken um  $1/2$  in der nicht skalierten Eingabe  $I$ . Weil in  $I$  alle Gewichte ganzzahlig sind, ändert eine solche Veränderung der Schranken die Menge der erlaubten Lösungen nicht. Wenn wir nun eine Perturbation durchführen, bei der jeder stochastische Koeffizient um weniger als  $1/(2nm_{\max}M)$  verändert wird, so wird der Wert einer Lösung und jedes Gewicht um weniger als  $1/(2M)$  verändert. Deswegen ändert sich durch eine solche Perturbation weder die Menge der erlaubten Lösungen noch die Reihenfolge der Lösungen, die durch die Zielfunktion induziert wird. Eine Perturbation mit dieser Eigenschaft nennen wir *passend*.

Wir werden nun  $\phi$  so groß wählen, dass die Perturbation mit einer Wahrscheinlichkeit von mindestens  $3/4$  passend ist. Wir nutzen dabei die Eigenschaft aus, dass die absoluten Erwartungswerte der Zufallsvariablen, die zu den Koeffizienten addiert werden, endlich sind. Die

### 3.2. Allgemeinerere Perturbationsmodelle

---

Dichte dieser Zufallsvariablen sei  $f_\phi$ , sie entspricht einer Skalierung der Dichte  $f$ , die durch das Perturbationsmodell vorgegeben ist. Es gelte  $E = \int_{\mathbb{R}} |s|f(s) ds < \infty$ . Man rechnet leicht nach, dass für die skalierte Dichte  $f_\phi$  gilt  $\int_{\mathbb{R}} |s|f_\phi(s) ds = E/\phi$ . Sei  $r$  eine Zufallsvariable mit Dichte  $f_\phi$  mit  $\phi = 8n^2m_{\max}kEM$ . Mit der Markoff-Ungleichung erhält man

$$\Pr \left[ |r| > \frac{1}{2nm_{\max}M} \right] = \Pr [|r| > 4nk\mathbf{E}[|r|]] \leq \frac{1}{4nk}.$$

Da die Perturbation aus  $nk$  vielen Zufallsvariablen mit der Dichte  $f_\phi$  besteht, ist die Wahrscheinlichkeit, dass die Perturbation nicht passend ist, nicht größer als  $1/4$ .

Falls die Perturbation passend ist, wird Algorithmus  $A$  mit der perturbierten Eingabe aufgerufen. Da  $A$  eine polynomielle geglättete Laufzeit besitzt, gibt es insbesondere ein Polynom  $P(N, \phi)$ , so dass die Wahrscheinlichkeit, dass Algorithmus  $A$  auf der Eingabe  $\mathcal{H}(I, \phi)$  mehr als  $P(N, \phi)$  Rechenschritt benötigt, höchstens  $1/4$  beträgt. Wir lassen Algorithmus  $A$  nun  $P(N, \phi)$  viele Schritte laufen. Ist er danach nicht beendet, so brechen wir ihn ab und geben kein Ergebnis zurück. Ansonsten geben wir die Lösung zurück, die Algorithmus  $A$  berechnet hat. Da die Perturbation die optimale Lösung nicht verändert hat und Algorithmus  $A$  sich nie irrt, ist das auch die optimale Lösung der Probleminstanz  $I$ .

Es bezeichne  $\mathcal{Q}$ , das Ereignis, dass die Perturbation passend ist. Da für zwei beliebige Ereignisse  $A$  und  $B$  gilt  $\Pr[A \wedge B] \geq \Pr[A] + \Pr[B] - 1$ , können wir die Wahrscheinlichkeit, dass ein Optimum zurückgegeben wird, abschätzen durch

$$\Pr[\mathcal{Q} \wedge (T_A(\mathcal{H}(I, \phi)) \leq P(N, \phi))] \geq \Pr[\mathcal{Q}] - \Pr[T_A(\mathcal{H}(I, \phi)) > P(N, \phi)] \geq \frac{1}{2}.$$

Die Wahrscheinlichkeit, keine Lösung zurückzugeben, ist also höchstens  $1/2$ . Die Laufzeit dieses Verfahrens ist wegen  $\phi = O(n^2m_{\max}kM)$  pseudopolynomiell, wenn  $m_{\max}$  polynomiell in  $n$  beschränkt ist.  $\square$

## 3.2 Allgemeinerere Perturbationsmodelle

### 3.2.1 Unterschiedlich perturbierter Koeffizienten

Ein Perturbationsmodell für ein ganzzahliges lineares Optimierungsproblem besteht, so wie wir es in Kapitel 1 definiert haben, aus einer Auswahl von stochastischen Komponenten und einer Wahrscheinlichkeitsdichte  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  mit  $\sup_{s \in \mathbb{R}} f(s) = 1$  und einem endlichem absoluten Erwartungswert  $E = \int_{\mathbb{R}} |s|f(s) ds < \infty$ . Eine  $\phi$ -Perturbation erfolgt dann dadurch, dass zu jedem Koeffizienten in den stochastischen Komponenten unabhängig eine Zufallsvariable addiert wird, deren Dichte  $f_\phi$  durch eine Skalierung aus  $f$  hervorgeht. Es seien  $k$  viele Komponenten als stochastisch ausgewählt und es gebe  $n$  ganzzahlige Variablen, dann bezeichne  $r_{j,i}$  für  $j \in [k]$  und  $i \in [n]$  die Zufallsvariable, die zum  $i$ -ten Koeffizienten der  $j$ -ten stochastischen Komponente addiert wird.

Bei der Analyse der Gaps in Kapitel 2 wird an keiner Stelle ausgenutzt, dass die Verteilungen der  $r_{j,i}$  identisch sind. Beim Beweis von Theorem 1.3.7 geht diese Eigenschaft jedoch an einer Stelle ein, es wird nämlich vorausgesetzt, dass der absolute Erwartungswert der  $r_{j,i}$  durch

eine von  $n$  unabhängige Konstante  $E$  nach oben beschränkt ist. Diese Konstante  $E$  geht polynomiell in die Laufzeit der konstruierten Algorithmen ein.

Unsere Analyse erlaubt es, Theorem 1.3.7 auf eine allgemeinere Klasse von Perturbationsmodellen zu erweitern, nämlich solche, bei denen die Verteilungen der  $r_{j,i}$  verschieden sein dürfen, die absoluten Erwartungswerte jedoch durch eine Konstante nach oben beschränkt sind.

Wir stellen diese erweiterte Klasse von Perturbationsmodellen nun formal dar. Für  $E > 0$  sei dazu  $\mathcal{F}(E)$  die Menge aller Dichten  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  mit  $\sup_{s \in \mathbb{R}} f(s) = 1$  und einem absoluten Erwartungswert, der nicht größer als  $E$  ist. Ein Perturbationsmodell  $\mathcal{H}$  ist beschrieben durch eine Auswahl von  $k \geq 1$  vielen Komponenten, die stochastisch sein sollen, eine Konstante  $E > 0$  und für jede stochastische Komponente eine Folge von Dichten aus  $\mathcal{F}(E)$ . Für  $j \in [k]$  sei die Folge  $f^{(j)} : \mathbb{N} \rightarrow \mathcal{F}(E)$  gegeben. Die Dichte der Zufallsvariablen  $r_{j,i}$ , die bei einer  $\phi$ -Perturbation zu dem  $i$ -ten Koeffizienten der  $j$ -ten stochastischen Komponente addiert wird, sei  $(f_i^{(j)})_\phi$ , wobei für alle  $s \in \mathbb{R}$  gelte  $(f_i^{(j)})_\phi(s) = \phi f_i^{(j)}(\phi s)$ .

### 3.2.2 Nullerhaltende Perturbationen

Ein Kritikpunkt an der geglätteten Analyse des Simplex-Algorithmus, der bereits in der Arbeit [9] von Spielman und Teng diskutiert wird, ist, dass durch die Perturbationen Koeffizienten, die durch den Gegner auf Null gesetzt werden, verändert werden. Dadurch kann bei vielen Problemen die Problemstruktur zerstört werden. Die gleiche Kritik lässt sich an unserer Analyse ganzzahliger linearer Optimierungsprobleme üben, wir werden jedoch zeigen, dass sich unsere Analyse auch dann noch anwenden lässt, wenn jeder Koeffizient in einer stochastischen Komponente entweder fest auf Null gesetzt wird oder durch eine Perturbation entsteht. Der Gegner im semi-zufälligen Eingabemodell wird also dadurch gestärkt, dass er Nullen vorgeben kann, die nicht perturbiert werden.

Wir betrachten zunächst den Fall, dass es genau eine stochastische Nebenbedingung gibt und zeigen, dass Theorem 2.3.17 auch dann noch gilt, wenn der Gegner in dieser Nebenbedingung Koeffizienten fest auf Null setzen darf. Der Gegner setze  $n - l < n$  viele Koeffizienten fest auf Null, wir können o. B. d. A. davon ausgehen, dass dies die Koeffizienten  $w_{l+1}, \dots, w_n$  sind. Es gilt also  $w = (w_1, \dots, w_l, 0, \dots, 0)$ . Wir bezeichnen zwei Lösungen  $x, y \in \mathcal{D}^n$  als *äquivalent*, wenn sie in den ersten  $l$  Komponenten übereinstimmen, in diesem Falle stimmt auch das Gewicht von  $x$  und  $y$  stets überein. Durch diese Relation wird die Menge der erlaubten Lösungen  $\mathcal{S} \subseteq \mathcal{D}^n$  in Äquivalenzklassen partitioniert. Weil die Gewichte aller Lösungen aus derselben Äquivalenzklasse stets übereinstimmen, sind entweder alle Lösungen aus einer Klasse gültig oder gar keine Lösung aus der Klasse. Deswegen können nur solche Lösungen pareto-optimal sein, die in ihrer Äquivalenzklasse den höchsten Rang haben. Diese Lösungen wählen wir als Repräsentanten für die Äquivalenzklassen und entfernen alle anderen Lösungen aus  $\mathcal{S}$ . Weiterhin entfernen wir die letzten  $n - l$  Variablen und erhalten somit ein neues Optimierungsproblem  $\Pi'$  mit einer Menge von erlaubten Lösungen  $\mathcal{S}' \subseteq \mathcal{D}^l$ . Das Gewicht einer Lösung  $x' \in \mathcal{S}'$  stimmt mit dem Gewicht der ihr entsprechenden Lösung  $x \in \mathcal{S}$  überein.

Die Reihenfolge der Lösungen in  $\mathcal{S}'$ , die durch die Zielfunktion von  $\Pi'$  vorgegeben wird, entspreche der Reihenfolge, die durch die Zielfunktion von  $\Pi$  induziert wird. Dann ist eine Lösung im Problem  $\Pi$  genau dann der Gewinner, wenn die zu ihr gehörige Lösung  $x'$  der Gewinner



### 3.3. Ganzzahlige Packungs- und Überdeckungsprobleme

---

in  $\Pi'$  ist. Das gleiche gilt für den minimalen Verlierer. Aus diesem Grunde stimmen auch das Verlierer- bzw. das Gültigkeits-Gap von  $\Pi$  und  $\Pi'$  überein. Es lässt sich Korollar 2.3.18 auf  $\Pi'$  anwenden und aus den oben genannten Gründen übertragen sich die Aussagen über das Verlier- und das Gültigkeits-Gap auf  $\Pi$ .

Wir betrachten nun den Fall, dass es  $k \geq 1$  viele stochastische Nebenbedingungen gibt. Da der Beweis von Theorem 2.4.3 nur auf Korollar 2.3.18 beruht, überträgt sich somit auch dieses Theorem auf den Fall, dass der Gegner in jeder Nebenbedingung Koeffizienten auf Null setzen darf, die nicht perturbiert werden.

Wir wollen nun noch zeigen, dass sich unsere Analyse auch dann noch anwenden lässt, wenn die Zielfunktion stochastisch ist und der Gegner einige ihrer Koeffizienten fest auf Null setzen darf. Auch in diesem Falle gehen wir davon aus, dass der Gegner die Koeffizienten  $c_{l+1}, \dots, c_n$  fest auf Null gesetzt hat, und teilen die Menge der erlaubten Lösungen  $\mathcal{S}$  gemäß der oben beschriebenen Relation in Äquivalenzklassen ein. Alle Lösungen aus derselben Äquivalenzklasse haben dann den gleichen Zielfunktionswert. Zunächst betrachten wir wieder das reduzierte Problem  $\Pi'$  und wenden darauf Theorem 2.2.2 an. In diesem Falle können die Repräsentanten der Äquivalenzklassen, die in  $\Pi'$  übernommen werden, beliebig gewählt werden, weil alle Lösungen aus einer Äquivalenzklasse gleich gut sind.

Das Gewinner-Gap des Problems  $\Pi'$  beschreibt bezogen auf das ursprüngliche Problem  $\Pi$ , wie groß die Differenz der Werte der besten und zweitbesten Äquivalenzklasse ist. In der Äquivalenzklasse, die die besten Lösungen enthält, können sehr viele Lösungen liegen, das heißt, das eigentliche Gewinner-Gap des Problems  $\Pi$ , also der Abstand der besten zur zweitbesten Lösung, ist im Allgemeinen Null. Wir können jedoch den Zertifizierer für stochastische Zielfunktionen so anpassen, dass er trotzdem noch wie gewünscht funktioniert. Statt bei Algorithmus 1 zu überprüfen, ob die Lösungen bezüglich der Koeffizienten  $[c]_b$  und  $\bar{c}$  übereinstimmen, muss lediglich geprüft werden, ob sie in der gleichen Äquivalenzklasse liegen. Nur falls das nicht der Fall ist, kann die Lösung  $x'$  nicht zertifiziert werden. Ferner dürfen bei Algorithmus 1 nur die ersten  $l$  Schleifendurchläufe durchgeführt werden, weil die letzten  $n - l$  Koeffizienten fest auf Null gesetzt wurden. Für den so modifizierten Zertifizierer ist also das Gewinner-Gap des Problems  $\Pi'$  entscheidend und nicht das von  $\Pi$ .

### 3.3 Ganzzahlige Packungs- und Überdeckungsprobleme

In diesem Abschnitt werden wir mit Hilfe von Theorem 1.3.7 die geglättete Komplexität von ganzzahligen Packungs- und Überdeckungsproblemen bestimmen. Es handelt sich dabei um zwei wichtige Klassen von ganzzahligen Optimierungsproblemen.

**Definition 3.3.1 (Packungs- und Überdeckungsproblem).** *Ein ganzzahliges Programm mit den Variablen  $x = (x_1, \dots, x_n)^T \in \mathbb{N}_0^n$  heißt **Packungsprogramm**, falls die lineare Zielfunktion  $c \cdot x$  mit  $c \in \mathbb{R}_{\geq 0}^n$  maximiert werden soll und die Menge der erlaubten Lösungen durch lineare Nebenbedingungen der Form  $Ax \leq t$  beschrieben ist, wobei  $A \in \mathbb{R}_{\geq 0}^{k \times n}$  und  $t \in \mathbb{R}_{\geq 0}^k$  gilt. Es heißt **Überdeckungsprogramm**, falls die lineare Zielfunktion minimiert werden soll und die Nebenbedingungen die Form  $Ax \geq t$  für  $A \in \mathbb{R}_{\geq 0}^{k \times n}$  und  $t \in \mathbb{R}_{\geq 0}^k$  haben.*

*Ein ganzzahliges Optimierungsproblem  $\Pi$  heißt **ganzzahliges Packungsproblem**, falls jede*

*Probleminstanz von  $\Pi$  ein Packungsprogramm ist.  $\Pi$  heißt **ganzzahliges Überdeckungsproblem**, falls jede Probleminstanz von  $\Pi$  ein Überdeckungsprogramm ist.*

Ganzzahlige Packungs- und Überdeckungsprobleme mit einer konstanten Anzahl von Nebenbedingungen können mit Hilfe von dynamischer Programmierung in pseudopolynomieller Zeit gelöst werden. Die Anzahl der Nebenbedingungen geht jedoch exponentiell in die Laufzeit dieser Algorithmen ein.

Um die geglättete Komplexität von ganzzahligen Packungs- und Überdeckungsproblemen zu bestimmen, müssen wir zunächst eine Klasse von Perturbationsmodellen festlegen, bezüglich derer die geglättete Komplexität bestimmt werden soll. Wir wählen die Klasse, die auch der Analyse von Theorem 1.3.7 zu Grunde liegt oder eine der Erweiterungen, die im letzten Abschnitt vorgestellt wurden. Für unsere Analyse genügt es bei Packungsproblemen zu verlangen, dass mindestens eine Nebenbedingung stochastisch ist, bei Überdeckungsproblemen müssen jedoch für die Analyse alle Nebenbedingungen stochastisch sein. Ferner muss das Perturbationsmodell so gewählt werden, dass nach dem Perturbieren immer noch alle Koeffizienten nicht negativ sind, damit durch das Perturbieren die Klasse der Packungs- bzw. Überdeckungsprobleme nicht verlassen wird.

Um unsere Analyse der geglätteten Komplexität anwenden zu können, müssen wir zunächst die stochastischen Nebenbedingungen und die Zielfunktion, sofern auch sie stochastisch ist, des betrachteten Problems so skalieren, dass die größte vorkommende Zahl in jeder stochastischen Komponente 1 ist. Wir können Theorem 1.3.7 an dieser Stelle nicht direkt anwenden, weil die Voraussetzung, dass die Größe des Wertebereiches der ganzzahligen Variablen polynomiell in der Anzahl der Variablen  $n$  beschränkt ist, nicht erfüllt ist, denn es gilt  $\mathcal{D} = \mathbb{N}_0$ . Wir werden jedoch zeigen, dass sich der Wertebereich der ganzzahligen Variablen mit großer Wahrscheinlichkeit polynomiell in  $n$  beschränken lässt.

**Lemma 3.3.2.** *Es sei  $\Pi$  ein ganzzahliges Packungsproblem und  $I$  eine Probleminstanz mit  $n$  ganzzahligen Variablen. Ferner sei  $\varepsilon \in (0, 1]$  beliebig und  $\mathcal{H}(I, \phi)$  bezeichne die Zufallsvariable, die eine  $\phi$ -Perturbation von  $I$  beschreibt. Die Wahrscheinlichkeit, dass es in der optimalen Lösung von  $\mathcal{H}(I, \phi)$  eine Variable gibt, die einen Wert annimmt, der größer als  $\varepsilon^{-1}$  ist, beträgt höchstens  $\varepsilon n \phi$ .*

*Beweis.* Mindestens eine Nebenbedingung in  $\Pi$  ist von stochastischer Natur, diese hat die Form  $w_1 x_1 + \dots + w_n x_n \leq t$ . Da die Nebenbedingung so skaliert wurde, dass die größte vorkommende Zahl 1 ist, können wir o. B. d. A. davon ausgehen, dass  $t = 1$  gilt und jeder Koeffizient  $w_i$  im Intervall  $[0, 1]$  liegt, denn gäbe es einen Koeffizienten  $w_i > t$ , so wäre zwangsläufig  $x_i = 0$  in jeder erlaubten Lösung und wir könnten die  $i$ -te Variable aus dem Problem entfernen.

Nach dem Skalieren und Perturbieren hat die Nebenbedingung somit die Form  $w'_1 x_1 + \dots + w'_n x_n \leq 1$ . Da wir die Klasse der Packungsprobleme nicht verlassen möchten, muss auch nach dem Perturbieren stets  $w'_i \geq 0$  gelten. Damit erhält man

$$\Pr [\exists i \in [n] : w'_i \leq \varepsilon] \leq \sum_{i=1}^n \Pr [w'_i \in [0, \varepsilon]] \leq \varepsilon n \phi.$$

### 3.3. Ganzzahlige Packungs- und Überdeckungsprobleme

---

Falls jeder Koeffizient größer als  $\varepsilon$  ist, ist der Wert jeder ganzzahligen Variable in jeder gültigen Lösung durch  $\varepsilon^{-1}$  nach oben beschränkt, weil sonst die Gewichtsschranke überschritten würde. Damit folgt das Lemma.  $\square$

**Lemma 3.3.3.** *Es sei  $\Pi$  ein ganzzahliges Überdeckungsproblem mit  $k$  vielen Nebenbedingungen und  $I$  eine Probleminstanz mit  $n$  ganzzahligen Variablen. Ferner sei  $\varepsilon \in (0, 1]$  beliebig und  $\mathcal{H}(I, \phi)$  bezeichne die Zufallsvariable, die eine  $\phi$ -Perturbation von  $I$  beschreibt. Die Wahrscheinlichkeit, dass es in der optimalen Lösung von  $\mathcal{H}(I, \phi)$  eine Variable gibt, die einen Wert annimmt, der größer als  $\lceil \varepsilon^{-1} \rceil$  ist, beträgt höchstens  $\varepsilon nk\phi$ .*

*Beweis.* Alle Nebenbedingungen in  $\Pi$  sind von stochastischer Natur und nach dem Skalieren liegen alle Zahlen, die in den Nebenbedingungen vorkommen, in dem Intervall  $[0, 1]$ . Es bezeichne  $w_{j,i}$  den  $i$ -ten Koeffizienten der  $j$ -ten Nebenbedingung und  $t_j$  die Schranke der  $j$ -ten Nebenbedingung. In Abhängigkeit der  $w_{j,i}$  und der Schranken  $t_j$  wollen wir die Werte der ganzzahligen Variablen in einer optimalen Lösung beschränken. Wir betrachten die  $i$ -te Variable  $x_i$  und setzen  $u_i = \max_{j \in [k]} \lceil 1/w_{j,i} \rceil$ . Wählt man  $x_i = u_i$ , so sind wegen  $t_j \leq 1$  alle Nebenbedingungen erfüllt. Da die Zielfunktion minimiert werden soll, macht es also unter keinen Umständen Sinn,  $x_i > u_i$  zu wählen.

Das bedeutet, dass der Wertebereich der ganzzahligen Variablen auf die Menge  $\{0, \dots, \lceil \varepsilon^{-1} \rceil\}$  eingeschränkt ist, falls  $w_{j,i} \geq \varepsilon$  für alle  $j \in [k]$  und  $i \in [n]$  gilt. Für die Wahrscheinlichkeit, dass dieses Ereignis nach dem Perturbieren nicht eintritt, erhält man

$$\Pr[\exists j \in [k], i \in [n] : w_{j,i} \leq \varepsilon] \leq \sum_{j \in [k]} \sum_{i \in [n]} \Pr[w_{j,i} \in [0, \varepsilon]] \leq \varepsilon nk\phi.$$

Somit ist das Lemma gezeigt.  $\square$

**Theorem 3.3.4 (Die geglättete Komplexität ganzzahliger Packungs- und Überdeckungsprobleme).** *Ganzzahlige Packungsprobleme, bei denen mindestens eine Nebenbedingung stochastisch ist, besitzen eine polynomielle geglättete Komplexität. Ganzzahlige Überdeckungsprobleme besitzen eine polynomielle geglättete Komplexität, falls jede Nebenbedingung von stochastischer Natur ist.*

*Beweis.* Beim Beweis von Theorem 1.3.7 gehen  $m$  und  $m_{\max}$  polynomiell in die Laufzeit des adaptiven Rundens ein. Wir haben dort angenommen, dass  $m_{\max}$  und damit auch  $m$  polynomiell in  $n$  beschränkt sind. Wir haben gezeigt, dass es ein Polynom  $P$  gibt, so dass für die Laufzeit  $T_{\text{AR}}$  des adaptiven Rundens gilt

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \forall \varepsilon \in (0, 1] : \forall I \in \mathcal{I}_N : \Pr \left[ T_{\text{AR}}(\mathcal{H}(I, \phi)) \geq P \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \leq \varepsilon.$$

Weil  $m \leq 2m_{\max}$  gilt, können wir  $P$  o. B. d. A. in der Form  $P(N, \phi, 1/\varepsilon) = m_{\max}^c p(N, \phi, 1/\varepsilon)$  schreiben, wobei  $p$  von  $\mathcal{D}$  unabhängig ist. Wir betrachten zunächst Packungsprobleme mit einer stochastischen Nebenbedingung und setzen

$$P' \left( N, \phi, \frac{1}{\varepsilon} \right) = \left( \frac{2n\phi}{\varepsilon} \right)^c p \left( N, 2\phi, \frac{1}{2\varepsilon} \right).$$

Das Ereignis, dass alle Koeffizienten der stochastischen Nebenbedingung größer als  $\alpha$  sind, bezeichnen wir im Folgenden mit  $\mathcal{K}(\alpha)$ . Es gilt

$$\begin{aligned} & \Pr \left[ T_{\text{AR}}(\mathcal{H}(I, \phi)) \geq P' \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \\ \leq & \Pr \left[ -\mathcal{K} \left( \frac{\varepsilon}{2n\phi} \right) \right] + \Pr \left[ T_{\text{AR}}(\mathcal{H}(I, \phi)) \geq P' \left( N, \phi, \frac{1}{\varepsilon} \right) \mid \mathcal{K} \left( \frac{\varepsilon}{2n\phi} \right) \right] \end{aligned} \quad (3.2)$$

Analog zu Lemma 3.3.2 rechnet man nach, dass die erste dieser Wahrscheinlichkeiten durch  $\varepsilon/2$  nach oben beschränkt ist. Wir betrachten nun die bedingte Wahrscheinlichkeit. Unter der Bedingung  $\mathcal{K}(\varepsilon/(2n\phi))$  können wir nach Lemma 3.3.2 den Wertebereich der ganzzahligen Variablen auf die Menge  $\mathcal{D} = \{0, \dots, \lfloor 2n\phi/\varepsilon \rfloor\}$  beschränken, ohne dabei die optimale Lösung zu verändern. Die Verteilungen der Koeffizienten  $w_i$  werden durch diese Bedingung jedoch nur geringfügig geändert. Es sei  $f_i$  eine Dichte, durch die die Verteilung des Koeffizienten  $w_i$  beschrieben ist, und  $\bar{f}_i$  eine Dichte, durch die die Verteilung von  $w_i$  unter der Bedingung  $\mathcal{K}(\varepsilon/(2n\phi))$  beschrieben ist. Dann gilt

$$\begin{aligned} \bar{f}_i(x) &= \begin{cases} 0 & \text{falls } x \in [0, \varepsilon/(2n\phi)] \\ \frac{f_i(x)}{1 - \Pr[w_i \in [0, \varepsilon/(2n\phi)]]} & \text{sonst} \end{cases} \\ &\leq \begin{cases} 0 & \text{falls } x \in [0, \varepsilon/(2n\phi)] \\ 2f_i(x) & \text{sonst} \end{cases} . \end{aligned}$$

Unter der Bedingung  $\mathcal{K}(\varepsilon/(2n\phi))$  erhöhen wir die maximale Dichte  $\phi$  und den absoluten Erwartungswert also um höchstens den Faktor 2. Wir können somit auf Theorem 1.3.7 zurückgreifen, um die bedingte Wahrscheinlichkeit in (3.2) durch  $\varepsilon/2$  abzuschätzen. Damit haben wir insgesamt gezeigt, dass für das Polynom  $P'$  gilt

$$\forall N \in \mathbb{N} : \forall \phi \geq 1 : \forall \varepsilon \in (0, 1] : \forall I \in \mathcal{I}_N : \Pr \left[ T_{\text{AR}}(\mathcal{H}(I, \phi)) \geq P' \left( N, \phi, \frac{1}{\varepsilon} \right) \right] \leq \varepsilon.$$

Somit besitzen Packungsprobleme mit mindestens einer stochastischen Nebenbedingung eine polynomielle geglättete Komplexität. Die Aussage über die Überdeckungsprobleme kann analog gezeigt werden.  $\square$

## KAPITEL 4

# Ausblick

Wir haben in dieser Arbeit die geglättete Komplexität von ganzzahligen Optimierungsproblemen untersucht. Die Resultate geben einen Hinweis darauf, welche Probleme der betrachteten Klasse in der Praxis effizient gelöst werden können. Es gibt jedoch einige Kritikpunkte, die man an unserer Analyse anbringen kann, und viele offene Fragen im Themenbereich dieser Diplomarbeit.

- Ein Kritikpunkt an unserer Analyse betrifft das Perturbationsmodell. Die Idee der geglätteten Komplexität ist es, dass die Veränderungen, die durch eine Perturbation erfolgen, klein sind. Das ist bei unserem Modell nur teilweise erfüllt, denn eine stochastische Komponente wird zunächst so skaliert, dass die größte vorkommende Zahl 1 ist, und anschließend wird zu jedem Koeffizienten eine Zufallsvariable addiert, deren absoluter Erwartungswert  $E/\phi$  ist. Da  $\phi$  polynomiell in die Laufzeit des adaptiven Rundens eingeht, sind für praktische Anwendungen nur solche Werte von  $\phi$  interessant, die polynomiell in der Eingabegröße beschränkt sind. Betrachten wir den Fall  $E = 1$  und  $\phi = n^k$ . Ist in diesem Falle die größte vorkommende Zahl vor dem Skalieren  $2^n$ , so bedeutet eine Änderung eines Koeffizienten um  $r$  in der skalierten Komponente bezogen auf das ursprüngliche Problem eine Änderung um  $r2^n$ . Liegt  $r$  in der Größenordnung  $1/n^k$ , so ist das eine Änderung, die für große  $n$  zwar klein ist in Relation zu dem unperturbierten Wert  $2^n$ , die aber extrem groß ist bezogen auf Zahlen, die vor dem Skalieren bereits klein waren. Es wäre interessant, *relative Perturbationen* zu untersuchen, also solche, bei denen die Änderung jedes Koeffizienten klein bezogen auf seine Größe ist. Unser Perturbationsmodell sorgt nämlich dafür, dass nach dem Perturbieren alle Zahlen in etwa in der gleichen Größenordnung liegen.
- Weiterhin wäre es interessant die Zufallsvariablen Gewinner-, Verlierer- und Gültigkeits-Gap weiter zu untersuchen und die unteren Schranken für die Größe, die wir in den Theoremen 2.2.2 und 2.4.3 erhalten haben, durch passende obere Schranken zu ergänzen.
- Beier und Vöcking haben das binäre Rucksackproblem untersucht und eine Schranke für die erwartete Anzahl von pareto-optimalen Rucksackfüllungen gezeigt, falls entweder die Gewichte oder die Profite der Objekte unabhängig zufällig gewählt werden [1]. Sie haben gezeigt, dass es im Erwartungswert nur polynomiell viele pareto-optimale Rucksackfüllungen gibt. Daraus ergibt sich sofort, dass das Rucksackproblem in diesem Falle mit dem *Nemhauser/Ullman Algorithmus* im Erwartungswert in polynomieller Zeit gelöst werden kann. Es wäre interessant, zu untersuchen, ob sich die Anzahl der pareto-optimalen Rucksackfüllungen auch dann noch beschränken lässt, wenn jedes Objekt auch mehrfach eingepackt werden darf, wenn also der Wertebereich der ganzzahligen Variablen von  $\{0, 1\}$  auf  $\{0, \dots, m - 1\}$  erweitert wird.

- Das adaptive Runden ist wegen seiner Laufzeit nicht besonders praktikabel. Bei pseudopolynomiellen Algorithmen für ganzzahlige Optimierungsprobleme geht in der Regel die Anzahl der Nebenbedingungen exponentiell in die Laufzeit ein, weil diese Algorithmen in der Regel auf dynamischer Programmierung beruhen. In der Praxis beobachtet man jedoch, dass auch solche Instanzen von Packungsproblemen effizient gelöst werden können, die sehr viele Nebenbedingungen enthalten. Solche Probleme treten in vielen konkreten Anwendungen auf, beispielsweise bei der Konstruktion optimierender Compiler [10]. Die vorherrschenden Algorithmen, die dazu benutzt werden, sind *Branch and Bound-Algorithmen*. Eine Erklärung, warum diese Algorithmen in praktischen Anwendungen so effizient sind, wäre wünschenswert.

# Grundlagen der Wahrscheinlichkeitstheorie

In vielen Bereichen der Informatik spielen randomisierte Algorithmen und zufällige Prozesse eine immer größere Rolle, deswegen sind Vorlesungen, in denen Grundlagen der Stochastik vermittelt werden, mittlerweile fester Bestandteil eines Informatikstudiums. In diesen Vorlesungen werden aber in der Regel nur diskrete Wahrscheinlichkeitsräume ausführlich betrachtet, weil diskrete Modelle in der Informatik eine größere Bedeutung haben als kontinuierliche. Bei der Analyse ganzzahliger linearer Optimierungsprobleme begegnen wir jedoch oft kontinuierlichen Wahrscheinlichkeitsräumen, da die stochastischen Koeffizienten der Zielfunktion und der Nebenbedingungen reelle Zufallsvariablen sind. Deswegen werden in diesem Anhang die für diese Arbeit wesentlichen Begriffe über nicht diskrete Wahrscheinlichkeitstheorie explizit dargestellt. Dabei wird auf Beweise verzichtet und einige Sätze werden nicht in voller Allgemeinheit formuliert, sondern nur so allgemein wie es für diese Arbeit notwendig ist.

Dieser Anhang basiert im Wesentlichen auf den Büchern über Wahrscheinlichkeitstheorie von Patrick Billingsley [3], William Feller [5] und Ulrich Krengel [6]. Ausführliche Details und Beweise der hier dargestellten Lemmata und Sätze können in diesen Büchern nachgelesen werden.

## A.1 Wahrscheinlichkeitsräume

Um ein mathematisches Modell eines *Zufallsexperimentes* zu entwickeln, ist es grundlegend, festzulegen, welches die möglichen Ergebnisse des Experimentes sind. Die Menge dieser Ergebnisse bezeichnet man üblicherweise mit  $\Omega$ . Im allgemeinen Fall stellt man keine Forderung an  $\Omega$ , insbesondere ist nicht die Abzählbarkeit von  $\Omega$  verlangt. Die Elemente aus  $\Omega$  heißen *Elementarereignisse*.

Im diskreten Fall ist es üblich, jeder Teilmenge von  $\Omega$  eine Wahrscheinlichkeit zuzuordnen, wir werden jedoch im Folgenden sehen, dass das im Allgemeinen nur eingeschränkt möglich ist. Deswegen betrachten wir auch *Wahrscheinlichkeitsmaße*, die nur auf einer Teilmenge  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  definiert sind. Die Teilmenge  $\mathcal{A}$  darf jedoch nicht beliebig sein, sondern muss gewisse Eigenschaften erfüllen, damit wir weiterhin auf den Elementen von  $\mathcal{A}$  abzählbare Mengenoperationen durchführen können, ohne  $\mathcal{A}$  dabei zu verlassen. Diese Eigenschaften sind im Begriff der  $\sigma$ -Algebra zusammengefasst.

**Definition A.1.1.** Es sei  $\Omega \neq \emptyset$  beliebig. Eine Menge  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  heißt  **$\sigma$ -Algebra**, wenn gilt:

$$\Omega \in \mathcal{A}, \tag{A.1}$$

$$A \in \mathcal{A} \Rightarrow \neg A := \Omega \setminus A \in \mathcal{A}, \tag{A.2}$$

$$A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}. \tag{A.3}$$

Aus den drei Eigenschaften einer  $\sigma$ -Algebra ergeben sich weitere Eigenschaften, insbesondere der Abschluss gegen abzählbar viele Durchschnittsbildungen.

Auf einer  $\sigma$ -Algebra  $\mathcal{A}$  können *Wahrscheinlichkeitsmaße* definiert werden, das sind Funktionen, die den Ereignissen aus  $\mathcal{A}$  Wahrscheinlichkeiten zuordnen.

**Definition A.1.2.** Ein **messbarer Raum** ist ein Paar  $(\Omega, \mathcal{A})$  bestehend aus einer nicht leeren Menge  $\Omega$  und einer  $\sigma$ -Algebra  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ . Ein **Wahrscheinlichkeitsmaß**  $P$  ist eine Funktion  $P : \mathcal{A} \rightarrow [0, 1]$  mit  $P(\Omega) = 1$ , die  **$\sigma$ -additiv** ist, das heißt, für disjunkte  $A_1, A_2, \dots \in \mathcal{A}$  muss gelten

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

$(\Omega, \mathcal{A}, P)$  heißt dann **Wahrscheinlichkeitsraum** und  $P$  auch **Wahrscheinlichkeitsverteilung**. Teilmengen  $A \subseteq \Omega$ , die zu  $\mathcal{A}$  gehören, heißen **Ereignisse**.

Von besonderem Interesse im Kontext dieser Arbeit ist der Fall  $\Omega = \mathbb{R}$ . Offenbar ist  $\mathcal{P}(\mathbb{R})$  eine  $\sigma$ -Algebra und kann mit einem Wahrscheinlichkeitsmaß  $P$  gemäß Definition A.1.2 versehen werden. Der folgende Satz zeigt jedoch, dass ein solches Wahrscheinlichkeitsmaß einer grundlegenden Einschränkung unterliegt.

**Theorem A.1.3.** Unter der Voraussetzung der Gültigkeit der Kontinuumshypothese, die ebenso wie ihre Negation mit den Axiomen der Mengenlehre vereinbar ist, gibt es keine auf ganz  $\mathcal{P}(\mathbb{R})$  definierte  $\sigma$ -additive Funktion  $P$  mit  $P(\mathbb{R}) = 1$ , die jeder aus einem einzigen Punkt bestehenden Menge  $\{\omega\}$  den Wert  $P(\{\omega\}) = 0$  zuordnet. Unter dieser Voraussetzung gibt es also auf  $\mathcal{P}(\mathbb{R})$  nur diskrete Wahrscheinlichkeitsmaße.

Die Kontinuumshypothese sagt aus, dass es keine überabzählbare Teilmenge der reellen Zahlen gibt, die eine kleinere Mächtigkeit als  $\mathbb{R}$  hat. Sie kann ebenso wie ihre Negation der Mengenlehre als Axiom hinzugefügt werden, ohne dadurch einen Widerspruch zu erzeugen.

Da es aber gerade die nicht diskreten Wahrscheinlichkeitsmaße sind, die uns interessieren, und da auch nur solche durch *Dichten* beschrieben werden können, dürfen wir keine Wahrscheinlichkeitsmaße auf  $\mathcal{P}(\mathbb{R})$  betrachten, sondern nur solche auf geeigneten Teilmengen  $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R})$ . Eine typische Wahl ist die *borelsche  $\sigma$ -Algebra*  $\mathcal{B}(\mathbb{R})$ ; das ist die  $\sigma$ -Algebra über  $\mathbb{R}$ , die von den halboffenen Intervallen  $(a, b]$  mit  $a, b \in \mathbb{R}$  erzeugt wird.  $\mathcal{B}(\mathbb{R})$  ist also die kleinste  $\sigma$ -Algebra, die jedes dieser Intervalle enthält. Die Mengen aus  $\mathcal{B}(\mathbb{R})$  heißt auch *borelsche Mengen* oder *Borel-Mengen*. Es ist jedes abgeschlossene Intervall borelsch, jede abzählbare Vereinigung offener und abgeschlossener Intervalle und jede geometrisch vorstellbare Teilmenge von  $\mathbb{R}$ . Es



können nicht borelsche Mengen konstruiert werden, diese spielen jedoch für unsere Betrachtungen keine Rolle. In analoger Weise definiert man auch auf  $\mathbb{R}^n$  eine borelsche  $\sigma$ -Algebra  $\mathcal{B}(\mathbb{R}^n)$  als die durch die halboffenen Intervalle  $(a_1, b_1] \times \dots \times (a_n, b_n]$  erzeugte  $\sigma$ -Algebra. Auch die Mengen aus  $\mathcal{B}(\mathbb{R}^n)$  werden wir im Folgenden als borelsch bezeichnen.

Zur Berechnung von Wahrscheinlichkeiten ist das folgende Lemma nützlich.

**Lemma A.1.4.** *Es sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Ist  $B_1 \subseteq B_2 \subseteq \dots$  eine wachsende Folge von Ereignissen und  $B$  deren Vereinigung, so gilt*

$$P(B) = \lim_{i \rightarrow \infty} P(B_i). \quad (\text{A.4})$$

*Ist  $C_1 \supseteq C_2 \supseteq \dots$  eine fallende Folge von Ereignissen und  $C$  deren Durchschnitt, so gilt*

$$P(C) = \lim_{i \rightarrow \infty} P(C_i). \quad (\text{A.5})$$

Zum Schluss dieses Abschnitts wollen wir noch Aussagen über die Wahrscheinlichkeit einer endlichen Vereinigung von Ereignissen treffen.

**Lemma A.1.5.** *Es sei  $(\Omega, \mathcal{A}, P)$  ein beliebiger Wahrscheinlichkeitsraum und  $A_1, \dots, A_n \in \mathcal{A}$  seien beliebige Ereignisse. Für  $k \in [n]$  sei*

$$S_k = \sum_{\substack{i_1 < \dots < i_k \\ i_1 \in [n], \dots, i_k \in [n]}} P(A_{i_1} \cap \dots \cap A_{i_k}).$$

*Es gilt dann*

$$P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n (-1)^{i-1} S_i.$$

*Für jedes gerade  $k \in [n]$  gilt*

$$P(A_1 \cup \dots \cup A_n) \geq \sum_{i=1}^k (-1)^{i-1} S_i$$

*und für jedes ungerade  $k \in [n]$  gilt*

$$P(A_1 \cup \dots \cup A_n) \leq \sum_{i=1}^k (-1)^{i-1} S_i.$$

*Diese beiden Ungleichungen sind auch als die **Ungleichungen von Boole und Bonferroni** bekannt.*

## A.2 Verteilungen und Dichten

Ein Wahrscheinlichkeitsmaß  $P$  auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  ist eindeutig beschrieben, wenn die Wahrscheinlichkeiten der Erzeuger, also der Intervalle  $(a, b]$  festgelegt sind. Dies kann im reellen Fall durch eine Verteilungsfunktion geschehen.

**Definition A.2.1.** Eine Funktion  $F : \mathbb{R} \rightarrow [0, 1]$  heißt **Verteilungsfunktion**, wenn sie rechtsstetig und monoton wachsend ist und wenn  $F(x) \rightarrow 0$  für  $x \rightarrow -\infty$  und  $F(x) \rightarrow 1$  für  $x \rightarrow \infty$  gilt.

Ist  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathcal{B}(\mathbb{R})$  und setzt man  $F(x) = P((-\infty, x])$ , so ist  $F$  eine Verteilungsfunktion. Umgekehrt kann man von einer beliebigen Verteilungsfunktion  $F$  ausgehen und erhält ein Wahrscheinlichkeitsmaß  $P$ , indem man  $P((a, b]) = F(b) - F(a)$  für jedes Intervall  $(a, b]$  setzt. Da die borelsche  $\sigma$ -Algebra von den halboffenen Intervallen erzeugt wird, ist  $P$  dadurch vollständig festgelegt.

**Definition A.2.2.** Eine nicht negative Lebesgue-integrierbare Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  mit

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

heißt **Dichte**.

Jede Dichte  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  beschreibt eine Verteilungsfunktion  $F : \mathbb{R} \rightarrow [0, 1]$  durch

$$F(x) = \int_{-\infty}^x f(t) dt. \tag{A.6}$$

Ist  $F$  eine beliebige stetige Verteilungsfunktion auf  $\mathbb{R}$  und bilden die Stellen  $C \subseteq \mathbb{R}$ , an denen  $F$  nicht differenzierbar ist, eine Nullmenge, so wird durch

$$f(x) := \frac{d}{dx} F(x) \text{ für } x \in \mathbb{R} \setminus C$$

eine Dichte  $f$  definiert, die  $F$  beschreibt; diese kann auf  $C$  beliebig fortgesetzt werden. Die zu  $F$  gehörende Dichte ist auch an den Stellen aus  $\mathbb{R} \setminus C$  nicht eindeutig bestimmt, denn sie kann auf einer beliebigen Nullmenge undefiniert werden, ohne dabei den Wert des Integrals in Formel (A.6) zu verändern.

### A.3 Eindimensionale Zufallsvariablen

Oftmals sind wir nicht an dem Ergebnis des Zufallsexperimentes interessiert, sondern nur an einer charakteristischen Kenngröße. Werfen wir beispielsweise zwei Würfel, so ist unter Umständen nur die Summe der Augenzahlen entscheidend. Deswegen betrachten wir Abbildungen, die aus  $\Omega$  nur die gewünschten Informationen extrahieren. Solche Abbildungen können durch *Zufallsvariablen* realisiert werden, wobei für diese Arbeit nur reelle Zufallsvariablen relevant sind.

**Definition A.3.1.** Es sei  $(\Omega, \mathcal{A})$  ein messbarer Raum. Eine Abbildung  $X : \Omega \rightarrow \mathbb{R}$  heißt **reelle Zufallsvariable**, falls für jede borelsche Menge  $A \subset \mathbb{R}$  gilt  $X^{-1}(A) \in \mathcal{A}$ .

Die Bedingung, die an reelle Zufallsvariablen gestellt wird, lässt sich dadurch erklären, dass wir ein Wahrscheinlichkeitsmaß  $P$ , das auf  $(\Omega, \mathcal{A})$  definiert ist, auf den Raum  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  übertragen möchten. Dafür ist es notwendig, dass die Urbilder der borelschen Mengen in  $\mathcal{A}$  liegen, denn dann kann ein Wahrscheinlichkeitsmaß  $P_X$  auf  $\mathbb{R}$  dadurch definiert werden, dass  $P_X(A) = P(X^{-1}(A))$  für jede Borel-Menge  $A$  gesetzt wird.

Die Verteilung einer reellwertigen Zufallsvariable  $X$  kann durch ihre *Verteilungsfunktion*  $F_X : \mathbb{R} \rightarrow [0, 1]$  mit  $F_X(x) := P_X((-\infty, x])$  eindeutig beschrieben werden. Wenn das zu Grunde liegende Wahrscheinlichkeitsmaß  $P_X$  aus dem Kontext hervorgeht, so werden wir statt  $P_X((-\infty, x])$  auch  $\Pr[X \leq x]$  schreiben und für borelsche Mengen  $A \subseteq \mathbb{R}$  statt  $P_X(A)$  auch  $\Pr[X \in A]$ . Kann die Verteilungsfunktion  $F_X$  der Zufallsvariablen  $X$  durch eine Dichte  $f_X$  beschrieben werden, so heißt  $X$  auch *stetige Zufallsvariable*.

Es wird nun der Begriff der *Unabhängigkeit* zweier Zufallsvariablen definiert. Anschaulich sind zwei Zufallsvariablen  $X$  und  $Y$  unabhängig, wenn die Kenntnis von  $X$  keine Rückschlüsse irgendeiner Art auf  $Y$  zulässt und umgekehrt.

**Definition A.3.2.** *Zwei reelle Zufallsvariablen  $X : \Omega \rightarrow \mathbb{R}$  und  $Y : \Omega \rightarrow \mathbb{R}$  heißen **unabhängig**, falls für alle Wahlen von  $A_1, A_2 \in \mathcal{B}(\mathbb{R})$  gilt*

$$P(X^{-1}(A_1) \cap Y^{-1}(A_2)) = P_X(A_1) \cdot P_Y(A_2).$$

*In der oben eingeführten Notation lässt sich diese Bedingung auch schreiben als*

$$\Pr[X \in A_1 \wedge Y \in A_2] = \Pr[X \in A_1] \cdot \Pr[Y \in A_2].$$

Oftmals spielen auch Transformationen von Zufallsvariablen eine Rolle. Sind  $X$  und  $Y$  Zufallsvariablen, die beide durch Dichten beschrieben sind, so ist die Frage nach einer Dichte von  $X+Y$  und von anderen skalierten und verschobenen Varianten von Interesse. Die wesentlichen Spezialfälle sind in folgendem Lemma festgehalten.

**Lemma A.3.3.** *Es seien  $X_1$  und  $X_2$  unabhängige reelle Zufallsvariablen und  $a, c \in \mathbb{R}$  beliebige Konstanten mit  $c \neq 0$ . Die Dichte von  $X_1$  sei  $f_1$ , die von  $X_2$  sei  $f_2$ .*

1. *Die Funktion  $g_1 : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  mit  $g_1(x) = f_1(x - a)$  für alle  $x \in \mathbb{R}$  ist eine Dichte der Zufallsvariablen  $Y_1 = X_1 + a$ .*
2. *Eine Dichte  $g_2$  der Zufallsvariablen  $Y_2 = cX_1$  erhält man, indem man*

$$g_2(x) = \frac{1}{|c|} f_1\left(\frac{x}{c}\right)$$

*für alle  $x \in \mathbb{R}$  setzt.*

3. *Eine Dichte  $g_3$  der Zufallsvariablen  $Y_3 = X_1 + X_2$  erhält man durch die Faltung von  $f_1$  und  $f_2$ , man setzt also für jedes  $x \in \mathbb{R}$*

$$g_3(x) = \int_{-\infty}^{\infty} f_1(x - t) f_2(t) dt.$$

Ein weiterer wesentlicher Begriff ist der *Erwartungswert* einer Zufallsvariablen.

**Definition A.3.4.** *Sei  $X$  eine reelle Zufallsvariable mit Dichte  $f$ . Falls  $\int_{-\infty}^{\infty} |x|f(x) dx$  existiert und endlich ist, so heißt*

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

*der **Erwartungswert** von  $X$  und*

$$E(|X|) = \int_{-\infty}^{\infty} |x|f(x) dx$$

*der **absolute Erwartungswert** von  $X$ .*

An vielen Stellen wird sich die folgende Abschätzung als nützlich erweisen.

**Lemma A.3.5 (Markoff-Ungleichung).** *Es sei  $X$  eine reelle Zufallsvariable, für die stets  $X \geq 0$  gilt, und  $k > 0$  beliebig. Dann gilt*

$$\Pr[X \geq k\mathbf{E}[X]] \leq \frac{1}{k}.$$

## A.4 Zweidimensionale Zufallsvariablen

Das Konzept reeller Zufallsvariablen kann kanonisch auf mehr als eine Dimension verallgemeinert werden. Sei dazu  $(\Omega, \mathcal{A})$  ein beliebiger messbarer Raum. Eine Abbildungen  $Z : \Omega \rightarrow \mathbb{R}^n$  heißt *n-dimensionale reelle Zufallsvariable*, falls für jede borelsche Menge  $A \in \mathcal{B}(\mathbb{R}^n)$  gilt  $Z^{-1}(A) \in \mathcal{A}$ .

Sind  $X : \Omega \rightarrow \mathbb{R}$  und  $Y : \Omega \rightarrow \mathbb{R}$  reelle Zufallsvariablen, so ist  $Z : \Omega \rightarrow \mathbb{R}^2$  mit  $Z(x) = (X(x), Y(x))$  für alle  $x \in \Omega$  eine zweidimensionale reelle Zufallsvariable. Ein Wahrscheinlichkeitsmaß  $P$  auf dem Raum  $(\Omega, \mathcal{A})$  wird im zweidimensionalen Fall durch  $Z$  auf den Raum  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$  übertragen. Es ergibt sich das Maß  $P_Z : \mathcal{B}(\mathbb{R}^2) \rightarrow [0, 1]$  mit  $P_Z(A) = P(Z^{-1}(A))$  für alle  $A \in \mathcal{B}(\mathbb{R}^2)$ .

Genauso wie im eindimensionalen Fall kann das Wahrscheinlichkeitsmaß  $P_Z$  durch eine Verteilungsfunktion  $F_Z : \mathbb{R}^2 \rightarrow \mathbb{R}$  mit  $F_Z(x, y) = P_Z((-\infty, x] \times (-\infty, y])$  eindeutig beschrieben werden, diese wird auch *gemeinsame Verteilung* von  $X$  und  $Y$  genannt. Wenn  $P_Z$  aus dem Kontext hervor geht, benutzen wir auch hier wieder die Schreibweise  $F_Z(x, y) = \Pr[X \leq x \wedge Y \leq y]$ .

**Definition A.4.1.** *Ein besonders wichtiger Spezialfall ist der, dass  $F_Z$  durch eine Dichte  $f_Z : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$  beschrieben werden kann.  $f_Z$  heißt **gemeinsame Dichte** von  $X$  und  $Y$ , wenn für alle  $a, b \in \mathbb{R}$  gilt*

$$F(a, b) = \int_{-\infty}^a \int_{-\infty}^b f(x, y) dy dx.$$

Durch die gemeinsame Dichte sind zwei reelle Zufallsvariablen vollständig beschrieben, insbesondere lässt sich daran erkennen, ob sie unabhängig sind.

**Satz A.4.2.** *Zwei reelle Zufallsvariablen  $X$  und  $Y$  mit den Dichten  $f_X$  und  $f_Y$  sind genau dann unabhängig, wenn die Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  mit*

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

*eine gemeinsame Dichte der zweidimensionalen reellen Zufallsvariable  $Z = (X, Y)$  ist.*

**Definition und Lemma A.4.3.** *Sei  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  eine gemeinsame Dichte der reellen Zufallsvariablen  $X$  und  $Y$ . Die Funktionen  $f_X$  und  $f_Y$  mit  $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$  und  $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$  heißen **Randdichten**. Mit diesen Randdichten gilt*

$$\begin{aligned} \Pr[X \leq a] &= \Pr[X \leq a \wedge Y < \infty] = \int_{-\infty}^a \left( \int_{-\infty}^{\infty} f(x, y) dy \right) dx = \int_{-\infty}^a f_X(x) dx \\ \Pr[Y \leq a] &= \Pr[X < \infty \wedge Y \leq a] = \int_{-\infty}^a \left( \int_{-\infty}^{\infty} f(x, y) dx \right) dy = \int_{-\infty}^a f_Y(y) dy. \end{aligned}$$

Ähnlich wie im eindimensionalen Fall, wo für reelle Zufallsvariablen  $X_1$  und  $X_2$  und für Konstanten  $a, c \in \mathbb{R}$  die Dichten von  $X_1 + X_2$ ,  $X_1 + a$  und  $cX_1$  bestimmt wurden, interessieren auch im zweidimensionalen Fall die Dichten von Zufallsvariablen, die durch eine Transformation aus  $(X_1, X_2)$  hervorgehen, diese können mit Hilfe des folgenden Satzes bestimmt werden.

**Satz A.4.4.** *Es seien  $X_1$  und  $X_2$  zwei reelle Zufallsvariablen mit der gemeinsamen Dichte  $f_X : \mathbb{R}^2 \rightarrow \mathbb{R}$  und es sei  $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  eine bijektive Transformation mit den Komponenten  $\Phi_1$  und  $\Phi_2$ . Ferner seien  $\Phi_1^{-1}$  und  $\Phi_2^{-1}$  die stetig differenzierbaren Komponenten von  $\Phi^{-1}$ . Für  $y = (y_1, y_2) \in \mathbb{R}^2$  sei  $d(y)$  die Determinante der Jacobimatrix von  $\Phi^{-1}$ , also*

$$d(y) = \det \begin{pmatrix} \frac{\partial \Phi_1^{-1}}{\partial y_1}(y) & \frac{\partial \Phi_1^{-1}}{\partial y_2}(y) \\ \frac{\partial \Phi_2^{-1}}{\partial y_1}(y) & \frac{\partial \Phi_2^{-1}}{\partial y_2}(y) \end{pmatrix}.$$

Für jedes  $y \in \mathbb{R}^2$  gelte  $d(y) \neq 0$  und die Zufallsvariablen  $Y_1$  und  $Y_2$  seien definiert durch  $Y_1 = \Phi_1(X_1, X_2)$  und  $Y_2 = \Phi_2(X_1, X_2)$ . Dann ist eine gemeinsame Dichte  $f_Y$  von  $Y_1$  und  $Y_2$  gegeben durch

$$f_Y(y) = |d(y)| \cdot f_X(\Phi^{-1}(y)).$$

## A.5 Bedingte Dichten und bedingte Verteilungen

Es gibt viele Situationen, in denen man Zufallsvariablen unter bestimmten Bedingungen betrachtet. Bei einem Würfel kann man sich beispielsweise die Frage stellen, wie groß die Wahrscheinlichkeit ist, eine 2 zu würfeln, unter der Bedingung, dass eine gerade Zahl gewürfelt wird. Es gilt in diesem Falle  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , das Ereignis, das uns interessiert, ist  $A = \{2\}$  und die Bedingung ist  $B = \{2, 4, 6\}$ . Für die bedingte Wahrscheinlichkeit gilt im diskreten Fall

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{2, 4, 6\})} = \frac{1}{3}. \quad (\text{A.7})$$

Im diskreten Fall ist es nicht sinnvoll, Wahrscheinlichkeiten unter einer Bedingung  $B$  mit  $P(B) = 0$  zu betrachten, weil eine solche Bedingung nie eintritt. Im kontinuierlichen Fall ist das anders, dort bedeutet  $P(B) = 0$  nicht, dass das Ereignis  $B$  nicht eintreten kann, und deswegen ist es sinnvoll, sich auch solche bedingte Wahrscheinlichkeiten anzuschauen. Nehmen wir als Beispiel einen Telefonanruf, dessen Dauer sich aus der Zeit, die man in der Warteschleife verbringt, und der eigentlichen Gesprächsdauer zusammensetzt. In einem einfachen, nicht besonders realitätsnahen Modell könnte die Wartezeit uniform zufällig aus dem Intervall zwischen 0 und 5 Minuten gewählt werden, und die Gesprächszeit unabhängig von der Wartezeit uniform zufällig aus dem Intervall zwischen 4 und 6 Minuten. Es macht nun anschaulich durchaus Sinn, sich nach der erwarteten Dauer des Anrufes zu fragen unter der Bedingung, dass man genau 3 Minuten in der Warteschleife verbringt, auch wenn dieses Ereignis die Wahrscheinlichkeit 0 hat. Anschaulich ist klar, dass man als erwartete Anrufdauer in diesem Falle 8 Minuten erhalten sollte.

Offenbar lässt sich Formel (A.7) in einem solchen Szenario nicht mehr anwenden, stattdessen gilt die folgende Verallgemeinerung.

**Definition und Satz A.5.1.** *Es seien  $X$  und  $Y$  zwei stetige reelle Zufallsvariablen mit der gemeinsamen Dichte  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  und es sei  $y \in \mathbb{R}$  so gewählt, dass*

$$f_Y(y) := \int_{-\infty}^{\infty} f(x, y) dx > 0$$

*gilt. Dann heißt die Funktion  $f_{X|Y=y} : \mathbb{R} \rightarrow \mathbb{R}$  mit*

$$f_{X|Y=y}(x) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} = \frac{f(x, y)}{f_Y(y)}$$

*bedingte Dichte von  $X$  unter der Bedingung  $Y = y$ . Es handelt sich dabei um eine Dichte nach Definition A.2.2 und für jede Borel-Menge  $A \in \mathbb{R}$  gilt*

$$\Pr[X \in A | Y = y] = \int_A f_{X|Y=y}(t) dt.$$

Wir wollen nun noch den Satz über *totale Wahrscheinlichkeit* und eine leichte Folgerung daraus im kontinuierlichen Fall formulieren.

**Satz A.5.2.** *Es seien  $X$  und  $Y$  zwei stetige reelle Zufallsvariablen mit der gemeinsamen Dichte  $f$ . Ferner sei  $A \subseteq \mathbb{R}$  eine Borel-Menge und  $f_Y$  bezeichne die Randdichte von  $Y$ . Dann gilt*

$$\Pr[X \in A] = \int_{-\infty}^{\infty} \Pr[X \in A | Y = y] f_Y(y) dy.$$

**Korollar A.5.3.** *Es seien  $X$  und  $Y$  zwei stetige reelle Zufallsvariablen mit der gemeinsamen Dichte  $f$ . Ferner seien  $A \subseteq \mathbb{R}$  und  $B \subseteq \mathbb{R}$  Borel-Mengen mit  $\Pr[Y \in B] > 0$ . Dann gilt*

$$\Pr[X \in A | Y \in B] = \frac{\int_B \Pr[X \in A | Y = y] f_Y(y) dy}{\Pr[Y \in B]},$$

*wobei  $f_Y$  die Randdichte von  $Y$  bezeichne.*

Wir werden den Satz über totale Wahrscheinlichkeit nun noch weiter verallgemeinern.

**Satz A.5.4.** *Es seien  $X$  und  $Y$  eindimensionale reelle Zufallsvariablen und es sei  $Z$  eine  $n$ -dimensionale reelle Zufallsvariable, so dass  $Y$  und  $Z$  unabhängig sind. Es existiere eine gemeinsame Dichte  $f : \mathbb{R}^{n+2} \rightarrow \mathbb{R}$  von  $X, Y$  und  $Z$  und es bezeichnen  $f_Y$  und  $f_Z$  die Randdichten von  $Y$  bzw.  $Z$ . Dann gilt für jede Borel-Menge  $A \in \mathcal{B}(\mathbb{R})$  und für jedes  $y \in \mathbb{R}$  mit  $f_Y(y) > 0$*

$$\Pr[X \in A | Y = y] = \int_{\mathbb{R}^n} \Pr[X \in A | Y = y \wedge Z = z] f_Z(z) dz,$$

*wobei wir  $\Pr[X \in A | Y = y \wedge Z = z] = 0$  für solche  $z \in \mathbb{R}$  mit  $f_Z(z) = 0$  definieren.*

## Ergänzung zum Beweis von Lemma 2.3.10

An dieser Stelle möchten wir den Beweis der Abschätzung

$$\int_{M_{i,j}^{*,m_3,m_4}(p)} 1 \, db \leq 4m_{\max}s$$

nachholen.

Dazu betrachten wir noch einmal die Menge  $M_{i,j}^{*,m_3,m_4}(p)$ , es ist

$$\begin{aligned} M_{i,j}^{*,m_3,m_4}(p) &= \left\{ b \in \mathbb{R} \mid 0 < f_{A_{i,j}^{m_3,m_4}}(b) \leq \frac{1}{2n^2m^2m_{\max}sp} \right\} \\ &\subseteq \left\{ b \in \mathbb{R} \mid f_{A_{i,j}^{m_3,m_4}}(b) > 0 \right\}. \end{aligned}$$

**1. Fall:**  $m_3 = 0$  und  $m_4 \neq 0$

In diesem Fall gilt  $f_{A_{i,j}^{m_3,m_4}} = f_{m_4w_j}$ . Damit ergibt sich

$$\begin{aligned} M_{i,j}^{*,m_3,m_4}(p) &\subseteq \{ b \in \mathbb{R} \mid f_{m_4w_j}(b) > 0 \} \\ &= \left\{ b \in \mathbb{R} \mid \frac{1}{|m_4|} \cdot f_j\left(\frac{b}{m_4}\right) > 0 \right\} \\ &\subseteq \left\{ b \in \mathbb{R} \mid -s \leq \frac{b}{m_4} \leq s \right\} \\ &= [-s|m_4|, s|m_4|]. \end{aligned}$$

Also gilt in diesem Fall

$$\int_{M_{i,j}^{*,m_3,m_4}(p)} 1 \, db \leq 2s|m_4| \leq 2m_{\max}s,$$

wie gewünscht.

**2. Fall:**  $m_3 \neq 0$  und  $m_4 = 0$

Dieser Fall kann analog zum ersten Fall behandelt werden.

### Vorbereitung der weiteren Fälle

In den noch nicht behandelten Fällen gilt  $m_3 \neq 0$  und  $m_4 \neq 0$ . Die Dichte  $f_{A_{i,j}^{m_3,m_4}}(b)$  lässt sich dann auch folgendermaßen schreiben:

$$\begin{aligned} f_{A_{i,j}^{m_3,m_4}}(b) &= f_{m_3w_i+m_4w_j}(b) \\ &= \int_{-\infty}^{\infty} f_{m_3w_i}(x)f_{m_4w_j}(b-x) dx \\ &= \frac{1}{|m_3| \cdot |m_4|} \cdot \int_{-\infty}^{\infty} f_i\left(\frac{x}{m_3}\right) f_j\left(\frac{b-x}{m_4}\right) dx \\ &= \frac{1}{|m_4|} \cdot \int_{-\infty}^{\infty} f_i(x) \cdot f_j\left(\frac{b-m_3x}{m_4}\right) dx. \end{aligned}$$

Im Falle  $m_3 \neq 0$  und  $m_4 \neq 0$  gilt somit

$$\begin{aligned} M_{i,j}^{*,m_3,m_4}(p) &\subseteq \left\{ b \in \mathbb{R} \mid \frac{1}{|m_4|} \cdot \int_{-\infty}^{\infty} f_i(x) \cdot f_j\left(\frac{b-m_3x}{m_4}\right) dx > 0 \right\} \\ &\subseteq \left\{ b \in \mathbb{R} \mid \exists x \in \mathbb{R} : f_i(x) \cdot f_j\left(\frac{b-m_3x}{m_4}\right) > 0 \right\} \\ &= \left\{ b \in \mathbb{R} \mid \exists x \in \mathbb{R} : f_i(x) > 0 \wedge f_j\left(\frac{b-m_3x}{m_4}\right) > 0 \right\} \\ &\subseteq \left\{ b \in \mathbb{R} \mid \exists x \in \mathbb{R} : -s \leq x \leq s \wedge -s \leq \frac{b-m_3x}{m_4} \leq s \right\}. \end{aligned} \quad (\text{B.1})$$

### 3. Fall: $m_3 \neq 0, m_4 \neq 0$ und $m_3 \cdot m_4 > 0$

Wir beginnen diesen Fall damit, die zweite Ungleichung in (B.1) umzuformen. Es gilt

$$-s \leq \frac{b-m_3x}{m_4} \leq s \iff -\frac{m_4}{m_3}s + \frac{b}{m_3} \leq x \leq \frac{m_4}{m_3}s + \frac{b}{m_3}.$$

Nimmt man nun beide Ungleichungen in (B.1) zusammen, so erhält man für  $x$  die unteren Schranken

$$x_{l,1} = -s \text{ und } x_{l,2} = -\frac{m_4}{m_3}s + \frac{b}{m_3}$$

und die oberen Schranken

$$x_{u,1} = s \text{ und } x_{u,2} = \frac{m_4}{m_3}s + \frac{b}{m_3}.$$

Für vorgegebenes  $b$  darf  $x$  also in das Intervall  $I = [\max\{x_{l,1}, x_{l,2}\}, \min\{x_{u,1}, x_{u,2}\}]$  fallen. Ist dieses Intervall leer, so gibt es kein geeignetes  $x$ . Um zu bestimmen, wann dieses Intervall leer wird, lösen wir zunächst die Gleichungen  $x_{l,1} = x_{u,2}$  und  $x_{l,2} = x_{u,1}$  jeweils nach  $b$  auf. Es gilt

$$x_{l,1} = x_{u,2} \iff b = -(m_3 + m_4)s$$

und

$$x_{l,2} = x_{u,1} \iff b = (m_3 + m_4)s.$$



---

**Unterfall 3a:**  $m_3 > 0$  und  $m_4 > 0$

In diesem Unterfall gilt

$$\begin{aligned} b < -(m_3 + m_4)s &\Rightarrow x_{l,1} > x_{u,2} \text{ und} \\ b > (m_3 + m_4)s &\Rightarrow x_{l,2} > x_{u,1}. \end{aligned}$$

Zusammenfassend können wir also festhalten, dass  $I = \emptyset$  ist, falls  $b < -(m_3 + m_4)s$  oder  $b > (m_3 + m_4)s$  gilt. Demnach ist in diesem Unterfall

$$M_{i,j}^{*,m_3,m_4}(p) \subseteq [-(m_3 + m_4)s, (m_3 + m_4)s],$$

also

$$\int_{M_{i,j}^{*,m_3,m_4}(p)} 1 \, db \leq 2(m_3 + m_4)s \leq 4m_{\max}s.$$

**Unterfall 3b:**  $m_3 < 0$  und  $m_4 < 0$

In diesem Unterfall gilt

$$\begin{aligned} b > -(m_3 + m_4)s &\Rightarrow x_{l,1} > x_{u,2} \text{ und} \\ b < (m_3 + m_4)s &\Rightarrow x_{l,2} > x_{u,1}. \end{aligned}$$

Zusammenfassend können wir also festhalten, dass  $I = \emptyset$  ist, falls  $b > -(m_3 + m_4)s$  oder  $b < (m_3 + m_4)s$  gilt. Demnach ist in diesem Unterfall

$$M_{i,j}^{*,m_3,m_4}(p) \subseteq [(m_3 + m_4)s, -(m_3 + m_4)s],$$

also

$$\int_{M_{i,j}^{*,m_3,m_4}(p)} 1 \, db \leq 2(-m_3 - m_4)s \leq 4m_{\max}s.$$

**4. Fall:**  $m_3 \neq 0, m_4 \neq 0$  und  $m_3 \cdot m_4 < 0$

Wir beginnen auch diesen Fall damit, die zweite Ungleichung in (B.1) umzuformen. Es gilt

$$-s \leq \frac{b - m_3x}{m_4} \leq s \iff \frac{m_4}{m_3}s + \frac{b}{m_3} \leq x \leq -\frac{m_4}{m_3}s + \frac{b}{m_3}.$$

Nimmt man nun beide Ungleichungen in (B.1) zusammen, so erhält man für  $x$  die unteren Schranken

$$x_{l,1} = -s \text{ und } x_{l,2} = \frac{m_4}{m_3}s + \frac{b}{m_3}$$

und die oberen Schranken

$$x_{u,1} = s \text{ und } x_{u,2} = -\frac{m_4}{m_3}s + \frac{b}{m_3}.$$

Für vorgegebenes  $b$  darf  $x$  also in das Intervall  $I = [\max\{x_{l,1}, x_{l,2}\}, \min\{x_{u,1}, x_{u,2}\}]$  fallen. Ist dieses Intervall leer, so gibt es kein geeignetes  $x$ . Um zu bestimmen, wann dieses Intervall leer wird, lösen wir zunächst die Gleichungen  $x_{l,1} = x_{u,2}$  und  $x_{l,2} = x_{u,1}$  jeweils nach  $b$  auf. Es gilt

$$x_{l,1} = x_{u,2} \iff b = (m_4 - m_3)s$$

und

$$x_{l,2} = x_{u,1} \iff b = -(m_4 - m_3)s.$$

**Unterfall 4a:**  $m_3 > 0$  und  $m_4 < 0$

In diesem Unterfall gilt

$$\begin{aligned} b < (m_4 - m_3)s &\Rightarrow x_{l,1} > x_{u,2} \text{ und} \\ b > -(m_4 - m_3)s &\Rightarrow x_{l,2} > x_{u,1}. \end{aligned}$$

Zusammenfassend können wir also festhalten, dass  $I = \emptyset$  ist, falls  $b < (m_4 - m_3)s$  oder  $b > -(m_4 - m_3)s$  gilt. Demnach ist in diesem Unterfall

$$M_{i,j}^{*,m_3,m_4}(p) \subseteq [(m_4 - m_3)s, -(m_4 - m_3)s],$$

also

$$\int_{M_{i,j}^{*,m_3,m_4}(p)} 1 \, db \leq 2(m_3 - m_4)s \leq 4m_{\max}s.$$

**Unterfall 4b:**  $m_3 < 0$  und  $m_4 > 0$

In diesem Unterfall gilt

$$\begin{aligned} b > (m_4 - m_3)s &\Rightarrow x_{l,1} > x_{u,2} \text{ und} \\ b < -(m_4 - m_3)s &\Rightarrow x_{l,2} > x_{u,1}. \end{aligned}$$

Zusammenfassend können wir also festhalten, dass  $I = \emptyset$  ist, falls  $b > (m_4 - m_3)s$  oder  $b < -(m_4 - m_3)s$  gilt. Demnach ist in diesem Unterfall

$$M_{i,j}^{*,m_3,m_4}(p) \subseteq [-(m_4 - m_3)s, (m_4 - m_3)s],$$

also

$$\int_{M_{i,j}^{*,m_3,m_4}(p)} 1 \, db \leq 2(-m_3 + m_4)s \leq 4m_{\max}s.$$

# Literaturverzeichnis

- [1] R. Beier und B. Vöcking (2003). *Random Knapsack in Expected Polynomial Time*. In *Proc. of the 35th Annual ACM Symposium on Theory of Computing (STOC-2003)*, S. 232–241.
- [2] R. Beier und B. Vöcking (2004). *Typical Properties of Winners and Losers in Discrete Optimization*. In *Proc. of the 36th Annual ACM Symposium on Theory of Computing (STOC-2004)*, S. 343–352.
- [3] P. Billingsley (1995). *Probability and Measure*. John Wiley & Sons, New York, dritte Auflage.
- [4] S. Cook (1971). *The Complexity of Theorem-Proving Procedures*. In *Conference Record of Third Annual ACM Symposium on Theory of Computing*, S. 151–158.
- [5] W. Feller (1971). *An Introduction to Probability Theory and Its Applications*, Band II. John Wiley & Sons, New York, zweite Auflage.
- [6] U. Krengel (2002). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, Wiesbaden, 6. Auflage.
- [7] K. Mulmuley, U. Vazirani und V. Vazirani (1987). *Matching is as Easy as Matrix Inversion*. In *Combinatorica 7(1)*, S. 105–113.
- [8] D. Spielman. *Smoothed Analysis Homepage*. Internet-Seite. <http://www-math.mit.edu/~spielman/SmoothedAnalysis/>.
- [9] D. A. Spielman und S.-H. Teng (2001). *Smoothed Analysis of Algorithms: Why The Simplex Algorithm Usually Takes Polynomial Time*. In *Proc. of the 33rd Annual ACM Symposium on Theory of Computing*, S. 296–305.
- [10] M. Verma, L. Wehmeyer und P. Marwedel (2004). *Cache-Aware Scratchpad Allocation*. In *Proc. of DATE 2004*, S. 1264–1269.
- [11] J. Wang (1997). *Average-Case Computational Complexity Theory*. In *Complexity Theory Retrospective II*, S. 295–328. Springer, Berlin.



# Symbolverzeichnis

$[n]$	$\{1, 2, \dots, n\}$
$\square, \diamond$	Ende eines Beweises bzw. Ende eines Beispiels
$\Delta, \Lambda, \Gamma$	Größe des Gewinner-, Verlierer-, bzw. Gültigkeits-Gaps
$\lfloor x \rfloor_b, \lceil x \rceil_b$	die Zahl $x$ nach dem $b$ -ten Nachkommabit ab- bzw. aufgerundet
$\log$	Logarithmus zur Basis 2
$\mathbb{N}$	Menge der natürlichen Zahlen, $\{1, 2, 3, \dots\}$
$\mathbb{N}_0$	$\{0, 1, 2, 3, \dots\}$
$\mathbb{R}_{\geq 0}$	Menge der nicht negativen reellen Zahlen
$\mathbb{Z}, \mathbb{Q}, \mathbb{R}$	Menge der ganzen, rationalen bzw. reellen Zahlen
$\mathcal{D}$	Wertebereich der ganzzahligen Variablen
$\mathcal{H}$	Perturbationsmodell
$\mathcal{H}(I, \phi)$	Perturbation der Eingabe $I$ gemäß $\mathcal{H}$
$\mathcal{S}$	Menge der erlaubten Lösungen aus $\mathcal{D}^n$
$\phi$	Supremum der Dichte des Perturbationsmodells
$m$	$ \mathcal{D} $
$m_{\max}$	$\max\{ a  \mid a \in \mathcal{D}\}$
$N$	Eingabegröße
$n$	Anzahl der ganzzahligen Variablen

# Index

- adaptives Runden, 13, 44
- average-case Komplexität, 2
- bikriteriell, 28
- binäres Optimierungsproblem, 3
- binäres Programm, 3
- Borel-Menge, 58
- borelsch, 58
- Branch and Bound Algorithmus, 56
- Constrained Shortest Path Problem, 5
- CSP, 5
- Dichte, 58
  - bedingte, 64
  - gemeinsame, 62
- Elementarereignis, 57
- Ereignis, 58
- Erwartungswert, 61
  - absoluter, 61
- Gültigkeits-Gap, 12, 19
  - bei mehreren Nebenbedingungen, 33
- ganzzahliges lineares Optimierungsproblem, 6
- ganzzahliges Optimierungsproblem, 3
- ganzzahliges Programm, 4
- geglättete Analyse, 2
- Gewicht, 12
- Gewinner, 12, 16
  - bei mehreren Nebenbedingungen, 33
- Gewinner-Gap, 11, 16
- Isolating Lemma, 12
- Komplexitätsmaß
  - durchschnittliche Komplexität, 1
  - geglättete Komplexität, 2
  - maximale Rechenzeit, 1
- Las-Vegas-Algorithmus, 11
- linear abhängig, 20
- linear unabhängig, 20
- lineares Optimierungsproblem, 1
- lineares Programm, 2
- Markoff-Ungleichung, 62
- messbarer Raum, 58
- Nemhauser/Ullman Algorithmus, 55
- Nullmenge, 19
- Orakel, 7
- Packungsproblem, 51
- Packungsprogramm, 51
- pareto-optimal, 28
- Perturbation, 2
  - relative, 55
- Perturbationsmodell, 2
- $\phi$ -Perturbation, 7
- polynomielle geglättete Komplexität, 3, 7
- Randdichte, 62
- semi-zufälliges Eingabemodell, 2
- $\sigma$ -additiv, 58
- $\sigma$ -Algebra, 57, 58
  - borelsche, 58
- Simplex-Algorithmus, 2
- smoothed Komplexität, 2

- totale Wahrscheinlichkeit, 64
- Träger, 20
- Traveling Salesperson Problem, 4
- TSP, 4
  
- Überdeckungsproblem, 52
- Überdeckungsprogramm, 51
- UFC, 4
- Unabhängigkeit, 61
- Uncapacitated Facility Location Problem,  
4
- Ungleichung
  - Boole und Bonferroni, 59
  
- Verlierer, 12, 19
  - minimaler, 12, 19
- Verlierer-Gap, 12, 19
  - bei mehreren Nebenbedingungen, 34
- Verteilung
  - gemeinsame, 62
- Verteilungsfunktion, 60, 61
  
- Wahrscheinlichkeitsmaß, 57, 58
- Wahrscheinlichkeitsraum, 57, 58
- Wahrscheinlichkeitsverteilung, 58
- worst-case Komplexität, 1
  
- Zertifizierer, 13
  - stochastische Nebenbedingungen, 42
  - stochastische Zielfunktion, 40
  - stochastische Zielfunktion und Neben-  
bedingungen, 43
- ZPP, 10
- Zufallsexperiment, 57
- Zufallsvariable, 60
  - $n$ -dimensionale, 62
  - reelle, 60
  - stetige, 61