Stephan Wagner

Anwendungen der Hardy-Littlewood-Methode

Partitionen und das Problem von Waring

Diplomarbeit*

Technische Mathematik

Studienzweig Operations Research, Statistik, Finanz- und Versicherungsmathematik

Verfasst am
Institut für Mathematik A
Technische Universität Graz
unter Anleitung von

O.Univ.-Prof. Dr. Robert F. Tichy

Graz, Juli 2004

*Diese Arbeit wurde unterstützt vom FWF-Projekt S-8307-MAT.

Ich versichere, diese Arbeit selbständig verfasst, andere als angegebene Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfsmittel bedient zu haben.

Stephan Wagner

Inhaltsverzeichnis

Vorwort

Für unzählige Probleme der asymptotischen Zahlentheorie und Kombinatorik kann die "circle method" von Hardy und Littlewood Anwendung finden. Viele davon, wie die beiden hier vorgestellten Fragestellungen, gehen bis auf die großen Mathematiker des 17. und 18. Jahrhunderts zurück.

Unter dieser Methode versteht man zunächst nur die prinzipiell einfache, aber unwahrscheinlich elegante und wirkungsvolle Idee, aus der erzeugenden Funktion einer gegebenen Zahlenfolge durch Anwendung der Cauchy'schen Integralformel die Koeffizienten zurückzugewinnen. Durch den Einsatz diverser zusätzlicher Ideen und "Tricks" lässt sich diese Methode auf eine Fülle von Problemstellungen anwenden. Eine besondere Faszination geht vom Ineinandergreifen von analytischen Methoden mit Techniken der elementaren Zahlentheorie und Kombinatorik aus.

Es wird im Zuge dieser Arbeit auf zwei der klassischen Aufgaben eingegangen – die Formel von Rademacher für die Partitionsfunktion und das Problem von Waring. Neben den Originalfragestellungen sollen aber auch neue Varianten und Anwendungen behandelt werden – dies wird insbesondere Bestandteil der Kapitel 4 und 8 sein.

Rademachers Formel für die Partitionsfunktion $p(n)$, die die Anzahl verschiedener Zerlegungen einer natürlichen Zahl n ohne Rücksicht auf die Reihenfolge der einzelnen Summanden zählt, nimmt die ersten beiden Kapitel der Arbeit ein. In Kapitel 1 werden einige elementare Tatsachen über Partitionen bereitgestellt, Kapitel 2 behandelt den Beweis der berühmten Formel.

Mit der Betrachtung von Verteilungseigenschaften von Partitionen, die auf Szekeres zurückgeht, beschäftigt sich Kapitel 3. Kapitel 4 wendet sich schließlich Anwendungen von Partitionen in der Graphentheorie zu, die überraschenderweise sogar mit Fragestellungen aus der Chemie in Verbindung gebracht werden können.

Als zweites klassisches Problem beschreibe ich die Lösung des Waring'schen Problems, in dem es um die Darstellbarkeit von natürlichen Zahlen durch Summen k-ter Potenzen geht, mit Hilfe der Hardy-Littlewood-Methode. Während im einführenden Kapitel 5 einige einfache Spezialfälle beschrieben werden, zeigen die Kapitel 6 und 7 die Lösung von Hardy und Littlewood. Kapitel 8 ist einer Variante gewidmet, bei der zusätzliche Bedingungen über Ziffernentwicklungen eingebracht werden.

Mein besonderer Dank gilt an dieser Stelle all jenen, die in verschiedenster Weise zur Entstehung dieser Arbeit beigetragen haben, insbesondere natürlich O.Univ.-Prof. Dr. Robert F. Tichy für die Betreuung der Arbeit und die Unterstützung durch das FWF-Forschungsprojekt S-8307-MAT.

Abschließend möchte ich mich auch bei allen bedanken, die meine Liebe und Begeisterung für die Mathematik in all den Jahren gefördert haben.

Graz, im Juli 2004

Stephan Wagner

Teil I

Partitionen

Kapitel 1

Allgemeines über Partitionen

1.1 Definitionen und Notation

Definition 1.1 Unter einer *Partition* einer natürlichen Zahl n verstehen wir eine endliche Folge $c = (c_1, \ldots, c_d)$ natürlicher Zahlen $c_i \geq 1$, sodass

- $c_1 + \ldots + c_d = n$,
- $c_1 \geq \ldots \geq c_d$.

Als übliche Schreibweise wird $c \vdash n$ für "c ist Partition von n" verwendet.

Eine gute Möglichkeit zur Darstellung von Partitionen sind die sogenannten *Ferrer-Diagramme* oder *Young-Tableaus*. Dabei wird eine Partition durch ein Gebilde aus Kästchen dargestellt, bei dem die Anzahl der Kästchen in der i-ten Zeile das i-te Glied der Folge c darstellt. Die Abbildung zeigt das Ferrer-Diagramm zur Partition $(6, 5, 3, 3, 1)$ von 18.

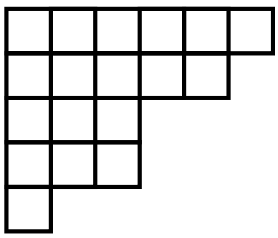

Abbildung 1.1: Beispiel eines Ferrer-Diagramms

Im Folgenden interessieren wir uns vor allem für die Anzahl verschiedener Partitionen in gewissen Klassen. Wir führen die folgenden Notationen ein:

Definition 1.2 Es bezeichne $p(n)$ die Zahl der Partitionen von n; weiters sei $P(n, k)$ die Zahl der Partitionen mit $\leq k$ Teilen und $p(n, k)$ die Zahl der Partitionen mit genau k Teilen, und es sei $p(n, k, l)$ die Zahl der Partitionen mit genau k Teilen und Maximum l.

Ferrer-Diagramme stellen ein gutes Mittel zur Veranschaulichung von Bijektionen zwischen Klassen von Partitionen dar. So gilt etwa folgende Tatsache:

Proposition 1.1 $P(n,k)$, die Anzahl der Partitionen von n in $\leq k$ Teile, ist gleich der Anzahl $p(n+k,k)$ der Partitionen von $n+k$ in k Teile.

Beweis: Man betrachte das Ferrer-Diagramm einer Partition $c \vdash n$ mit $\leq k$ Teilen. Fügt man vorne noch einen $k \times 1$-Block hinzu, so entsteht eine Partition $c' \vdash n+k$ mit genau k Teilen. Gibt man umgekehrt eine Partition von $n+k$ in genau k Teile vor, so kann

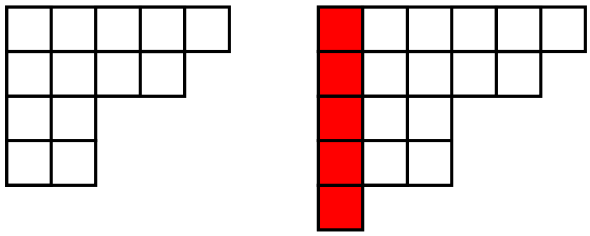

Abbildung 1.2: Zum Beweis von Proposition 1.1

man den $k \times 1$-Block entfernen und erhält eine Partition von n mit $\leq k$ Teilen. Dadurch wird eine Bijektion zwischen diesen beiden Klassen von Partitionen hergestellt, womit die Mächtigkeiten gleich sein müssen. $\square$

Eine sehr wesentliche Operation im Zusammenhang mit Ferrer-Diagrammen ist die Konjugation, die hier kurz formal definiert wird:

Definition 1.3 Sei $c = (c_1, \ldots, c_d) \vdash n$ eine Partition. Dann ist durch $\overline{c}$, definiert durch

$$\overline{c}_i := |\{1 \leq j \leq d : \ c_j \geq i\}|,$$

eine weitere Partition von n gegeben. Diese Operation entspricht dem Spiegeln des Ferrer-Diagramms entlang der Diagonalen. Es gilt $\overline{\overline{c}} = c$.

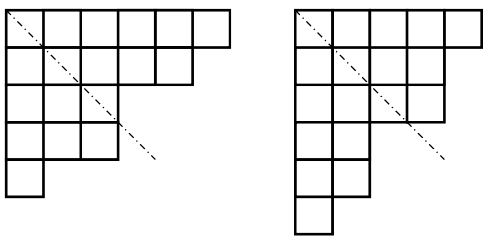

Abbildung 1.3: Konjugation einer Partition

Die Konjugation vertauscht offenbar die Rollen von Länge und Maximum. Daher ist $P(n,k)$ auch die Zahl der Partitionen von n mit Maximum $\leq k$, $p(n,k)$ ist auch die Zahl der Partitionen mit Maximum k, und $p(n,k,l) = p(n,l,k)$.

1.2 Erzeugende Funktionen

Lemma 1.2 Die erzeugende Funktion für die Partitionsfunktion $p(n)$ ist gegeben durch

$$\sum_{n=0}^{\infty} p(n)x^n = \prod_{m=1}^{\infty}(1-x^m)^{-1}. \tag{1.1}$$

Beweis: Es gilt

$$(1-x^m)^{-1} = \sum_{k=0}^{\infty} x^{km}.$$

Wird das Produkt ausfaktorisiert, so ergeben sich genau Summanden der Form

$$x^{\sum_{m=1}^{\infty} k_m m}.$$

Interpretiert man k_m als die Häufigkeit des Auftretens von m in einer Partition c, so gehört zu jeder Partition von n genau ein Summand x^n und umgekehrt. Damit ergibt sich

$$\prod_{m=1}^{\infty}(1-x^m)^{-1} = \sum_{n=0}^{\infty} p(n)x^n.$$

□

Lemma 1.3 Die erzeugende Funktion für $P(n,k)$ ist gegeben durch

$$\sum_{n=0}^{\infty} P(n,k)x^n = \prod_{m=1}^{k}(1-x^m)^{-1}. \tag{1.2}$$

Beweis: $P(n,k)$ beschreibt, wie oben erwähnt, auch die Anzahl aller Partitionen von n mit Maximum $\leq k$, d.h. ausschließlich Teilen $\leq k$. Das bedeutet, dass in der erzeugenden Funktion von $p(n)$ genau die Faktoren wegfallen, die Teilen $> k$ entsprechen. □

Korollar 1.4 Die erzeugende Funktion für $p(n,k)$ ist gegeben durch

$$\sum_{n=0}^{\infty} p(n,k)x^n = x^k \prod_{m=1}^{k}(1-x^m)^{-1}. \tag{1.3}$$

Beweis: Es wurde in Proposition 1.1 bereits gezeigt, dass $P(n,k) = p(n+k,k)$ oder $P(n-k,k) = p(n,k)$ gilt. Daraus folgt unmittelbar

$$\sum_{n=0}^{\infty} p(n,k)x^n = \sum_{n=k}^{\infty} P(n-k,k)x^k x^{n-k} = x^k \sum_{n=0}^{\infty} P(n,k)x^n = x^k \prod_{m=1}^{k}(1-x^m)^{-1}.$$

□

Lemma 1.5 Die erzeugende Funktion für $p(n,k,l)$ ist gegeben durch

$$\sum_{n=0}^{\infty} p(n,k,l)x^n = x^{k+l-1} \prod_{m=1}^{k-1} \frac{1-x^{l+m-1}}{1-x^m}. \tag{1.4}$$

Beweis: Zunächst bestimmen wir die erzeugende Funktion von $P(n,k,l)$, der Anzahl der Partitionen von n in $\leq k$ Teile, die allesamt $\leq l$ sind. Es soll gezeigt werden, dass

$$\sum_{n=0}^{\infty} P(n,k,l)x^n = \frac{(x)_{k+l}}{(x)_k(x)_l} = \prod_{m=1}^{k} \frac{1-x^{l+m}}{1-x^m} = \prod_{m=1}^{l} \frac{1-x^{k+m}}{1-x^m},$$

wobei $(x)_k = \prod_{m=1}^{k}(1-x^m)$ eine sogenannte *q-Reihe* ist. $\frac{(x)_{k+l}}{(x)_k(x)_l}$ bezeichnet man als *Gaußschen Binomialkoeffizienten.* Er erfüllt einige algebraische Relationen und stimmt im Grenzwert $x \to 1$ mit dem Binomialkoeffizienten $\binom{k+l}{k}$ überein.

Sei $G(k,l,x) = \sum_{n=0}^{\infty} P(n,k,l)x^n$ die erzeugende Funktion von $P(n,k,l)$ und $g(k,l,x) = \frac{(x)_{k+l}}{(x)_k(x)_l}$ der Ausdruck auf der rechten Seite. Wir zeigen, dass $G(k,l,x)$ und $g(k,l,x)$ dieselbe Rekursion mit denselben Startwerten erfüllen. Zunächst gilt offenbar

$$P(n,k,0) = P(n,0,l) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

und damit $G(k,0,x) = G(0,l,x) = 1$, denn nur die leere Partition, die eine Partition von 0 ist, hat Länge bzw. Maximum $= 0$. Da $(x)_0 = 1$, gilt auch $g(k,0,x) = g(0,l,x) = 1$.

Weiters haben wir die folgende Rekursionsbeziehung:

$$\begin{aligned} g(k,l,x) - g(k,l-1,x) &= \frac{(x)_{k+l}}{(x)_k(x)_l} - \frac{(x)_{k+l-1}}{(x)_k(x)_{l-1}} \\ &= \frac{(x)_{k+l-1}}{(x)_k(x)_{l-1}}\Big(\frac{1-x^{k+l}}{1-x^l} - 1\Big) \\ &= \frac{(x)_{k+l-1}}{(x)_k(x)_{l-1}} x^l \frac{1-x^k}{1-x^l} \\ &= x^l \frac{(x)_{k+l-1}}{(x)_{k-1}(x)_l} \\ &= x^l g(k-1,l,x). \end{aligned}$$

$P(n,k,l) - P(n,k,l-1)$ beschreibt die Anzahl aller Partitionen von n mit $\leq k$ Teilen und Maximum l. Entfernt man aus dem Ferrer-Diagramm einer solchen Partition die erste Zeile, so ergibt sich eine Partition von $n-l$ mit $\leq k-1$ Teilen und Maximum $\leq l$ und umgekehrt (wie im Beweis von Proposition 1.1). Also folgt $P(n,k,l) - P(n,k,l-1) = P(n-l,k-1,l)$

und damit

$$\begin{aligned}
G(k,l,x) - G(k,l-1,x) &= \sum_{n=0}^{\infty} P(n,k,l)x^n - \sum_{n=0}^{\infty} P(n,k,l-1)x^n \\
&= \sum_{n=l}^{\infty} P(n-l,k-1,l)x^n \\
&= x^l \sum_{n=0}^{\infty} P(n,k-1,l)x^n \\
&= x^l G(k-1,l,x).
\end{aligned}$$

$g(k,l,x)$ und $G(k,l,x)$ sind durch die Startwerte und die Rekursion eindeutig festgelegt. Da diese übereinstimmen, gilt auch $g(k,l,x) = G(k,l,x)$, wie behauptet.

Entfernt man die erste Zeile und Spalte im Ferrer-Diagramm einer Partition von n mit k Teilen und Maximum l, so ergibt sich eine Partition von $n-k-l+1$ mit $\leq k-1$ Teilen und Maximum $\leq l-1$ und umgekehrt (wieder wie im Beweis von Proposition 1.1), es gilt somit $p(n,k,l) = P(n-k-l+1, k-1, l-1)$. Daher folgt schließlich

$$\begin{aligned}
\sum_{n=0}^{\infty} p(n,k,l)x^n &= \sum_{n=k+l-1}^{\infty} P(n-k-l+1,k-1,l-1)x^n \\
&= x^{k+l-1} \sum_{n=0}^{\infty} P(n,k-1,l-1)x^n \\
&= x^{k+l-1} \prod_{m=1}^{k-1} \frac{1-x^{l+m-1}}{1-x^m}.
\end{aligned}$$

□

Zuletzt sei noch auf eine berühmte Identität hingewiesen, die auf Euler zurückgeht, und die eine rekursive Berechnung von $p(n)$ ermöglicht.

Satz 1.6 *Es sei $F(x) = \prod_{m=1}^{\infty}(1-x^m)^{-1}$ die erzeugende Funktion von $p(n)$. Dann gilt*

$$F(x)^{-1} = \prod_{m=1}^{\infty}(1-x^m) = 1 + \sum_{m=1}^{\infty}(-1)^m \left(x^{\frac{3m^2-m}{2}} + x^{\frac{3m^2+m}{2}}\right). \tag{1.5}$$

Beweis: Der Beweis wird kombinatorisch geführt, wobei Ferrer-Diagramme wiederum hilfreich sein werden. Auf der linken Seite ergeben sich beim Ausmultiplizieren des Produkts Summanden von der Form

$$(-1)^k x^{m_1+m_2+\ldots+m_k},$$

wobei die m_i paarweise verschieden sind. Daher ist der Koeffizient von x^n genau $p_g(n) - p_u(n)$, wobei $p_g(n)$ die Anzahl der Partitionen von n in eine gerade Zahl paarweise verschiedener Summanden bezeichnet, $p_u(n)$ die Anzahl der Partitionen von n in eine ungerade Zahl paarweise verschiedener Summanden.

Es sei $c = (c_1, \ldots, c_d)$ eine Partition von n in paarweise verschiedene Teile. Wir betrachten das Ferrer-Diagramm von c. Wir bezeichnen die unterste Zeile (die dem minimalen Teil c_d entspricht) mit Basis. Weiters sei r maximal mit der Eigenschaft, dass $c_j = c_{j+1} + 1$ für alle $1 \leq j < r$ gilt. Die ganz rechts liegenden Kästchen der Zeilen $1, 2, \ldots, r$ bilden die sogenannte Böschung des Diagramms.

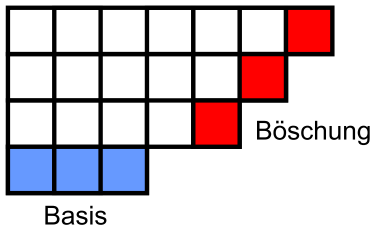

Abbildung 1.4: Basis und Böschung eines Ferrer-Diagramms

Sei b die Größe der Basis und g die Größe der Böschung. Nun ordnen wir c eine andere Partition c^* in folgender Weise zu:

- Falls $b \leq g$, so entfernen wir die Basis und hängen sie an die Böschung an. Die Partition, die dem entstehenden Diagramm entspricht, hat dann lauter verschiedene Teile, und die Anzahl der Teile sinkt um 1. Die Größe der neuen Böschung ist gleich der Größe der alten Basis und daher kleiner als die der neuen Basis.

- Falls $b > g$, so entfernen wir die Böschung und hängen sie unten an die Basis an. Wieder hat die neue Partition lauter verschiedene Teile, und die Anzahl der Teile wächst um 1. Die Größe der neuen Böschung ist zumindest so groß wie die der alten und damit $\geq$ der Größe der neuen Basis.

Aus dem Gesagten ergibt sich auch, dass $(c^*)^* = c$, womit sich eine Bijektion zwischen den Partitionen von n in eine gerade Zahl paarweise verschiedener Summanden und den Partitionen von n in eine ungerade Zahl paarweise verschiedener Summanden ergibt. Ausnahmen kann es nur dann geben, wenn Basis und Böschung ein gemeinsames Kästchen haben und $b = g$ oder $b = g+1$ ist, wie in den folgenden Bildern dargestellt ist (Abbildung 1.5). Im ersten Fall ergibt sich keine Partition mit verschiedenen Teilen. Hier ist

$$n = \sum_{i=1}^{g} (g+i) = g^2 + \frac{g(g+1)}{2} = \frac{3g^2+g}{2}.$$

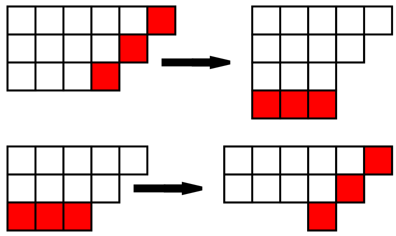

Abbildung 1.5: Ausnahmen zur konstruierten Bijektion

Im zweiten Fall erhält man kein zulässiges Diagramm, und es ist

$$n = \sum_{i=1}^{g} (g + i - 1) = g^2 + \frac{g(g-1)}{2} = \frac{3g^2 - g}{2}.$$

Ist g gerade, so folgt $p_g(n) - p_u(n) = 1$, ist g ungerade, $p_g(n) - p_u(n) = -1$. Daher gilt

$$p_g(n) - p_u(n) = \begin{cases} (-1)^g & n = \frac{3g^2 \pm g}{2} \\ 0 & \text{sonst}, \end{cases}$$

womit die gewünschte Formel folgt. □

Korollar 1.7 Für $n \geq 1$ gilt

$$p(n) = \sum_{m=1}^{\infty} (-1)^{m+1} \left(p\left(n - \frac{3m^2 - m}{2}\right) + p\left(n - \frac{3m^2 + m}{2}\right)\right), \tag{1.6}$$

wobei $p(k) = 0$ für $k < 0$ gesetzt wird.

Bemerkung: Die Summe ist endlich, weil für hinreichend großes m $n - \frac{3m^2-m}{2}$ und $n - \frac{3m^2+m}{2}$ negativ sind.

Beweis: Aus der Identität

$$\left(\sum_{n=0}^{\infty} p(n) x^n\right) \left(1 + \sum_{m=1}^{\infty} (-1)^m \left(x^{\frac{3m^2-m}{2}} + x^{\frac{3m^2+m}{2}}\right)\right) = 1$$

erhält man durch Ausmultiplizieren der linken Seite als Koeffizienten von x^n

$$p(n) - \sum_{m=1}^{\infty} (-1)^{m+1} \left(p\left(n - \frac{3m^2 - m}{2}\right) + p\left(n - \frac{3m^2 + m}{2}\right)\right).$$

Dies muss gleich 0 sein, womit das behauptete Ergebnis folgt. □

Kapitel 2

Rademachers Formel für die Partitionsfunktion

Es stellt sich uns nun die Frage, wie sich die Größe $p(n)$ asymptotisch verhält. Wie in diesem Kapitel ausgeführt wird, gilt

$$p(n) \sim \frac{\exp(\pi\sqrt{2n/3})}{4\sqrt{3}n}, \tag{2.1}$$

was von Hardy und Ramanujan 1918 [8] mit Hilfe der "circle method", um die es in dieser Arbeit geht, bewiesen wurde. In der Tat konnten sie sogar zeigen, dass eine asymptotische Formel von der Form

$$p(n) = \sum_{k<\alpha\sqrt{n}} P_k(n) + O(n^{-1/4}) \tag{2.2}$$

gilt, wobei $P_1(n) \sim \frac{\exp(\pi\sqrt{2n/3})}{4\sqrt{3}n}$ den Hauptterm darstellt und alle weiteren von ähnlicher Bauart sind. Da der Fehler von der Ordnung $O(n^{-1/4})$ ist, ist die Formel für hinreichend großes n sogar exakt, da ja $p(n)$ eine ganze Zahl ist. Der Nachteil an dieser Formel ist, dass die unendliche Summe der $P_k(n)$ nicht konvergiert.

Dieser Nachteil konnte von Rademacher 1937 [23] ausgemerzt werden, indem er einen anderen Integrationspfad im Beweis wählte. Er erhielt dadurch eine konvergente Reihe

$$p(n) = \sum_{k=1}^{\infty} R_k(n). \tag{2.3}$$

Zum Beweis werden gewisse Eigenschaften der erzeugenden Funktion benötigt, um das Kurvenintegral

$$\frac{1}{2\pi i}\int_{\mathcal{C}} \frac{F(x)}{x^{n+1}}\,dx \tag{2.4}$$

auszuwerten, wobei $F(x) = \prod_{m=1}^{\infty}(1-x^m)^{-1} = \sum_{n=0}^{\infty} p(n)x^n$ und $\mathcal{C}$ eine Kurve um den Nullpunkt ist. Nach der Cauchy'schen Integralformel entspricht dies genau $p(n)$. Der hier gebrachte Beweis folgt im Wesentlichen [2].

2.1 Die Dedekind'sche η-Funktion

$F(x)$ hängt mit der sogenannten *Dedekind'schen η-Funktion* zusammen, die folgendermaßen gegeben ist:

Definition 2.1

$$\eta(\tau) := e^{\pi i\tau/12} \prod_{n=1}^{\infty} (1 - e^{2\pi i n\tau}). \tag{2.5}$$

Offensichtlich gilt $\eta(\tau + m) = e^{\pi i m/12}\eta(\tau)$ für ganzzahliges m. Die Funktion erfüllt sogar noch mehr, nämlich die folgende bemerkenswerte Funktionalgleichung:

Satz 2.1 *Es seien $a, b, c, d \in \mathbb{Z}$ mit $c > 0$, $ad - bc = 1$, und $\tau \in H = \{\tau \in \mathbb{C} : \ \operatorname{Im} \tau > 0\}$. Dann gilt*

$$\eta\Big(\frac{a\tau + b}{c\tau + d}\Big) = \varepsilon(a, b, c, d)(-i(c\tau + d))^{1/2}\eta(\tau), \tag{2.6}$$

wobei

$$\varepsilon(a, b, c, d) = \exp\Big(\pi i\Big(\frac{a + d}{12c} + s(-d, c)\Big)\Big)$$

und

$$s(h, k) = \sum_{r=1}^{k-1} \frac{r}{k}\Big(\frac{hr}{k} - \Big\lfloor\frac{hr}{k}\Big\rfloor - \frac{1}{2}\Big)$$

eine sogenannte Dedekind-Summe ist.

Beweis: Diese Funktionalgleichung kann auf verschiedene Arten bewiesen werden, etwa als Folgerung zu einer allgemeineren Formel von Iseki (s. [2]). Hier soll ein Beweis gezeigt werden, der auf Basil Gordon zurückgeht: zunächst betrachten wir die modulare Gruppe, definiert durch

$$\Gamma = \Big\{A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Big| \, a, b, c, d \in \mathbb{Z}, \ \det A = 1\Big\},$$

wobei die Matrizen A und $-A$ identifiziert werden (entspricht der Bildung der Faktorgruppe bezüglich dem Normalteiler, der aus der Einheitsmatrix I und $-I$ besteht). Diese Gruppe wirkt auf der oberen Halbebene H durch Möbius-Transformation:

$$z \mapsto \frac{az + b}{cz + d}$$

Man rechnet leicht nach, dass die Bildung der inversen Abbildung dem Invertieren der Matrix und die Verknüpfung der Multiplikation entspricht. Wir schreiben daher $Az = \frac{az+b}{cz+d}$, falls $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Lemma 2.2 Γ wird von den beiden Matrizen

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{und} \qquad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

erzeugt. Jede Matrix $A \in \Gamma$ kann in der Form

$$A = T^{n_1} S T^{n_2} S \dots S T^{n_k}$$

mit ganzen Zahlen n_i geschrieben werden.

Beweis: Da A und $-A$ identifiziert werden, können wir uns auf Matrizen mit $c \geq 0$ beschränken. Wir benutzen nun Induktion nach c.

Falls $c = 0$, dann gilt $ad = 1$ und damit $a = d = \pm 1$, daher

$$A = \begin{pmatrix} \pm 1 & b \\ 0 & \pm 1 \end{pmatrix} = \begin{pmatrix} 1 & \pm b \\ 0 & 1 \end{pmatrix} = T^{\pm b}.$$

Nun folgt der Induktionsschritt. Es sei $c > 0$. Division mit Rest ergibt $d = cq + r$ für gewisse $q, r \in \mathbb{Z}$, wobei $0 \leq r < c$. Dann gilt

$$AT^{-q} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & -q \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & -aq + b \\ c & r \end{pmatrix}$$

und weiter

$$AT^{-q}S = \begin{pmatrix} a & -aq + b \\ c & r \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -aq + r & -a \\ r & -c \end{pmatrix}.$$

Nach der Induktionsvoraussetzung kann diese Matrix durch T und S ausgedrückt werden, und damit auch A. $\square$

Wir wollen also zeigen, dass die Funktionalgleichung (2.6) für beliebiges $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ mit $c > 0$ gilt. Zunächst zeigen wir sie für den Spezialfall $A = S$:

Lemma 2.3 Für $\tau \in H = \{\tau \in \mathbb{C} : \ \operatorname{Im} \tau > 0\}$ gilt

$$\eta\Big(-\frac{1}{\tau}\Big) = (-i\tau)^{1/2} \eta(\tau). \tag{2.7}$$

Beweis: Der hier gebrachte Beweis geht auf Siegel [26] zurück. Wir beweisen die Gleichung zunächst für rein imaginäres $\tau = iy$. Da beide Seiten analytisch auf ganz H sind, folgt die Gleichheit dann aus der Eindeutigkeit der analytischen Fortsetzung. Die Gleichung ist in diesem Fall äquivalent zu

$$\log \eta(i/y) - \log \eta(iy) = \frac{1}{2} \log y.$$

Wir entwickeln die linke Seite in zwei Reihen:

$$\begin{aligned}\log \eta(iy) &= -\frac{\pi y}{12} + \log \prod_{n=1}^{\infty}(1 - e^{-2\pi n y}) = -\frac{\pi y}{12} + \sum_{n=1}^{\infty} \log(1 - e^{-2\pi n y}) \\ &= -\frac{\pi y}{12} - \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{e^{-2\pi n m y}}{m} = -\frac{\pi y}{12} - \sum_{m=1}^{\infty} \frac{e^{-2\pi m y}}{m(1 - e^{-2\pi m y})} \\ &= -\frac{\pi y}{12} + \sum_{m=1}^{\infty} \frac{1}{m(1 - e^{2\pi m y})}\,.\end{aligned}$$

Wir müssen also die Gleichung

$$\sum_{m=1}^{\infty} \frac{1}{m(1 - e^{2\pi m y})} - \sum_{m=1}^{\infty} \frac{1}{m(1 - e^{2\pi m / y})} - \frac{\pi}{12}\Big(y - \frac{1}{y}\Big) = -\frac{1}{2}\log y \tag{2.8}$$

beweisen. Zu diesem Zweck betrachten wir für festes $y > 0$ die Funktion

$$F_n(z) = -\frac{1}{8z}\cot(\pi i N z)\cot\frac{\pi N z}{y},$$

wobei $N = n + \frac{1}{2}$. Diese Funktion integrieren wir entlang des Parallelogramms $\mathcal{C}$, das die Punkte $y, i, -y, -i$ verbindet. Das Integral ist gleich $2\pi i$ mal der Summe aus den Residuen im Inneren von $\mathcal{C}$. F_n hat einfache Pole in $z = \frac{ik}{N}$ und $z = \frac{ky}{N}$ für $k = \pm 1, \pm 2, \dots, \pm n$. Zudem gibt es einen Pol dritter Ordnung mit Residuum $i(y - y^{-1})/24$ (ergibt sich einfach aus der Reihenentwicklung des Cotangens) in 0. In $z = \frac{ik}{N}$ ist das Residuum

$$\frac{1}{8\pi k}\cot\frac{\pi i k}{y}.$$

Diese Funktion ist gerade in k. Daher gilt

$$\sum_{\substack{k=-n \\ k\neq 0}}^{n} \operatorname*{Res}_{z=ik/N} F_n(z) = 2\sum_{k=1}^{n} \frac{1}{8\pi k}\cot\frac{\pi i k}{y}.$$

Für den Cotangens schreiben wir

$$\cot i\theta = \frac{\cos i\theta}{\sin i\theta} = i\frac{e^{-\theta} + e^{\theta}}{e^{-\theta} - e^{\theta}} = -i\frac{e^{2\theta} + 1}{e^{2\theta} - 1} = \frac{1}{i}\Big(1 - \frac{2}{1 - e^{2\theta}}\Big).$$

Dann ergibt sich

$$\sum_{\substack{k=-n \\ k\neq 0}}^{n} \operatorname*{Res}_{z=ik/N} F_n(z) = \frac{1}{4\pi i}\sum_{k=1}^{n}\frac{1}{k} - \frac{1}{2\pi i}\sum_{k=1}^{n} \frac{1}{k(1 - e^{2\pi k / y})}\,.$$

Analog zeigt man

$$\sum_{\substack{k=-n\\k\neq 0}}^{n} \operatorname*{Res}_{z=ky/N} F_n(z) = \frac{i}{4\pi}\sum_{k=1}^{n}\frac{1}{k} - \frac{i}{2\pi}\sum_{k=1}^{n}\frac{1}{k(1-e^{2\pi ky})}.$$

Daraus folgt, dass das Integral von $F_n(z)$ über $\mathcal{C}$ für $n \to \infty$ der linken Seite der Gleichung (2.8) entspricht. Zu zeigen bleibt noch

$$\lim_{n\to\infty}\int_{\mathcal{C}} F_n(z)\,dz = -\frac{1}{2}\log y.$$

Der Grenzwert von $zF_n(z)$ für $n \to \infty$ ist auf den Kanten zwischen y, i bzw. $-y, -i$ (die Ecken ausgenommen) gleich $\frac{1}{8}$, auf den beiden anderen $-\frac{1}{8}$. Dies ergibt sich einfach aus den bekannten Tatsachen, dass

$$\cos(nz) \sim \begin{cases} \frac{e^{n\operatorname{Im} z - in\operatorname{Re} z}}{2} & \operatorname{Im} z > 0\\ \frac{e^{-n\operatorname{Im} z + in\operatorname{Re} z}}{2} & \operatorname{Im} z < 0\end{cases}$$

bzw.

$$\sin(nz) \sim \begin{cases} \frac{ie^{n\operatorname{Im} z - in\operatorname{Re} z}}{2} & \operatorname{Im} z > 0\\ -\frac{ie^{-n\operatorname{Im} z + in\operatorname{Re} z}}{2} & \operatorname{Im} z < 0\end{cases}$$

und daher

$$\cot(nz) \overset{n\to\infty}{\to} \begin{cases} -i & \operatorname{Im} z > 0\\ i & \operatorname{Im} z < 0.\end{cases}$$

Damit ist $F_n(z)$ für alle n gleichmäßig beschränkt auf $\mathcal{C}$. Man kann daher Integral und Grenzwert vertauschen und erhält

$$\begin{aligned}\lim_{n\to\infty}\int_{\mathcal{C}} F_n(z)\,dz &= \int_{\mathcal{C}} \lim_{n\to\infty} zF_n(z)\frac{dz}{z}\\ &= \frac{1}{8}\Big(-\int_{-i}^{y} + \int_{y}^{i} - \int_{i}^{-y} + \int_{-y}^{-i}\Big)\frac{dz}{z}\\ &= \frac{1}{4}\Big(-\int_{-i}^{y} + \int_{y}^{i}\Big)\frac{dz}{z}\\ &= \frac{1}{4}\Big(-\log y + \log(-i) + \log i - \log y\Big)\\ &= -\frac{1}{2}\log y,\end{aligned}$$

womit alles bewiesen ist. □

2.2 Eigenschaften von Dedekind-Summen

Schließlich benötigen wir einige Eigenschaften der Dedekind-Summen

$$s(h,k) = \sum_{r=1}^{k-1} \frac{r}{k}\Big(\frac{hr}{k} - \Big\lfloor \frac{hr}{k} \Big\rfloor - \frac{1}{2}\Big), \tag{2.9}$$

bevor wir endgültig zum Beweis von Satz 2.1 übergehen können. Wir führen dazu die Schreibweise

$$((x)) = \begin{cases} x - \lfloor x \rfloor - \frac{1}{2} & x \notin \mathbb{Z} \\ 0 & x \in \mathbb{Z} \end{cases}$$

ein. Diese Funktion ist offenbar periodisch und ungerade. Es gilt daher

$$\sum_{r \mod k} \Big(\Big(\frac{hr}{k}\Big)\Big) = 0, \text{ falls } (h,k) = 1.$$

Damit folgt

$$\sum_{r \mod k} \Big(\Big(\frac{r}{k}\Big)\Big)\Big(\Big(\frac{hr}{k}\Big)\Big) = \sum_{r \mod k} \Big(\frac{r}{k} - \frac{1}{2}\Big)\Big(\Big(\frac{hr}{k}\Big)\Big) = \sum_{r \mod k} \frac{r}{k}\Big(\Big(\frac{hr}{k}\Big)\Big).$$

Man kann daher die Dedekindsumme $s(h,k)$ in der folgenden Weise schreiben:

$$s(h,k) = \sum_{r \mod k} \Big(\Big(\frac{r}{k}\Big)\Big)\Big(\Big(\frac{hr}{k}\Big)\Big). \tag{2.10}$$

Diese Schreibweise zeigt, dass $s(h,k)$ nur von der Kongruenzklasse von h modulo k abhängt.

Lemma 2.4

1. Falls $h' \equiv \pm h \mod k$, dann ist $s(h',k) = \pm s(h,k)$.
2. Falls $h\bar{h} = \pm 1 \mod k$, dann ist $s(\bar{h},k) = \pm s(h,k)$.
3. Falls $h^2 + 1 \equiv 0 \mod k$, dann ist $s(h,k) = 0$.

Beweis: Aus der Darstellung (2.10) sieht man sofort 1. und 2.: 1., weil $((x))$ ungerade ist, 2., indem man r durch $\bar{h}r$ ersetzt (dabei wird wegen $(\bar{h},k) = 1$ ebenso ein vollständiges Restsystem durchlaufen). Schließlich ergibt sich aus $h^2 + 1 \equiv 0 \mod k$ nach 2. $s(h,k) = -s(h,k) = 0$. □

Proposition 2.5 (Reziprozitätsgesetz für Dedekind-Summen) Falls $h, k > 0$ und $(h,k) = 1$, dann gilt

$$12hks(h,k) + 12khs(k,h) = h^2 + k^2 - 3hk + 1.$$

Beweis: Wir berechnen die Summe $\sum_{r=1}^{k}((hr/k))^2$ auf zwei verschiedene Arten: zunächst ist

$$\sum_{r=1}^{k}\left(\left(\frac{hr}{k}\right)\right)^2 = \sum_{r \mod k}\left(\left(\frac{hr}{k}\right)\right)^2 = \sum_{r \mod k}\left(\left(\frac{r}{k}\right)\right)^2 = \sum_{r=1}^{k-1}\left(\frac{r}{k}-\frac{1}{2}\right)^2.$$

Andererseits gilt

$$\begin{aligned}
\sum_{r=1}^{k}\left(\left(\frac{hr}{k}\right)\right)^2 &= \sum_{r=1}^{k-1}\left(\frac{hr}{k}-\left\lfloor\frac{hr}{k}\right\rfloor-\frac{1}{2}\right)^2\\
&= \sum_{r=1}^{k-1}\left(\frac{h^2r^2}{k^2}+\left\lfloor\frac{hr}{k}\right\rfloor^2+\frac{1}{4}-\frac{hr}{k}+\left\lfloor\frac{hr}{k}\right\rfloor-\frac{2hr}{k}\left\lfloor\frac{hr}{k}\right\rfloor\right)\\
&= 2h\sum_{r=1}^{k-1}\frac{r}{k}\left(\frac{hr}{k}-\left\lfloor\frac{hr}{k}\right\rfloor-\frac{1}{2}\right)+\sum_{r=1}^{k-1}\left\lfloor\frac{hr}{k}\right\rfloor\left(\left\lfloor\frac{hr}{k}\right\rfloor+1\right)-\frac{h^2}{k^2}\sum_{r=1}^{k-1}r^2+\frac{1}{4}\sum_{r=1}^{k-1}1.
\end{aligned}$$

Aus diesen beiden und der Definition der Dedekind-Summen ergibt sich

$$2hs(h,k)+\sum_{r=1}^{k-1}\left\lfloor\frac{hr}{k}\right\rfloor\left(\left\lfloor\frac{hr}{k}\right\rfloor+1\right) = \frac{h^2+1}{k^2}\sum_{r=1}^{k-1}r^2-\frac{1}{k}\sum_{r=1}^{k-1}r. \tag{2.11}$$

In der Summe auf der linken Seite betrachten wir all jene Terme, für die $\lfloor\frac{hr}{k}\rfloor$ einen fixen Wert hat. Wegen $0<r<k$ gilt $0<\frac{hr}{k}<h$ und daher $0\le\lfloor\frac{hr}{k}\rfloor\le h-1$. Es sei nun $N(\nu)$ die Anzahl jener Indices r, für die $\lfloor\frac{hr}{k}\rfloor=\nu-1$ gilt ($1\le\nu\le h$). Wegen $(h,k)=1$ und $0<r<k$ ist kein $\frac{hr}{k}$ ganzzahlig. Daher gilt

$$\left\lfloor\frac{hr}{k}\right\rfloor=\nu-1 \iff \nu-1<\frac{hr}{k}<\nu \iff \frac{k(\nu-1)}{h}<r<\frac{k\nu}{h}$$

und somit

$$N(\nu)=\begin{cases}\lfloor\frac{k\nu}{h}\rfloor-\lfloor\frac{k(\nu-1)}{h}\rfloor & 1\le\nu\le h-1\\ k-1-\lfloor\frac{k(h-1)}{h}\rfloor & \nu=h\end{cases}$$

da $\frac{k\nu}{h}$ für $0<\nu<h$ nicht ganzzahlig ist. Es ergibt sich

$$\begin{aligned}
\sum_{r=1}^{k-1}\left\lfloor\frac{hr}{k}\right\rfloor\left(\left\lfloor\frac{hr}{k}\right\rfloor+1\right) &= \sum_{\nu=1}^{h}(\nu-1)\nu N(\nu)\\
&= \sum_{\nu=1}^{h}(\nu-1)\nu\left(\left\lfloor\frac{k\nu}{h}\right\rfloor-\left\lfloor\frac{k(\nu-1)}{h}\right\rfloor\right)-h(h-1)\\
&= \sum_{\nu=1}^{h-1}\left\lfloor\frac{k\nu}{h}\right\rfloor((\nu-1)\nu-\nu(\nu+1))+kh(h-1)-h(h-1)\\
&= -2\sum_{\nu=1}^{h-1}\nu\left\lfloor\frac{k\nu}{h}\right\rfloor+h(h-1)(k-1).
\end{aligned}$$

Andererseits erhält man aus der Definition von $s(k,h)$

$$2hs(k,h) = 2\sum_{\nu=1}^{h-1} \nu\Big(\frac{k\nu}{h} - \Big\lfloor \frac{k\nu}{h}\Big\rfloor - \frac{1}{2}\Big) = -2\sum_{\nu=1}^{h-1} \nu \Big\lfloor \frac{k\nu}{h}\Big\rfloor + \frac{2k}{h}\sum_{\nu=1}^{h-1}\nu^2 - \sum_{\nu=1}^{h-1}\nu \tag{2.12}$$

und damit

$$\sum_{r=1}^{k-1} \Big\lfloor \frac{hr}{k}\Big\rfloor \Big(\Big\lfloor \frac{hr}{k}\Big\rfloor + 1\Big) = 2hs(k,h) - \frac{2k}{h}\sum_{\nu=1}^{h-1}\nu^2 + \sum_{\nu=1}^{h-1}\nu + h(h-1)(k-1).$$

Setzen wir dies in (2.11) ein, so erhalten wir nach Multiplikation mit $6k$ unter Verwendung der bekannten Summenformeln genau das Reziprozitätsgesetz. $\square$

2.3 Beweis von Dedekinds Funktionalgleichung

Nun können wir zum eigentlichen Beweis von Satz 2.1 kommen. Wir wollen

$$\eta(A\tau) = \varepsilon(A)(-i(c\tau + d))^{1/2}\eta(\tau) \tag{2.13}$$

für alle $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ mit $c > 0$ zeigen, wobei

$$\varepsilon(A) = \exp\Big(\pi i\Big(\frac{a+d}{12c} + s(-d,c)\Big)\Big).$$

Für $A = S$ wurde dies bereits in Lemma 2.3 gezeigt, und wegen $\eta(\tau + m) = e^{\pi m i/12}\eta(\tau)$ gilt sie auch für $T^m S$, was der Transformation $\tau \mapsto m - \frac{1}{\tau}$ entspricht. Weiters wissen wir auch schon, dass Γ von S und T erzeugt wird: jedes $A \in \Gamma$ hat eine Darstellung

$$A = T^{n_1} S T^{n_2} S \dots S T^{n_r}.$$

Wir erhalten die Gleichung für beliebiges A in folgenden Schritten:

1. Wenn $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ und $c > 0$, dann gilt für alle $m \in \mathbb{Z}$

$$\varepsilon(AT^m) = e^{\pi i m/12}\varepsilon(A).$$

2. Wenn $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ und $c > 0$, dann gilt

$$\varepsilon(AS) = \begin{cases} e^{-\pi i/4}\varepsilon(A) & d > 0 \\ e^{\pi i/4}\varepsilon(A) & d < 0. \end{cases}$$

3. Wenn (2.13) für ein A gilt, dann auch für AT^m und für AS.

Damit folgt dann (2.13) induktiv für alle A.

Ad 1.: es ist $AT^m = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & am+b \\ c & cm+d \end{pmatrix}$, also

$$\varepsilon(AT^m) = \exp\Big(\pi i\Big(\frac{a+cm+d}{12c} + s(-cm-d, c)\Big)\Big).$$

Da $s(-d, c)$ aber nur von der Kongruenzklasse von d modulo c abhängt, folgt 1.

Ad 2.: es ist $AS = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b & -a \\ d & -c \end{pmatrix}$. Falls $d > 0$, wird AS durch $\begin{pmatrix} b & -a \\ d & -c \end{pmatrix}$ repräsentiert, sonst durch $\begin{pmatrix} -b & a \\ -d & c \end{pmatrix}$. Für $d > 0$ ergibt sich dann

$$\varepsilon(AS) = \exp\Big(\pi i\Big(\frac{b-c}{12d} + s(c, d)\Big)\Big).$$

Aus dem Reziprozitätsgesetz erhalten wir

$$s(c, d) + s(d, c) = \frac{c}{12d} + \frac{d}{12c} - \frac{1}{4} + \frac{1}{12cd}.$$

Im letzten Bruch ersetzen wir den Zähler durch $1 = ad - bc$ und formen um zu

$$\frac{b-c}{12d} + s(c, d) = \frac{a+d}{12c} - s(d, c) - \frac{1}{4}.$$

Daraus ergibt sich sofort $\varepsilon(AS) = e^{-\pi i/4}\varepsilon(A)$. Für $d < 0$ verläuft der Beweis analog.

Ad 3.: wir haben zu zeigen, dass aus

$$\eta(A\tau) = \varepsilon(A)(-i(c\tau + d))^{1/2}\eta(\tau) \tag{2.14}$$

die Gleichungen

$$\eta(AT^m\tau) = \varepsilon(AT^m)(-i(c\tau + d + mc))^{1/2}\eta(\tau) \tag{2.15}$$
$$\eta(AS\tau) = \varepsilon(AS)(-i(d\tau - c))^{1/2}\eta(\tau) \ (d > 0) \tag{2.16}$$
$$\eta(AS\tau) = \varepsilon(AS)(-i(-d\tau + c))^{1/2}\eta(\tau) \ (d < 0) \tag{2.17}$$

folgen. Für die erste Gleichung setzen wir $T^m\tau$ für τ in (2.14) ein und erhalten unter Verwendung von 1.

$$\begin{aligned} \eta(AT^m\tau) &= \varepsilon(A)(-i(cT^m\tau + d))^{1/2}\eta(T^m\tau) \\ &= \varepsilon(A)(-i(c\tau + mc + d))^{1/2}e^{\pi i m/12}\eta(\tau) \\ &= \varepsilon(AT^m)(-i(c\tau + d + mc))^{1/2}\eta(\tau). \end{aligned}$$

Ersetzen wir ebenso τ durch $S\tau$ in (2.14), dann ergibt sich für $d > 0$ (der andere Fall verläuft wieder analog) unter Verwendung von 2.

$$\begin{aligned}
\eta(AS\tau) &= \varepsilon(A)(-i(-\frac{c}{\tau}+d))^{1/2}\eta(S\tau) \\
&= \varepsilon(A)(-i(-\frac{c}{\tau}+d))^{1/2}(-i\tau)^{1/2}\eta(\tau) \\
&= \varepsilon(A)(-(d\tau-c))^{1/2}\eta(\tau) \\
&= \varepsilon(A)e^{-\pi i/4}(-i(d\tau-c))^{1/2}\eta(\tau) \\
&= \varepsilon(AS)(-i(d\tau-c))^{1/2}\eta(\tau),
\end{aligned}$$

womit der Beweis von Satz 2.1 vollständig ist. □

2.4 Die Transformationsformel für die erzeugende Funktion von $p(n)$

Proposition 2.6 Sei $F(x)$ wie zu Beginn des Kapitels die erzeugende Funktion der Partitionsfunktion $p(n)$ und seien x, x' gegeben durch

$$x = \exp\Big(\frac{2\pi ih}{k} - \frac{2\pi z}{k^2}\Big) \text{ und } x' = \exp\Big(\frac{2\pi iH}{k} - \frac{2\pi}{z}\Big),$$

wobei $\mathrm{Re}(z) > 0$ und h, k, H ganze Zahlen mit $k > 0$, $(h,k) = 1$ und $hH \equiv -1 \bmod k$ sind. Dann gilt

$$F(x) = e^{\pi is(h,k)}\Big(\frac{z}{k}\Big)^{1/2}\exp\Big(\frac{\pi}{12z} - \frac{\pi z}{12k^2}\Big)F(x'). \tag{2.18}$$

Beweis: Offenbar gilt $F(e^{2\pi i\tau}) = e^{\pi i\tau/12}/\eta(\tau)$. Aus der Funktionalgleichung für $\eta(\tau)$ ergibt sich für $\tau' = \frac{a\tau+b}{c\tau+d}$

$$\frac{1}{\eta(\tau)} = \frac{1}{\eta(\tau')}(-i(c\tau+d))^{1/2}\exp\Big(\pi i\Big(\frac{a+d}{12c} + s(-d,c)\Big)\Big)$$

und damit weiters

$$F(e^{2\pi i\tau}) = F(e^{2\pi i\tau'})\exp\Big(\frac{\pi i(\tau-\tau')}{12}\Big)(-i(c\tau+d))^{1/2}\exp\Big(\pi i\Big(\frac{a+d}{12c} + s(-d,c)\Big)\Big).$$

Wählen wir nun $a = H$, $c = k$, $d = -h$, $b = -\frac{hH+1}{k}$ und $\tau = \frac{iz+h}{k}$, dann ergibt sich $\tau' = \frac{iz^{-1}+H}{k}$ und damit

$$F\Big(\exp\Big(\frac{2\pi ih}{k} - \frac{2\pi z}{k}\Big)\Big) = F\Big(\exp\Big(\frac{2\pi iH}{k} - \frac{2\pi}{kz}\Big)\Big)z^{1/2}\exp\Big(\frac{\pi}{12kz} - \frac{\pi z}{12k} + \pi is(h,k)\Big).$$

Ersetzt man z durch $\frac{z}{k}$, so ergibt sich die Behauptung. □

Bemerkung: Der aufwändigste Teil hierbei ist die Identifizierung des Faktors $\omega(h,k) = e^{\pi i s(h,k)}$. Durch den Zugang über elliptische Funktionen kommt man auch ohne die vorangegangenen Rechnungen bereits zu dem Schluss, dass dieser Faktor konstant sein muss (also nicht von z abhängt) und $\exp\left(\pi i \frac{H-h}{12k}\right)\omega(h,k)$ stets eine 24-te Einheitswurzel ist. Richmond und Szekeres [25] geben einen elementaren Beweis dafür an, dass $\omega(h,k) = e^{\pi i s(h,k)}$ gilt, wobei nur verwendet wird, dass $\omega(h,k)$ eine von h und k abhängige Konstante ist.

2.5 Farey-Brüche und Ford-Kreise

Bevor wir zur Herleitung von Rademachers Formel kommen können, muss noch der Integrationspfad beschrieben werden, der dazu verwendet wird. Dieser greift auf die sogenannten *Ford-Kreise* zurück, die ihrerseits auf *Farey-Brüche* zurückgehen:

Definition 2.2 Die Folge F_n der Farey-Brüche n-ter Ordnung ist die Menge der gekürzten Brüche im Intervall $[0,1]$ mit Nenner $\leq n$, in aufsteigender Reihenfolge angeschrieben.

Beispiel: $F_7 = (0, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, 1)$.

Lemma 2.7 Für zwei verschiedene Brüche $\frac{a}{b} < \frac{c}{d}$ $(b,d > 0)$ liegt der Mediant $\frac{a+c}{b+d}$ zwischen den beiden Brüchen.

Beweis:
$$\frac{a+c}{b+d} - \frac{a}{b} = \frac{bc-ad}{b(b+d)} > 0 \text{ und } \frac{a+c}{b+d} - \frac{c}{d} = \frac{ad-bc}{d(b+d)} < 0.$$

□

Proposition 2.8 Wenn zwei Brüche $0 \leq \frac{a}{b} < \frac{c}{d} \leq 1$ die unimodulare Bedingung $bc - ad = 1$ erfüllen, dann sind sie für $\max(b,d) \leq n \leq b+d-1$ aufeinanderfolgende Farey-Brüche in F_n.

Beweis: Aus $bc - ad$ folgt bereits, dass $(a,b) = (c,d) = 1$, d.h. die Brüche sind bereits gekürzt. Da $n \geq \max(b,d)$, liegen beide Brüche in F_n. Angenommen nun, es gäbe einen Bruch $\frac{h}{k}$ mit Nenner $< b+d$, der zwischen den beiden liegt. Dann gilt aber

$$k = (bc-ad)k = b(ck-dh) + d(bh-ak)$$

$ck - dh$ und $bh - ak$ sind ganze Zahlen, und wegen $\frac{a}{b} < \frac{h}{k} < \frac{c}{d}$ beide positiv. Damit folgt aber $k \geq b+d$, ein Widerspruch. □

Lemma 2.9 Wenn zwei Brüche $0 \leq \frac{a}{b} < \frac{c}{d} \leq 1$ die unimodulare Bedingung $bc - ad = 1$ erfüllen, dann erfüllen auch $\frac{a}{b}, \frac{h}{k}$ bzw. $\frac{h}{k}, \frac{c}{d}$ diese Bedingung, wobei $\frac{h}{k} = \frac{a+c}{b+d}$ der Mediant ist.

Beweis: Es gilt

$bh - ak = b(a + c) - a(b + d) = bc - ad = 1$ und $ck - dh = c(b + d) - d(a + c) = bc - ad = 1.$

□

Satz 2.10 *Jeder Bruch aus F_{n+1}, der nicht bereits in F_n liegt, ist Mediant zweier aufeinanderfolgender Brüche aus F_n. Zwei aufeinanderfolgende Brüche $\frac{a}{b} < \frac{c}{d}$ in F_n erfüllen die unimodulare Eigenschaft $bc - ad = 1$.*

Beweis: Durch Induktion nach n. Für $n = 1$ ist die Aussage trivial. Sei $\frac{h}{k}$ ein Bruch, der in F_{n+1}, aber nicht in F_n liegt, und zwischen zwei Brüche $\frac{a}{b} < \frac{c}{d}$ aus F_n eingefügt wird. Nach Induktionsvoraussetzung ist die unimodulare Eigenschaft $bc - ad = 1$ erfüllt, daher muss nach Proposition 2.8 $n + 1 \geq b + d$ gelten.

Andererseits gilt für den Medianten $\frac{a+c}{b+d}$, dass $b(a + c) - a(b + d) = 1$, also $(a + c, b + d) = 1$, d.h. der Bruch $\frac{a+c}{b+d}$ ist gekürzt und liegt somit in F_{n+1}. Da weiters nach Lemma 2.9 und Proposition 2.8 $\frac{a}{b}$ und $\frac{a+c}{b+d}$ bzw. $\frac{a+c}{b+d}$ und $\frac{c}{d}$ in F_{b+d} aufeinanderfolgen, muss $\frac{h}{k} = \frac{a+c}{b+d}$ gelten, d.h. $\frac{h}{k}$ ist tatsächlich Mediant von $\frac{a}{b}$ und $\frac{c}{d}$. Nach Lemma 2.9 ist auch weiterhin die unimodulare Eigenschaft für aufeinanderfolgende Brüche gewährleistet. □

Nun sind wir bereit, Ford-Kreise einzuführen:

Definition 2.3 Einem gekürzten Bruch $\frac{h}{k}$ wird der Ford-Kreis $C(h, k)$ zugeordnet, dessen Mittelpunkt in der komplexen Ebene $\frac{h}{k} + i\frac{1}{2k^2}$ und dessen Radius $\frac{1}{2k^2}$ ist.

Satz 2.11 *Zwei Fordkreise $C(a, b)$ und $C(c, d)$ schneiden sich entweder gar nicht, oder sie berühren einander. Sie berühren sich genau dann, wenn $|bc-ad| = 1$, wobei der Berührpunkt $\frac{ab+cd}{b^2+d^2} + i\frac{1}{b^2+d^2}$ ist. Insbesondere berühren einander die Fordkreise, die aufeinanderfolgenden Farey-Brüchen zugeordnet sind.*

Beweis: Der Abstand zwischen den beiden Mittelpunkten ist

$$D = \sqrt{\left(\frac{a}{b} - \frac{c}{d}\right)^2 + \left(\frac{1}{2b^2} - \frac{1}{2d^2}\right)^2}.$$

Wir zeigen, dass dies $\geq \frac{1}{2b^2} + \frac{1}{2d^2}$, der Summe der beiden Radien, ist:

$$\begin{aligned} D^2 - \left(\frac{1}{2b^2} + \frac{1}{2d^2}\right)^2 &= \left(\frac{a}{b} - \frac{c}{d}\right)^2 + \left(\frac{1}{2b^2} - \frac{1}{2d^2}\right)^2 - \left(\frac{1}{2b^2} + \frac{1}{2d^2}\right)^2 \\ &= \left(\frac{ad - bc}{bd}\right)^2 - \frac{1}{b^2} \cdot \frac{1}{d^2} \\ &= \frac{(ad - bc)^2 - 1}{b^2 d^2} \geq 0 \end{aligned}$$

mit Gleichheit nur, wenn $|bc - ad| = 1$. Zu zeigen bleibt daher nur noch die Formel für den Berührpunkt. Dazu genügt der Nachweis, dass der Abstand vom Mittelpunkt $\frac{a}{b} + i\frac{1}{2b^2}$ des einen Kreises zum angegebenen Punkt tatsächlich gleich dem Kreisradius $\frac{1}{2b^2}$ ist (aus Symmetriegründen folgt dies auch für den zweiten Kreis):

$$\begin{aligned}\Big(\frac{a}{b} - \frac{ab+cd}{b^2+d^2}\Big)^2 + \Big(\frac{1}{2b^2} - \frac{1}{b^2+d^2}\Big)^2 &= \Big(\frac{(ad-bc)d}{b(b^2+d^2)}\Big)^2 + \Big(\frac{d^2-b^2}{2b^2(b^2+d^2)}\Big)^2 \\ &= \frac{4b^2d^2}{4b^4(b^2+d^2)^2} + \frac{(d^2-b^2)^2}{4b^4(b^2+d^2)^2} \\ &= \frac{(b^2+d^2)^2}{4b^4(b^2+d^2)^2} \\ &= \Big(\frac{1}{2b^2}\Big)^2.\end{aligned}$$

□

Korollar 2.12 Sind $\frac{h_1}{k_1} < \frac{h}{k} < \frac{h_2}{k_2}$ drei aufeinanderfolgende Fareybrüche, dann berührt $C(h,k)$ die Kreise $C(h_1,k_1)$ und $C(h_2,k_2)$ in den Punkten

$$\alpha_1(h,k) = \frac{h}{k} - \frac{k_1}{k(k^2+k_1^2)} + i\frac{1}{k^2+k_1^2} \tag{2.19}$$

bzw.

$$\alpha_2(h,k) = \frac{h}{k} + \frac{k_2}{k(k^2+k_2^2)} + i\frac{1}{k^2+k_2^2}. \tag{2.20}$$

Beweis: Durch Einsetzen in die Berührpunktformel aus Satz 2.11:

$$\begin{aligned}\alpha_1(h,k) &= \frac{hk+h_1k_1}{k^2+k_1^2} + i\frac{1}{k^2+k_1^2} \\ &= \frac{h}{k} + \frac{hk+h_1k_1}{k^2+k_1^2} - \frac{h}{k} + i\frac{1}{k^2+k_1^2} \\ &= \frac{h}{k} + \frac{k_1(h_1k-hk_1)}{k(k^2+k_1^2)} + i\frac{1}{k^2+k_1^2} \\ &= \frac{h}{k} - \frac{k_1}{k(k^2+k_1^2)} + i\frac{1}{k^2+k_1^2}\end{aligned}$$

und analog für $\alpha_2(h,k)$. □

Nach diesen Vorbereitungen können wir den Rademacher-Pfad definieren, der im Beweis von Rademachers Formel die Schlüsselrolle spielt.

Definition 2.4 Sei $N \in \mathbb{N}$ und F_N die Folge der Farey-Brüche N-ter Ordnung. Für drei aufeinanderfolgende Farey-Brüche $\frac{h_1}{k_1} < \frac{h}{k} < \frac{h_2}{k_2}$ aus F_N wird der Ford-Kreis $C(h,k)$ durch die Berührpunkte $\alpha_1(h,k)$ und $\alpha_2(h,k)$ in einen oberen und einen unteren Bogen geteilt. Wir definieren nun die Kurve $P(N)$ zwischen i und $i+1$ als Vereinigung aller oberen Bögen, die mit $\gamma(h,k)$ bezeichnet werden.

Beispiel: Die Abbildung zeigt den Rademacher-Pfad $P(3)$.

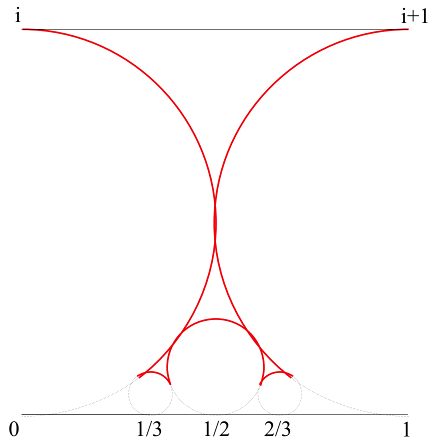

Abbildung 2.1: Der Rademacher-Pfad $P(3)$.

Lemma 2.13 Die Transformation $z = -ik^2\left(\tau - \frac{h}{k}\right)$ führt den Kreis $C(h,k)$ aus der τ-Ebene in einen Kreis vom Radius $\frac{1}{2}$ um $\frac{1}{2}$ in der z-Ebene über. Die Berührpunkte $\alpha_1(h,k)$ und $\alpha_2(h,k)$ werden dabei in die Punkte

$$z_1(h,k) = \frac{k^2}{k^2+k_1^2} + i\frac{kk_1}{k^2+k_1^2} \tag{2.21}$$

bzw.

$$z_2(h,k) = \frac{k^2}{k^2+k_2^2} - i\frac{kk_2}{k^2+k_2^2} \tag{2.22}$$

übergeführt. Der obere Bogen zwischen $\alpha_1(h,k)$ und $\alpha_2(h,k)$ geht dabei in jenen Bogen über, der 0 nicht enthält.

Beweis: Die Transformation besteht aus der Translation $\tau \mapsto \tau - \frac{h}{k}$, die den Mittelpunkt des Kreises $C(h,k)$ in den Punkt $i\frac{1}{2k^2}$ verschiebt, und der Multiplikation mit $-ik^2$, die den Radius um den Faktor k^2 auf $\frac{1}{2}$ vergrößert und gleichzeitig den Kreis um den Winkel $\frac{\pi}{2}$ im Uhrzeigersinn dreht. Damit ergibt sich der Kreis mit Radius $\frac{1}{2}$ um den Punkt $\frac{1}{2}$, wobei der Punkt $\frac{h}{k}$ des unteren Bogens in 0 übergeht. Die Ausdrücke für $z_1(h,k)$ und $z_2(h,k)$ ergeben sich durch direktes Einsetzen. □

Lemma 2.14 Für die Punkte $z_1(h,k)$ und $z_2(h,k)$ aus dem vorigen Lemma gilt

$$|z_1(h,k)| = \frac{k}{\sqrt{k^2+k_1^2}} \text{ und } |z_2(h,k)| = \frac{k}{\sqrt{k^2+k_2^2}}. \tag{2.23}$$

Zudem gilt für jeden Punkt z auf der Sehne zwischen $z_1(h,k)$ und $z_2(h,k)$ die Abschätzung

$$|z| < \frac{\sqrt{2}k}{N}. \tag{2.24}$$

Die Länge dieser Sehne ist $< \frac{2\sqrt{2}k}{N}$.

Beweis: Es gilt

$$\begin{aligned}|z_1(h,k)|^2 &= \frac{k^4}{(k^2+k_1^2)^2} + \frac{k^2k_1^2}{(k^2+k_1^2)^2}\\ &= \frac{k^2(k^2+k_1^2)}{(k^2+k_1^2)^2} = \frac{k^2}{k^2+k_1^2},\end{aligned}$$

woraus obige Formel folgt. Für $z_2(h,k)$ prüft man analog nach. Weiters gilt für jeden Punkt z auf der Sehne zwischen $z_1(h,k)$ und $z_2(h,k)$ jedenfalls $|z| \le \max(z_1(h,k), z_2(h,k))$. Es genügt daher, $|z_1(h,k)| < \frac{\sqrt{2}k}{N}$ zu beweisen, für $z_2(h,k)$ ergibt sich die Abschätzung wieder in analoger Weise.

Da $\frac{h}{k}$ und $\frac{h_1}{k_1}$ aufeinanderfolgende Farey-Brüche in F_N sind, liegt ihr Mediant nicht in F_N. Daher muss $k+k_1 > N$ gelten. Es folgt nach der Ungleichung zwischen dem arithmetischen und dem quadratischen Mittel

$$\sqrt{\frac{k^2+k_1^2}{2}} \ge \frac{k+k_1}{2} > \frac{N}{2}$$

und damit

$$|z_1(h,k)| = \frac{k}{\sqrt{k^2+k_1^2}} \le \frac{k}{N/\sqrt{2}} = \frac{\sqrt{2}k}{N}.$$

Nach der Dreiecksungleichung ist schließlich die Länge der Sehne $\le |z_1(h,k)| + |z_2(h,k)|$ und damit $< \frac{2\sqrt{2}k}{N}$. □

2.6 Herleitung der Formel

Satz 2.15 (Rademachers Formel für die Partitionsfunktion) *Die Partitionsfunktion $p(n)$ ist für $n \ge 1$ durch die folgende Reihe gegeben:*

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n)\sqrt{k}\frac{d}{dn}\left(\frac{\sinh\left(\frac{\pi}{k}\sqrt{\frac{2}{3}\left(n-\frac{1}{24}\right)}\right)}{\sqrt{n-1/24}}\right) \tag{2.25}$$

wobei

$$A_k(n) = \sum_{\substack{0\le h<k\\(h,k)=1}} e^{\pi i s(h,k) - 2\pi i n h/k}. \tag{2.26}$$

BEMERKUNG: Es gilt

$$\frac{d}{dn}\left(\frac{\sinh\left(\frac{\pi}{k}\sqrt{\frac{2}{3}\left(n-\frac{1}{24}\right)}\right)}{\sqrt{n-1/24}}\right) \sim \frac{\pi\sqrt{2}}{4kn\sqrt{3}}\exp\left(\frac{\pi}{k}\sqrt{\frac{2n}{3}}\right),$$

weswegen der Term bei $k=1$ der Hauptterm der Asymptotik sein muss. Für $k=1$ ist $A_k(n)=1$ und damit

$$p(n) \sim \frac{\exp(\pi\sqrt{2n/3})}{4\sqrt{3}n},$$

wie eingangs erwähnt wurde.

Beweis: Wir beginnen mit dem Integral (2.4):

$$p(n) = \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{F(x)}{x^{n+1}}\,dx.$$

$\mathcal{C}$ stellt eine noch festzulegende Kurve dar, die sich im Inneren der Kreisscheibe um 0 vom Radius 1 befindet. Im Inneren dieser Kreisscheibe ist $F(x)=\prod_{m=1}^{\infty}(1-x^m)^{-1}$ holomorph. Die Substitution $x=e^{2\pi i\tau}$ führt diesen Kreis in den Streifen mit Realteil $0 \le \operatorname{Re}\tau < 1$ und positivem Imaginärteil über. Wenn x entlang einem Kreis vom Radius $e^{-2\pi}$ um 0 läuft, dann läuft τ genau entlang der Horizontalen zwischen i und $i+1$. Diese wird durch den Rademacherpfad $P(N)$ ersetzt. Dann wird aus obiger Gleichung

$$p(n) = \int_i^{i+1} F(e^{2\pi i\tau})e^{-2\pi i n\tau}\,d\tau = \int_{P(N)} F(e^{2\pi i\tau})e^{-2\pi i n\tau}\,d\tau. \tag{2.27}$$

Dabei halten wir n fest und lassen N nach ∞ streben. Wir können $P(N)$ als Vereinigung der Bögen $\gamma(h,k)$ schreiben, womit sich für das Integral

$$\int_{P(N)} = \sum_{k=1}^{N}\sum_{\substack{0\le h<k\\(h,k)=1}}\int_{\gamma(h,k)}$$

ergibt. Für die Doppelsumme wird im Weiteren nur $\sum_{h,k}$ geschrieben. Für jeden der Bögen führen wir die Substitution $z=-ik^2\left(\tau-\frac{h}{k}\right)$ bzw. $\tau=\frac{h}{k}+\frac{iz}{k^2}$ durch. Nach Lemma 2.13 führt diese den Fordkreis $C(h,k)$ in einen Kreis mit Mittelpunkt $\frac{1}{2}$ und Radius $\frac{1}{2}$ über. Der Bogen $\gamma(h,k)$ wird dabei in den Bogen zwischen den Punkten $z_1(h,k)$ und $z_2(h,k)$ übergeführt, wobei $z_1(h,k)$ und $z_2(h,k)$ in Lemma 2.13 gegeben sind. Wir haben daher

$$\begin{aligned} p(n) &= \sum_{h,k}\int_{z_1(h,k)}^{z_2(h,k)} F\Big(\exp\Big(\frac{2\pi ih}{k}-\frac{2\pi z}{k^2}\Big)\Big)\frac{i}{k^2}e^{-2\pi inh/k}e^{2\pi nz/k^2}\,dz \\ &= \sum_{h,k} ik^{-2}e^{-2\pi inh/k}\int_{z_1(h,k)}^{z_2(h,k)} e^{2\pi nz/k^2}F\Big(\exp\Big(\frac{2\pi ih}{k}-\frac{2\pi z}{k^2}\Big)\Big)\,dz. \end{aligned} \tag{2.28}$$

Nun verwenden wir die Transformationsformel für $F(x)$ aus Proposition 2.6:

$$F(x) = \omega(h,k)\Big(\frac{z}{k}\Big)^{1/2} \exp\Big(\frac{\pi}{12z} - \frac{\pi z}{12k^2}\Big)F(x'),$$

wobei

$$x = \exp\Big(\frac{2\pi ih}{k} - \frac{2\pi z}{k^2}\Big) \text{ und } x' = \exp\Big(\frac{2\pi iH}{k} - \frac{2\pi}{z}\Big)$$

und $\omega(h,k) = e^{\pi is(h,k)}$ sowie $hH \equiv -1 \bmod k$. Wir bezeichnen den elementaren Faktor $z^{1/2}\exp\Big(\frac{\pi}{12z} - \frac{\pi z}{12k^2}\Big)$ mit $\Psi_k(z)$ und spalten das Integral via $F(x') = 1 + (F(x') - 1)$ in zwei Teile auf. Wir erhalten zunächst

$$p(n) = \sum_{h,k} ik^{-5/2}\omega(h,k)e^{-2\pi inh/k}(I_1(h,k) + I_2(h,k)), \tag{2.29}$$

wobei $I_1(h,k)$ und $I_2(h,k)$ die beiden Teile des Integrals bezeichnet:

$$I_1(h,k) = \int_{z_1(h,k)}^{z_2(h,k)} \Psi_k(z)e^{2\pi nz/k^2}\,dz,$$

$$I_2(h,k) = \int_{z_1(h,k)}^{z_2(h,k)} \Psi_k(z)e^{2\pi nz/k^2}\Big(F\Big(\exp\Big(\frac{2\pi iH}{k} - \frac{2\pi}{z}\Big)\Big) - 1\Big)\,dz.$$

Es stellt sich heraus, dass $I_2(h,k)$ für großes N klein und daher vernachlässigbar ist. Wir können hierfür als Integrationspfad die Sehne zwischen $z_1(h,k)$ und $z_2(h,k)$ verwenden, deren Länge in Lemma 2.14 mit $2\sqrt{2}k/N$ von oben abgeschätzt wurde. Für jeden Punkt dieser Sehne gilt nach demselben Lemma $|z| < \sqrt{2}k/N$. Weiters wird der Kreis um $\frac{1}{2}$ mit Radius $\frac{1}{2}$ durch die Abbildung $w = \frac{1}{z}$ auf die Halbebene $\mathrm{Re}(w) \geq 1$ abgebildet. Daher gilt auf der Sehne $0 < \mathrm{Re}(z) \leq 1$ und $\mathrm{Re}(1/z) \geq 1$. Der Integrand entlang der Sehne kann somit wie folgt abgeschätzt werden:

$$\begin{aligned}
&\Big|\Psi_k(z)e^{2\pi nz/k^2}\Big(F\Big(\exp\Big(\frac{2\pi iH}{k} - \frac{2\pi}{z}\Big)\Big) - 1\Big)\Big| = \\
&\qquad = |z|^{1/2}\exp\Big(\frac{\pi}{12}\mathrm{Re}\Big(\frac{1}{z}\Big) - \frac{\pi}{12k^2}\mathrm{Re}(z)\Big)\cdot e^{2\pi n\,\mathrm{Re}(z)/k^2}\Big|\sum_{m=1}^{\infty} p(m)e^{2\pi iHm/k}e^{-2\pi m/z}\Big| \\
&\qquad \leq |z|^{1/2}\exp\Big(\frac{\pi}{12}\mathrm{Re}\Big(\frac{1}{z}\Big)\Big)e^{2\pi n/k^2}\sum_{m=1}^{\infty} p(m)e^{-2\pi m\,\mathrm{Re}(1/z)} \\
&\qquad < |z|^{1/2}e^{2\pi n}\sum_{m=1}^{\infty} p(m)e^{-2\pi(m-1/24)\,\mathrm{Re}(1/z)} \\
&\qquad \leq |z|^{1/2}e^{2\pi n}\sum_{m=1}^{\infty} p(m)e^{-2\pi(m-1/24)} \\
&\qquad = c|z|^{1/2},
\end{aligned}$$

wobei

$$c = e^{2\pi n} \sum_{m=1}^{\infty} p(m) e^{-2\pi(m-1/24)} = e^{2\pi n+\pi/12}(F(e^{-2\pi}) - 1) < \infty$$

nur von n, nicht aber von N oder z abhängt. Daher ist der Integrand auf der Sehne zwischen $z_1(h,k)$ und $z_2(h,k)$ durch $c|z|^{1/2} \leq c(\sqrt{2}k/N)^{1/2} = c2^{1/4}(k/N)^{1/2}$ beschränkt. Da die Länge des Weges $\leq 2\sqrt{2}k/N$ ist, gilt insgesamt

$$|I_2(h,k)| < Ck^{3/2}N^{-3/2}$$

für eine gewisse Konstante C und damit

$$\begin{aligned}\Big|\sum_{h,k} ik^{-5/2}\omega(h,k)e^{-2\pi inh/k} I_2(h,k)\Big| &< \sum_{k=1}^{N} \sum_{\substack{0\leq h<k \\ (h,k)=1}} Ck^{-1}N^{-3/2} \\ &\leq CN^{-3/2}\sum_{k=1}^{N} 1 = CN^{-1/2}.\end{aligned}$$

Wir können daher auch

$$p(n) = \sum_{h,k} ik^{-5/2}\omega(h,k)e^{-2\pi inh/k} I_1(h,k) + O(N^{-1/2}) \tag{2.30}$$

schreiben. Um $I_1(h,k)$ berechnen zu können, dehnen wir den Integrationsbereich vom Bogen zwischen $z_1(h,k)$ und $z_2(h,k)$ auf den ganzen Kreis aus und zeigen, dass wieder nur ein Fehlerterm von $O(N^{-1/2})$ auftritt:

$$\int_{z_1(h,k)}^{z_2(h,k)} = \int_{K(-)} - \int_0^{z_1(h,k)} - \int_{z_2(h,k)}^0$$

wobei $\int_{K(-)}$ Integration in negativer Richtung über den Kreis K bedeutet. Sei J_1 das Integral über den Bogen zwischen 0 und $z_1(h,k)$. Die Länge dieses Bogens ist $\leq \pi|z_1(h,k)| \leq \pi\sqrt{2}k/N$, und auf dem Bogen gilt $|z| \leq |z_1(h,k)| \leq \sqrt{2}k/N$. Da weiters auf ganz K $0 \leq \mathrm{Re}(z) \leq 1$ und $\mathrm{Re}(1/z) = 1$ gilt, folgt für den Integranden

$$|\Psi_k(z)e^{2\pi nz/k^2}| = e^{2\pi n\,\mathrm{Re}(z)/k^2}|z|^{1/2}\exp\left(\frac{\pi}{12}\,\mathrm{Re}\left(\frac{1}{z}\right) - \frac{\pi}{12k^2}\,\mathrm{Re}(z)\right) \leq \frac{e^{2\pi n}2^{1/4}k^{1/2}e^{\pi/12}}{N^{1/2}}$$

und damit $|J_1| \leq C_1k^{3/2}N^{-3/2}$, wobei C_1 eine Konstante ist, die wiederum nur von n abhängt. Analog zeigt man $|J_2| \leq C_2k^{3/2}N^{-3/2}$ für das Integral über den Bogen zwischen $z_2(h,k)$ und 0. Wie in der Abschätzung für I_2 erhält man durch das Aufsummieren über alle h,k einen Fehlerterm der Ordnung $O(N^{-1/2})$. Damit ergibt sich

$$p(n) = \sum_{h,k} ik^{-5/2}\omega(h,k)e^{-2\pi inh/k}\int_{K(-)} \Psi_k(z)e^{2\pi nz/k^2}\,dz + O(N^{-1/2}).$$

Lassen wir nun N nach ∞ streben, so erhalten wir

$$p(n) = i \sum_{k=1}^{\infty} A_k(n) k^{-5/2} \int_{K(-)} z^{1/2} \exp\Big(\frac{\pi}{12z} + \frac{2\pi z}{k^2}\Big(n - \frac{1}{24}\Big)\Big)\, dz, \tag{2.31}$$

wobei

$$A_k(n) = \sum_{\substack{0 \le h < k \\ (h,k)=1}} e^{\pi i s(h,k) - 2\pi i n h/k}.$$

Es bleibt lediglich die Berechnung des Integrals übrig, wobei nunmehr aber über eine elementare Funktion integriert wird. Zunächst wird vermittels der Substitution $w = \frac{1}{z}$ $(dw = -w^2 dz)$ der Kreis K in die Gerade $\{1 + it : \ t \in \mathbb{R}\}$ übergeführt, womit sich

$$p(n) = -i \sum_{k=1}^{\infty} A_k(n) k^{-5/2} \int_{1-i\infty}^{1+i\infty} w^{-5/2} \exp\Big(\frac{\pi w}{12} + \frac{2\pi}{k^2 w}\Big(n - \frac{1}{24}\Big)\Big)\, dw \tag{2.32}$$

ergibt. Die weitere Substitution $t = \frac{\pi w}{12}$ macht daraus

$$p(n) = 2\pi \Big(\frac{\pi}{12}\Big)^{3/2} \sum_{k=1}^{\infty} A_k(n) k^{-5/2} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} t^{-5/2} \exp\Big(t + \frac{\pi^2}{6k^2 t}\Big(n - \frac{1}{24}\Big)\Big)\, dt, \tag{2.33}$$

wobei $c = \pi/12$. Dieses Integral zeigt eine Darstellungsform für die Bessel-Funktion erster Art (siehe [38]): diese ist definiert als

$$J_\nu(z) := \sum_{m=0}^{\infty} \frac{(-1)^m (z/2)^{\nu+2m}}{m!\Gamma(\nu + m + 1)}. \tag{2.34}$$

Setzt man für die Gamma-Funktion das bekannte Integral

$$\frac{1}{\Gamma(\nu + m + 1)} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} t^{-\nu-m-1} e^t \, dt$$

(wobei $c > 0$ ist) ein, so erhält man

$$\begin{aligned}
J_\nu(z) &= \frac{(z/2)^\nu}{2\pi i} \sum_{m=0}^{\infty} \int_{c-i\infty}^{c+i\infty} \frac{(-1)^m (z/2)^{2m}}{m!} t^{-\nu-m-1} e^t \, dt \\
&= \frac{(z/2)^\nu}{2\pi i} \int_{c-i\infty}^{c+i\infty} \sum_{m=0}^{\infty} \frac{(-1)^m (z/2)^{2m}}{m!} t^{-\nu-m-1} e^t \, dt \\
&= \frac{(z/2)^\nu}{2\pi i} \int_{c-i\infty}^{c+i\infty} t^{-\nu-1} e^t \sum_{m=0}^{\infty} \frac{1}{m!} \Big(-\frac{z^2}{4t}\Big)^m dt \\
&= \frac{(z/2)^\nu}{2\pi i} \int_{c-i\infty}^{c+i\infty} t^{-\nu-1} \exp\Big(t - \frac{z^2}{4t}\Big)\, dt,
\end{aligned}$$

und daher folgt mit $I_\nu(z) = i^{-\nu} J_\nu(iz)$

$$I_\nu(z) = \frac{(z/2)^\nu}{2\pi i} \int_{c-i\infty}^{c+i\infty} t^{-\nu-1} e^{t+z^2/4t}\, dt, \tag{2.35}$$

falls c, $\mathrm{Re}(\nu) > 0$. Setzt man $z = 2\Big(\frac{\pi^2}{6k^2}\Big(n - \frac{1}{24}\Big)\Big)^{1/2}$ und $\nu = 3/2$, so ergibt sich genau

$$\begin{aligned} p(n) &= 2\pi\Big(\frac{\pi}{12}\Big)^{3/2} \sum_{k=1}^{\infty} A_k(n) k^{-5/2} \frac{\pi^{-3/2}\Big(n-\frac{1}{24}\Big)^{-3/4}}{6^{-3/4}k^{-3/2}} I_{3/2}\Big(\frac{\pi}{k}\sqrt{\frac{2}{3}\Big(n-\frac{1}{24}\Big)}\Big) \\ &= \frac{2\pi\Big(n-\frac{1}{24}\Big)^{-3/4}}{24^{3/4}} \sum_{k=1}^{\infty} A_k(n) k^{-1} I_{3/2}\Big(\frac{\pi}{k}\sqrt{\frac{2}{3}\Big(n-\frac{1}{24}\Big)}\Big). \end{aligned} \tag{2.36}$$

Verwendet man schließlich noch, dass sich die Besselfunktion J_ν für Werte von der Form $\nu = l+1/2$ ($l \in \mathbb{Z}$) durch elementare Funktionen darstellen lässt, und zwar hier insbesondere

$$\begin{aligned} I_{3/2}(z) = i^{-3/2} J_{3/2}(iz) &= i^{-3/2} \sum_{m=0}^{\infty} \frac{i^{3/2}(z/2)^{2m+3/2}}{m!\Gamma(m+5/2)} \\ &= \sum_{m=0}^{\infty} \frac{(z/2)^{2m+3/2}}{m!\Gamma(1/2)2^{-m-2}(2m+3)!!} \\ &= \frac{1}{\sqrt{\pi}} \sum_{m=0}^{\infty} \frac{(m+1)!(z/2)^{2m+3/2}}{m!2^{-2m-3}(2m+3)!} \\ &= \sqrt{\frac{2z}{\pi}} \sum_{m=0}^{\infty} \frac{(2m+2)z^{2m+1}}{(2m+3)!} \\ &= \sqrt{\frac{2z}{\pi}} \frac{d}{dz}\Big(\sum_{m=0}^{\infty} \frac{z^{2m+2}}{(2m+3)!} + 1\Big) \\ &= \sqrt{\frac{2z}{\pi}} \frac{d}{dz}\Big(\frac{\sinh z}{z}\Big), \end{aligned}$$

so erhält man durch Einsetzen genau Rademachers Formel:

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n)\sqrt{k} \frac{d}{dn}\left(\frac{\sinh\left(\frac{\pi}{k}\sqrt{\frac{2}{3}\Big(n-\frac{1}{24}\Big)}\right)}{\sqrt{n-1/24}}\right). \tag{2.37}$$

□

Bemerkung: Eine ähnliche Formel wurde von Hua [12] für die Anzahl der Partitionen von n in lauter ungleiche Teile gefunden. Bezeichne $q(n)$ diese Anzahl, dann gilt

$$q(n) = \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} B_{2k-1}(n) \frac{d}{dn} \left(J_0\Big(\frac{\pi i}{2k-1} \sqrt{\frac{1}{3}\Big(n + \frac{1}{24}\Big)} \Big) \right), \tag{2.38}$$

wobei

$$B_k(n) = \sum_{\substack{0 \le h < k \\ (h,k)=1}} e^{\pi i (s(h,k) - s(2h,k)) - 2\pi i n h/k}$$

und J_0 wieder die Besselfunktion (0-ter Ordnung) bezeichnet. Der Hauptterm ist hier

$$q(n) \sim \frac{\exp(\pi \sqrt{n/3})}{4 \cdot 3^{1/4} n^{3/4}}. \tag{2.39}$$

Die folgende Tabelle zeigt einige Werte von $p(n)$ und $q(n)$. $p(n)$ ist Folge Nr. A000041 in "Sloanes On-Line Encyclopedia of Integer Sequences" [27], $q(n)$ Folge Nr. A000009.

n	$p(n)$	$q(n)$
1	1	1
2	2	1
3	3	2
4	5	2
5	7	3
6	11	4
7	15	5
8	22	6
9	30	8
10	42	10
20	627	64
50	204226	3658
100	190569292	444793
1000	24061467864032622473692149727991	8635565795744155161506
10000	$3.6167251325636292 \cdot 10^{106}$	$1.1226065745480385 \cdot 10^{75}$
100000	$2.749351056977570 \cdot 10^{346}$	$4.2494159403332316 \cdot 10^{244}$

Tabelle 2.1: Einige Werte von $p(n)$ und $q(n)$.

Kapitel 3

Verteilung von Partitionen nach Länge und Maximum

1941 untersuchten Erdös und Lehner [3] erstmals die Verteilung von Partitionen einer natürlichen Zahl n nach ihrer Länge. Sie konnten beweisen, dass für $k = \frac{1}{2c}\sqrt{n}\log n + x\sqrt{n}$ (wobei $c = \frac{\pi}{\sqrt{6}}$; diese Konstante wird im Folgenden von großer Bedeutung sein) gilt:

$$\lim_{n\to\infty} \frac{P(n,k)}{p(n)} = \exp\Big(- \frac{1}{c} e^{-cx} \Big). \tag{3.1}$$

Die Funktion auf der rechten Seite erfüllt alle Voraussetzungen einer Verteilungsfunktion: sie ist positiv und streng monoton, und sie strebt für $x \to \infty$ nach 1 sowie für $x \to -\infty$ nach 0. Dieses Resultat bedeutet auch, dass für eine beliebige Funktion $\omega(n)$, die mit n nach ∞ strebt, fast alle Partitionen von n im Intervall $\frac{1}{2c}\sqrt{n}\log n \pm \omega(n)\sqrt{n}$ liegen.

Szekeres verfeinerte die Ergebnisse von Erdös und Lehner in mehreren Arbeiten [28, 29, 31] und bestimmte auch eine Asymptotik für den Fall, dass Länge und Maximum gleichzeitig eingeschränkt werden. Dieser Satz soll das Hauptthema dieses Kapitels bilden. Andere Partitionsstatistiken, wie z.B. die kleinste Lücke oder der größte mehrfach vorkommende Teil wurden von Grabner und Knopfmacher in [6] behandelt.

Satz 3.1 (Szekeres [31]) *Es sei* $\lambda = \frac{ck}{\sqrt{n}}$ *und* $\mu = \frac{cl}{\sqrt{n}}$. *Falls* $(\frac{1}{4} + \varepsilon)\log n < \lambda, \mu < \frac{n^{1/4}}{\log n}$, *dann gilt*

$$\begin{aligned} p(n,k,l) \sim p(n) \exp\Big(&- (\lambda+\mu) - \frac{\sqrt{n}}{c}(e^{-\lambda} + e^{-\mu}) - \frac{\sqrt{n}}{4c}(e^{-2\lambda} + e^{-2\mu}) \\ &- \frac{\sqrt{n}}{4c^3}((\lambda+1)^2 e^{-2\lambda} + (\mu+1)^2 e^{-2\mu}) + \frac{\sqrt{n}}{c} e^{-(\lambda+\mu)} \\ &- \frac{\sqrt{n}}{2c^3}(\lambda+1)(\mu+1)e^{-(\lambda+\mu)} + \frac{1}{2c^2}\lambda\mu(e^{-\lambda} + e^{-\mu})\Big). \end{aligned} \tag{3.2}$$

Zum Beweis benötigen wir als wichtiges Hilfsmittel die Euler-Maclaurinsche Summenformel:

Lemma 3.2 (Euler-Maclaurinsche Summenformel) Seien $a, b, m \in \mathbb{N}$ mit $a \leq b$ und f auf einer Obermenge von $[a, b]$ m-mal differenzierbar, dann gilt

$$\sum_{a \leq k < b} f(k) = \int_a^b f(x)\,dx + \Big(\sum_{j=1}^{m} \frac{B_j}{j!} f^{(j-1)}(x)\Big)\Big|_a^b + (-1)^{m+1} \int_a^b \frac{B_m(\{x\})}{m!} f^{(m)}(x)\,dx, \tag{3.3}$$

wobei B_j die Bernoulli-Zahlen und $B_j(x) = \sum_{i=0}^{j} \binom{j}{i} B_i x^{n-i}$ die Bernoulli-Polynome sind.

Zum Beweis siehe z.B. [9], S.506 ff. Weiters werden die folgenden Integrale gebraucht:

Lemma 3.3

$$\int_0^\infty -\log(1 - e^{-u})\,du = \frac{\pi^2}{6} = c^2, \tag{3.4}$$

$$\int_0^\infty \frac{u}{e^u - 1}\,du = \frac{\pi^2}{6} = c^2, \tag{3.5}$$

$$\int_0^\infty \frac{u^2 e^u}{(e^u - 1)^2}\,du = \frac{\pi^2}{3} = 2c^2. \tag{3.6}$$

Beweis: Man entwickelt die Integrale jeweils in Reihen und integriert gliedweise. Auf $[\varepsilon, \infty)$ ist dies jedenfalls möglich, da dort die Reihen durch geometrische Reihen abgeschätzt werden können und daher sogar gleichmäßige Konvergenz vorliegt. Anschließend lässt man ε gegen 0 streben.

$$\begin{aligned}
\int_0^\infty -\log(1 - e^{-u})\,du &= \int_0^\infty \sum_{k=1}^{\infty} \frac{1}{k} e^{-uk}\,du \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}, \\
\int_0^\infty \frac{u}{e^u - 1}\,du &= \int_0^\infty \sum_{k=1}^{\infty} u e^{-uk}\,du \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}, \\
\int_0^\infty \frac{u^2 e^u}{(e^u - 1)^2}\,du &= \int_0^\infty \sum_{k=1}^{\infty} k u^2 e^{-uk}\,du \\
&= \sum_{k=1}^{\infty} \frac{2}{k^2} = \frac{\pi^2}{3}.
\end{aligned}$$

□

Das folgende Lemma liefert Information über die Funktion $L(x) := \int_x^\infty \frac{t}{e^t - 1}\,dt$, die ebenfalls eine Rolle im Beweis von Satz 3.1 spielen wird.

Lemma 3.4

$$L(x) = (1+x)e^{-x} + \frac{1}{4}(1+2x)e^{-2x} + O(xe^{-3x}) \tag{3.7}$$

für $x \to \infty$.

Beweis: Wir entwickeln wieder in eine Potenzreihe und integrieren gliedweise:

$$\begin{aligned} L(x) &= \int_x^\infty \frac{t}{e^t - 1}\,dt \\ &= \int_x^\infty \sum_{k=1}^\infty te^{-kt}\,dt \\ &= \sum_{k=1}^\infty \frac{1}{k^2}(1+kx)e^{-kx} \\ &= (1+x)e^{-x} + \frac{1}{4}(1+2x)e^{-2x} + O(xe^{-3x}). \end{aligned}$$

□

Nun folgen noch zwei Lemmata über die Asymptotik gewisser Summenausdrücke:

Lemma 3.5 Es strebe β nach 0 und βk nach ∞. Dann gilt

1.
$$-\sum_{j=1}^{k-1} \log(1-e^{-\beta j}) = \frac{c^2}{\beta} - \frac{1}{\beta}L(\beta k) - k\log(1-e^{-\beta k}) + \frac{1}{2}\log\left(\frac{\beta}{2\pi}\right) + o(1), \tag{3.8}$$

2.
$$\sum_{j=1}^{k-1} \frac{j}{e^{\beta j}-1} = \frac{c^2}{\beta^2} - \frac{1}{\beta^2}L(\beta k) - \frac{1}{2\beta} + o(\beta^{-1}), \tag{3.9}$$

3.
$$\sum_{j=1}^{k-1} \frac{j^2 e^{\beta j}}{(1-e^{\beta j})^2} = \frac{2c^2}{\beta^3} + o(\beta^{-3}), \tag{3.10}$$

wobei $c = \frac{\pi}{\sqrt{6}}$.

Beweis: Alle drei folgen aus der Euler-Maclaurinschen Summenformel (Lemma 3.2), wobei die uneigentlichen Integrale aus Lemma 3.3 einfließen. Darüber hinaus wird die Produktregel verwendet, um das Integral $\int \log(1-e^{-\beta x})\,dx$ zu behandeln.

$$
\begin{aligned}
\sum_{j=1}^{k-1} \log \frac{j}{1-e^{-\beta j}} &= \sum_{j=0}^{k-1} \log \frac{j}{1-e^{-\beta j}} - \log \frac{1}{\beta} \\
&= \int_0^k \log \frac{x}{1-e^{-\beta x}}\, dx - \frac{1}{2}\Big(\log \frac{k}{1-e^{-\beta k}} - \log \frac{1}{\beta}\Big) + \frac{1}{12}\Big(\frac{1}{k} - \frac{\beta}{e^{\beta k}-1} - \frac{\beta}{2}\Big) \\
&\quad - \int_0^k \frac{B_2(\{x\})}{2}\Big(\frac{\beta^2 e^{\beta x}}{(e^{\beta x}-1)^2} - \frac{1}{x^2}\Big)\, dx - \log\frac{1}{\beta} \\
&= k(\log k - 1) - \int_0^k \log(1-e^{-\beta x}) - \frac{1}{2}(\log k - \log \beta + o(1)) + o(1) \\
&\quad + O\Big(\int_0^\infty \Big(\frac{\beta^2 e^{\beta x}}{(e^{\beta x}-1)^2} - \frac{1}{x^2}\Big)\, dx\Big) \\
&= k(\log k - 1) - k\log(1-e^{-\beta k}) + \int_0^k \frac{\beta x}{e^{\beta x}-1}\, dx - \frac{1}{2}(\log k - \log\beta + o(1)) \\
&\quad + o(1) + O\Big(\beta \int_0^\infty \Big(\frac{e^u}{(e^u-1)^2} - \frac{1}{u^2}\Big)\, du\Big) \\
&= k(\log k - 1) - k\log(1-e^{-\beta k}) + \frac{1}{\beta}\int_0^{\beta k} \frac{u}{e^u-1}\, du + \frac{1}{2}\log\Big(\frac{\beta}{k}\Big) + o(1) \\
&= k(\log k - 1) - k\log(1-e^{-\beta k}) + \frac{1}{\beta}(c^2 - L(\beta k)) + \frac{1}{2}\log\Big(\frac{\beta}{k}\Big) + o(1).
\end{aligned}
$$

Aus der Stirlingschen Formel erhält man

$$
\sum_{j=1}^{k-1} \log j = \log(k-1)! = \log\Big(k^k e^{-k}\sqrt{\frac{2\pi}{k}} e^{o(1)}\Big) = k\log k - k + \frac{1}{2}\log\Big(\frac{2\pi}{k}\Big) + o(1)
$$

und damit

$$
\begin{aligned}
-\sum_{j=1}^{k-1} \log(1-e^{-\beta j}) &= \sum_{j=1}^{k-1} \log \frac{j}{1-e^{-\beta j}} - \sum_{j=1}^{k-1} \log j \\
&= k(\log k - 1) - k\log(1-e^{-\beta k}) + \frac{1}{\beta}(c^2 - L(\beta k)) + \frac{1}{2}\log\Big(\frac{\beta}{k}\Big) \\
&\quad - k\log k + k - \frac{1}{2}\log\Big(\frac{2\pi}{k}\Big) + o(1) \\
&= \frac{1}{\beta}(c^2 - L(\beta k)) - k\log(1-e^{-\beta k}) + \frac{1}{2}\log\Big(\frac{\beta}{2\pi}\Big) + o(1).
\end{aligned}
$$

Somit ist 1. bewiesen. 2. folgt ebenso:

$$\begin{aligned}
\sum_{j=1}^{k-1} \frac{j}{e^{\beta j}-1} &= \sum_{j=0}^{k-1} \frac{j}{e^{\beta j}-1} - \frac{1}{\beta} \\
&= \int_0^k \frac{x}{e^{\beta x}-1}\,dx - \frac{1}{2}\Big(\frac{k}{e^{\beta k}-1} - \frac{1}{\beta}\Big) + \frac{1}{12}\Big(\frac{-1+e^{\beta k}(1-\beta k)}{(e^{\beta k}-1)^2} + \frac{1}{2}\Big) \\
&\quad - \int_0^k \frac{B_2(\{x\})}{2} \frac{\beta e^{\beta x}(2+\beta x+e^{\beta x}(\beta x-2))}{(e^{\beta x}-1)^3}\,dx - \frac{1}{\beta} \\
&= \frac{1}{\beta^2}\int_0^{\beta k} \frac{u}{e^u-1}\,du - \frac{1}{2\beta} + o(\beta^{-1}) + O\Big(\int_0^\infty \frac{\beta e^{\beta x}(2+\beta x+e^{\beta x}(\beta x-2))}{(e^{\beta x}-1)^3}\,dx\Big) \\
&= \frac{c^2}{\beta^2} - \frac{1}{\beta^2}L(\beta k) - \frac{1}{2\beta} + o(\beta^{-1}) + O\Big(\int_0^\infty \frac{e^u(2+u+e^u(u-2))}{(e^u-1)^3}\,du\Big) \\
&= \frac{c^2}{\beta^2} - \frac{1}{\beta^2}L(\beta k) - \frac{1}{2\beta} + o(\beta^{-1}).
\end{aligned}$$

Schließlich ergibt auch auch 3. in gleicher Weise:

$$\begin{aligned}
\sum_{j=1}^{k-1} \frac{j^2 e^{\beta j}}{(1-e^{\beta j})^2} &= \sum_{j=0}^{k-1} \frac{j^2 e^{\beta j}}{(1-e^{\beta j})^2} - \frac{1}{\beta^2} \\
&= \int_0^k \frac{x^2 e^{\beta x}}{(1-e^{\beta x})^2}\,dx - \frac{1}{2}\Big(\frac{k^2 e^{\beta k}}{(1-e^{\beta k})^2} - \frac{1}{\beta^2}\Big) \\
&\quad + \int_0^k B_1(\{x\}) \frac{x e^{\beta x}(2+\beta x+e^{\beta x}(\beta x-2))}{(1-e^{\beta x})^3}\,dx - \frac{1}{\beta^2} \\
&= \frac{1}{\beta^3}\int_0^{\beta k} \frac{u^2 e^u}{(1-e^u)^2}\,du + O(\beta^{-2}) + O\Big(\int_0^\infty \frac{x e^{\beta x}(2+\beta x+e^{\beta x}(\beta x-2))}{(1-e^{\beta x})^3}\,dx\Big) \\
&= \frac{2c^2}{\beta^3} + o(\beta^{-3}) + O\Big(\beta^{-2}\int_0^\infty \frac{u e^u(2+u+e^u(u-2))}{(1-e^u)^3}\,du\Big) \\
&= \frac{2c^2}{\beta^3} + o(\beta^{-3}).
\end{aligned}$$

□

Lemma 3.6 Es strebe β nach 0 und $\beta k, \beta l$ nach ∞. Dann gilt

1.

$$\begin{aligned}
-\sum_{j=l}^{k+l-2} \log(1-e^{-\beta j}) &= \frac{1}{\beta}\Big(L(\beta l) - L(\beta(k+l))\Big) \\
&\quad - (k+l)\log(1-e^{-\beta(k+l)}) + l\log(1-e^{-\beta l}) + o(1), \qquad (3.11)
\end{aligned}$$

2.
$$\sum_{j=l}^{k+l-2} \frac{j}{e^{\beta j}-1} = \frac{1}{\beta^2}\Big(L(\beta l) - L(\beta(k+l))\Big) + o(\beta^{-1}), \tag{3.12}$$

3.
$$\sum_{j=l}^{k+l-2} \frac{j^2 e^{\beta j}}{(1-e^{\beta j})^2} = o(\beta^{-3}). \tag{3.13}$$

Beweis: Wir verwenden wiederum die Euler-Maclaurinsche Summenformel und werten das Integral wie in 1. des vorigen Lemmas aus:

$$\begin{aligned}
-\sum_{j=l}^{k+l-2} \log(1-e^{-\beta j}) &= -\sum_{j=l}^{k+l-1} \log(1-e^{-\beta j}) + \log(1-e^{-\beta(k+l-1)}) \\
&= -\int_l^{k+l} \log(1-e^{-\beta x})\,dx - \frac{1}{2}\Big(-\log(1-e^{-\beta(k+l)}) + \log(1-e^{-\beta l})\Big) \\
&\quad + \int_l^{k+l} B_1(\{x\})\frac{\beta}{1-e^{\beta x}}\,dx + \log(1-e^{-\beta(k+l-1)}) \\
&= -(k+l)\log(1-e^{-\beta(k+l)}) + l\log(1-e^{-\beta l}) + \int_l^{k+l} \frac{\beta x}{e^{\beta x}-1}\,dx \\
&\quad + o(1) + O\Big(\int_l^{k+l} \frac{\beta}{1-e^{\beta x}}\,dx\Big) \\
&= -(k+l)\log(1-e^{-\beta(k+l)}) + l\log(1-e^{-\beta l}) + \frac{1}{\beta}\int_{\beta l}^{\beta(k+l)} \frac{u}{e^u-1}\,du \\
&\quad + o(1) + O\Big(\int_{\beta l}^{\beta(k+l)} \frac{1}{1-e^u}\,du\Big) \\
&= -(k+l)\log(1-e^{-\beta(k+l)}) + l\log(1-e^{-\beta l}) \\
&\quad + \frac{1}{\beta}\Big(L(\beta l) - L(\beta(k+l))\Big) + o(1).
\end{aligned}$$

Dasselbe gilt für 2.:

$$\begin{aligned}
\sum_{j=l}^{k+l-2} \frac{j}{e^{\beta j}-1} &= \sum_{j=l}^{k+l-1} \frac{j}{e^{\beta j}-1} - \frac{k+l-1}{e^{\beta(k+l-1)}-1} \\
&= \int_l^{k+l} \frac{x}{e^{\beta x}-1}\,dx - \frac{1}{2}\Big(\frac{k+l}{e^{\beta(k+l)}-1} - \frac{l}{e^{\beta l}-1}\Big) \\
&\quad + \int_l^{k+l} B_1(\{x\})\frac{-1+e^{\beta x}(1-\beta x)}{(e^{\beta x}-1)^2}\,dx - \frac{k+l-1}{e^{\beta(k+l-1)}-1}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\beta^2}\int_{\beta l}^{\beta(k+l)} \frac{u}{e^u-1}\,du + o(\beta^{-1})\\
&\quad + O\Big(\int_l^{k+l} \frac{-1+e^{\beta x}(1-\beta x)}{(e^{\beta x}-1)^2}\,dx\Big)\\
&= \frac{1}{\beta^2}\Big(L(\beta l) - L(\beta(k+l))\Big) + o(\beta^{-1})\\
&\quad + O\Big(\beta^{-1}\int_{\beta l}^{\beta(k+l)} \frac{-1+e^{u}(1-u)}{(e^{u}-1)^2}\,du\Big)\\
&= \frac{1}{\beta^2}\Big(L(\beta l) - L(\beta(k+l))\Big) + o(\beta^{-1}).
\end{aligned}
$$

Für 3. bemerken wir nur

$$
\begin{aligned}
\sum_{j=l}^{k+l-2} \frac{j^2 e^{\beta j}}{(1-e^{\beta j})^2} &= O\Big(\int_l^{k+l-2} \frac{x^2 e^{\beta x}}{(1-e^{\beta x})^2}\,dx\Big)\\
&= O\Big(\beta^{-3}\int_{\beta l}^{\beta(k+l-2)} \frac{u^2 e^{u}}{(1-e^{u})^2}\,du\Big)\\
&= o(\beta^{-3}).
\end{aligned}
$$

Beweis von Satz 3.1: Wir verwenden die erzeugende Funktion aus Lemma 1.5. Wir wissen also, dass

$$\sum_{n=0}^{\infty} p(n,k,l)x^n = x^{k+l-1}\prod_{m=1}^{k-1}\frac{1-x^{l+m-1}}{(1-x^m)}. \tag{3.14}$$

Daher erhalten wir nach der Cauchy'schen Integralformel

$$p(n,k,l) = \frac{1}{2\pi i}\int_{\mathcal{C}} z^{k+l-n-2}\exp\Big(-\sum_{j=1}^{k-1}\log(1-z^j) + \sum_{j=l}^{l+k-2}\log(1-z^j)\Big)\,dz, \tag{3.15}$$

wobei $\mathcal{C}$ eine geeignete Kurve um 0 ist. Dieses Integral schätzen wir mit Hilfe der Sattelpunktsmethode ab: wir schreiben

$$z^{k+l-n-1}\exp\Big(-\sum_{j=1}^{k-1}\log(1-z^j) + \sum_{j=l}^{k+l-2}\log(1-z^j)\Big) = \exp g(z), \tag{3.16}$$

wobei $g(z) = -\sum_{j=1}^{k-1} \log(1-z^j) + \sum_{j=l}^{k+l-2} \log(1-z^j) + (k+l-n-1)\log z$. Die Ableitungen von g sind

$$\begin{aligned}
g'(z) &= \sum_{j=1}^{k-1} \frac{jz^{j-1}}{1-z^j} - \sum_{j=l}^{k+l-2} \frac{jz^{j-1}}{1-z^j} - \frac{n-k-l+1}{z} \\
&= \sum_{j=1}^{k-1} \frac{j}{z}\frac{1}{z^{-j}-1} - \sum_{j=l}^{k+l-2} \frac{j}{z}\frac{1}{z^{-j}-1} - \frac{n-k-l+1}{z}, \\
g''(z) &= \sum_{j=1}^{k-1} \frac{j(1+(j-1)z^{-j})}{z^2(z^{-j}-1)^2} - \sum_{j=l}^{k+l-2} \frac{j(1+(j-1)z^{-j})}{z^2(z^{-j}-1)^2} + \frac{n-k-l+1}{z^2} \\
&= \sum_{j=1}^{k-1} \Big(\frac{j}{z^2(1-z^{-j})} + \frac{j^2z^{-j}}{z^2(1-z^{-j})^2}\Big) - \sum_{j=l}^{k+l-2} \Big(\frac{j}{z^2(1-z^{-j})} + \frac{j^2z^{-j}}{z^2(1-z^{-j})^2}\Big) + \frac{n-k-l+1}{z^2}, \\
g'''(z) &= \sum_{j=1}^{k-1} \frac{jz^{-2j}(j^2(1+z^j) + 3j(-1+z^j) + 2(-1+z^j)^2)}{z^3(z^{-j}-1)^3} \\
&\quad - \sum_{j=l}^{k+l-2} \frac{jz^{-2j}(j^2(1+z^j) + 3j(-1+z^j) + 2(-1+z^j)^2)}{z^3(z^{-j}-1)^3} - \frac{2(n-k-l+1)}{z^3}.
\end{aligned}$$

Sei nun $z_0 = e^{-\beta}$ der Sattelpunkt, der somit die Bedingung

$$g'(e^{-\beta}) = 0 \iff \sum_{j=1}^{k-1} \frac{j}{e^{\beta j}-1} - \sum_{j=l}^{k+l-2} \frac{j}{e^{\beta j}-1} = n-k-l+1$$

erfüllen muss. Es ergibt sich

$$\int_0^\infty \frac{x}{e^{\beta x}-1}\,dx = \frac{1}{\beta^2}\int_0^\infty \frac{u}{e^u-1}\,du = \frac{c^2}{\beta^2} \sim n$$

und daher $\beta \sim \frac{c}{\sqrt{n}}$ bzw. genauer nach Lemma 3.5 und Lemma 3.6 (aufgrund der Voraussetzungen über k und l sind die Bedingungen zur Anwendung dieser Lemmata gewährleistet)

$$\frac{c^2}{\beta^2} - \frac{1}{\beta^2}L(\beta k) - \frac{1}{2\beta} + o(\beta^{-1}) - \frac{1}{\beta^2}\Big(L(\beta l) - L(\beta(k+l))\Big) + o(\beta^{-1}) = n-k-l+1 \quad (3.17)$$

oder

$$\frac{1}{\beta^2}\Big(c^2 - L(\beta k) - L(\beta l) + L(\beta(k+l))\Big) = n-k-l+\frac{1}{2\beta} + o(\sqrt{n})$$

und somit

$$\frac{1}{\beta^2} = \frac{n}{c^2}\Big(1 - \frac{k+l}{n} + \frac{1}{2\beta n}\Big)\Big(1 - \frac{1}{c^2}(L(\beta k) + L(\beta l)) + \frac{1}{c^2}L(\beta(k+l))\Big)^{-1} + o(\sqrt{n}). \quad (3.18)$$

Nun verwendet man $\frac{1}{2\beta n} \sim \frac{1}{2c\sqrt{n}}$ und die aus den Voraussetzungen über k und l folgenden Tatsachen

$$k, l = o(n^{3/4}) \ , \ L(\beta k), L(\beta l) = O(n^{-1/4+\delta}) \ , \ L(\beta(k+l)) = O(n^{-1/2+\delta}),$$

wobei $\delta > 0$ beliebig ist. Entwickelt man (3.18) nach n und setzt $\lambda = \frac{ck}{\sqrt{n}}$ und $\mu = \frac{cl}{\sqrt{n}}$ ein, so ergibt sich damit

$$\begin{aligned}\frac{1}{\beta} = \frac{\sqrt{n}}{c} - \frac{1}{2c^2}(\lambda+\mu) + \frac{1}{4c^2} + \frac{\sqrt{n}}{2c^3}(L(\beta k) + L(\beta l) - L(\beta(k+l)))&\\ + \frac{3\sqrt{n}}{8c^5}(L(\beta k) + L(\beta l))^2 - \frac{1}{4c^4}(\lambda+\mu)(L(\beta k) + L(\beta l)) + o(1)&\end{aligned} \tag{3.19}$$

oder durch Umformung

$$\beta = \frac{c}{\sqrt{n}} - \frac{1}{2c\sqrt{n}}(L(\beta k) + L(\beta l)) + \frac{\lambda+\mu}{2n} + O(n^{-1}\log^2 n). \tag{3.20}$$

Unter Verwendung von Lemma 3.5 und Lemma 3.6 erhält man für den Sattelpunkt

$$\begin{aligned}g(e^{-\beta}) &= -\sum_{j=1}^{k-1}\log(1-e^{-\beta j}) + \sum_{j=l}^{k+l-2}\log(1-e^{-\beta j}) + (n-k-l+1)\beta\\ &= \frac{c^2}{\beta} - \frac{1}{\beta}L(\beta k) - k\log(1-e^{-\beta k}) + \frac{1}{2}\log\left(\frac{\beta}{2\pi}\right) + o(1)\\ &\quad - \frac{1}{\beta}\Big(L(\beta l) - L(\beta(k+l))\Big) + (k+l)\log(1-e^{-\beta(k+l)}) - l\log(1-e^{-\beta l}) + o(1)\\ &\quad + (n-k-l+1)\beta.\end{aligned}$$

Wir multiplizieren Gleichung (3.17) mit β, um für $(n-k-l+1)\beta$ einen anderen Ausdruck zu erhalten, und setzen diesen dann ein:

$$\begin{aligned}g(e^{-\beta}) = \frac{2c^2}{\beta} - \frac{2}{\beta}(L(\beta k) + L(\beta l) - L(\beta(k+l))) - k\log(1-e^{-\beta k})&\\ - l\log(1-e^{-\beta l}) + (k+l)\log(1-e^{-\beta(k+l)}) - \frac{1}{2} + \frac{1}{2}\log\left(\frac{\beta}{2\pi}\right) + o(1).&\end{aligned} \tag{3.21}$$

Aus den Voraussetzungen über k und l, Lemma 3.4 sowie dem Ausdruck (3.20) erhalten wir der Reihe nach

$$\begin{aligned}k\beta &= \lambda - \frac{\lambda}{2c^2}\Big((\lambda+1)e^{-\lambda} + (\mu+1)e^{-\mu}\Big) + \frac{\lambda}{2c\sqrt{n}}(\lambda+\mu) + O(\lambda n^{-1/2}\log^2 n),\\ e^{-k\beta} &= e^{-\lambda}e^{\lambda-k\beta} = e^{-\lambda}\Big(1 + \lambda - k\beta + O((\lambda-k\beta)^2)\Big)\\ &= e^{-\lambda} + \frac{\lambda e^{-\lambda}}{2c^2}\Big((\lambda+1)e^{-\lambda} + (\mu+1)e^{-\mu}\Big) - \frac{\lambda e^{-\lambda}}{2c\sqrt{n}}(\lambda+\mu) + O(\lambda^2 e^{-\lambda}n^{-1/2}\log^2 n),\\ L(\beta k) &= (\lambda+1)e^{-\lambda} + \frac{\lambda^2 e^{-\lambda}}{2c^2}\Big((\lambda+1)e^{-\lambda} + (\mu+1)e^{-\mu}\Big) - \frac{\lambda^2\mu e^{-\lambda}}{2c\sqrt{n}} + \frac{1+2\lambda}{4}e^{-2\lambda} + o(n^{-1/2}).\end{aligned}$$

Analog erhält man auch $L(\beta l)$ und daher

$$\begin{aligned}L(\beta k) + L(\beta l) &= \Big((\lambda+1)e^{-\lambda} + (\mu+1)e^{-\mu}\Big)\Big(1 + \frac{1}{2c^2}(\lambda^2 e^{-\lambda} + \mu^2 e^{-\mu})\Big)\\ &\quad - \frac{\lambda\mu}{2c\sqrt{n}}(\lambda e^{-\lambda} + \mu e^{-\mu}) + \frac{1}{4}(e^{-2\lambda} + e^{-2\mu}) + \frac{1}{2}(\lambda e^{-2\lambda} + \mu e^{-2\mu}) + o(n^{-1/2}),\end{aligned}$$

und schließlich noch

$$L(\beta(k+l)) = (\lambda + \mu + 1)e^{-\lambda-\mu} + o(n^{-1/2}).$$

Setzt man all dies sowie $\frac{1}{\beta}$ aus Gleichung (3.19) ein, und verwendet man

$$\begin{aligned}-k\log(1 - e^{-\beta k}) &= ke^{-\beta k} + \frac{k}{2}e^{-2\beta k} + o(1),\\ -l\log(1 - e^{-\beta k}) &= le^{-\beta l} + \frac{l}{2}e^{-2\beta l} + o(1),\\ -(k+l)\log(1 - e^{-\beta(k+l)}) &= (k+l)e^{-\beta(k+l)} + o(1),\end{aligned}$$

so erhält man aus (3.21)

$$\begin{aligned}g(e^{-\beta}) &= 2c\sqrt{n} - (\lambda+\mu) - \frac{\sqrt{n}}{c}(e^{-\lambda} + e^{-\mu}) - \frac{\sqrt{n}}{4c}(e^{-2\lambda} + e^{-2\mu})\\ &\quad - \frac{\sqrt{n}}{4c^3}((\lambda+1)^2 e^{-2\lambda} + (\mu+1)^2 e^{-2\mu}) + \frac{\sqrt{n}}{c}e^{-(\lambda+\mu)}\\ &\quad - \frac{\sqrt{n}}{2c^3}(\lambda+1)(\mu+1)e^{-(\lambda+\mu)} + \frac{1}{2c^2}\lambda\mu(e^{-\lambda} + e^{-\mu}) + \frac{1}{2}\log\frac{\beta}{2\pi} + o(1). \qquad (3.22)\end{aligned}$$

Mit der zweiten Ableitung wird nun ebenso verfahren:

$$\begin{aligned}g''(e^{-\beta}) &= \sum_{j=1}^{k-1} \frac{je^{2\beta}}{(1 - e^{\beta j})} + \sum_{j=1}^{k-1} \frac{j^2 e^{(2+j)\beta}}{(1 - e^{\beta j})^2}\\ &\quad - \sum_{j=l}^{k+l-2} \frac{je^{2\beta}}{(1 - e^{\beta j})} - \sum_{j=l}^{k+l-2} \frac{j^2 e^{(2+j)\beta}}{(1 - e^{\beta j})^2} + (n-k-l+1)e^{2\beta}\\ &= e^{2\beta}\Big(O(\beta^{-2}) + \frac{2c^2}{\beta^3} + o(\beta^{-3}) + o(\beta^{-2}) + o(\beta^{-3}) + O(n)\Big)\\ &= (1 + O(\beta))\Big(\frac{2c^2}{\beta^3} + o(\beta^{-3}) + O(n)\Big)\\ &= \frac{2}{c}n^{3/2} + o(n^{3/2}).\end{aligned}$$

Schließlich ergibt sich für die dritte Ableitung

$$\begin{aligned} g'''(e^{-\beta}) = &\sum_{j=1}^{k-1} \frac{je^{2\beta j+3\beta}(j^2(1+e^{-\beta j})+3j(-1+e^{-\beta j})+2(-1+e^{-\beta j})^2)}{(e^{\beta j}-1)^3} \\ &- \sum_{j=l}^{k+l-2} \frac{je^{2\beta j+3\beta}(j^2(1+e^{-\beta j})+3j(-1+e^{-\beta j})+2(-1+e^{-\beta j})^2)}{(e^{\beta j}-1)^3} - 2(n-k-l+1)e^{3\beta}. \end{aligned}$$

Der Zähler kann leicht abgeschätzt werden, indem man $1 \geq e^{-\beta j} \geq 0$ verwendet:

$$-3j \leq j^2(1+e^{-\beta j})+3j(-1+e^{-\beta j})+2(-1+e^{-\beta j})^2 \leq 2j^2+2 \leq 4j^2.$$

Es folgt

$$\begin{aligned} |g'''(e^{-\beta})| &\leq 2\sum_{j=1}^{\infty} \frac{4j^3e^{2\beta j+3\beta}}{(e^{\beta j}-1)^3} + O(n) \\ &\sim 8e^{3\beta} \int_0^{\infty} \frac{x^3e^{2\beta x}}{(e^{\beta j}-1)^3}\,dx + O(n) \\ &= \frac{8e^{3\beta}}{\beta^4} \int_0^{\infty} \frac{u^3e^{2u}}{(e^u-1)^3}\,du + O(n) \\ &= O(\beta^{-4}) = O(n^2) \end{aligned}$$

und daher

$$g(z) = g(z_0) + \frac{g''(z_0)}{2}(z-z_0)^2 + O(n^2(z-z_0)^3).$$

Wir wählen nun $\mathcal{C} = \{z_0e^{it} \mid -\pi < t \leq \pi\}$. Dann ergibt sich

$$\begin{aligned} p(n,k,l) &= \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{\exp g(z)}{z}\,dz \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp g(z_0e^{it})\,dt. \end{aligned}$$

Für $-n^{-5/7} \leq t \leq n^{-5/7}$ gilt $z - z_0 = z_0(it + O(t^2)) = (1+O(n^{-1/2}))(it+O(n^{-10/7})) = it + O(n^{-17/14})$ und damit

$$g(z) = g(z_0) - \frac{g''(z_0)}{2}t^2 + O(n^{-1/7}). \tag{3.23}$$

Daher ist ein Teil des Integrals

$$\frac{1}{2\pi}\int_{-n^{-5/7}}^{n^{-5/7}} \exp g(z_0e^{it})\,dt = \frac{1}{2\pi}(\exp g(z_0))(1+O(n^{-1/7}))\int_{-n^{-5/7}}^{n^{-5/7}} e^{-\frac{g''(z_0)}{2}t^2}\,dt,$$

wobei

$$\begin{aligned}\int_{n^{-5/7}}^{\infty} e^{-\frac{g''(z_0)}{2}t^2}\,dt &\le \int_{n^{-5/7}}^{\infty} e^{-\frac{g''(z_0)}{2}n^{-5/7}t}\,dt\\ &= \int_{n^{-5/7}}^{\infty} \exp\Big(-\Big(\frac{n^{11/14}}{c}+o(n^{11/14})\Big)t\Big)\,dt \sim cn^{-11/14}e^{-\frac{1}{c}n^{1/14}+o(n^{11/14})},\end{aligned}$$

und daher

$$\begin{aligned}\int_{-n^{-5/7}}^{n^{-5/7}} e^{-\frac{g''(z_0)}{2}t^2}\,dt &= \int_{-\infty}^{\infty} e^{-\frac{g''(z_0)}{2}t^2}\,dt - 2\int_{n^{-5/7}}^{\infty} e^{-\frac{g''(z_0)}{2}t^2}\,dt\\ &= \sqrt{\frac{2\pi}{g''(z_0)}}+o(1)\\ &\sim \sqrt{\frac{\pi c}{n^{3/2}}}.\end{aligned}$$

Für diesen Teil des Integrals erhalten wir somit

$$\frac{1}{2\pi}\int_{-n^{-5/7}}^{n^{-5/7}} \exp g(z_0e^{it})\,dt \sim \sqrt{\frac{c}{4\pi n^{3/2}}}\exp g(z_0) = \frac{1}{2\cdot 6^{1/4}n^{3/4}}\exp g(z_0). \tag{3.24}$$

Der Rest des Integrals ist vernachlässigbar: dies wird gezeigt, indem wir zunächst beweisen, dass $\left|\frac{e^{g(z_0)}}{e^{g(z_0e^{it})}}\right|$ für $n^{-5/7} \le |t| \le \pi$ schneller als jede Potenz von n wächst. Es gilt

$$\begin{aligned}\left|\frac{e^{g(z_0)}}{e^{g(z_0e^{it})}}\right|^2 &= \Big|\prod_{j=1}^{k-1}\frac{1-z_0^je^{itj}}{1-z_0^j}\Big|^2\cdot\Big|\prod_{j=l}^{k+l-2}\frac{1-z_0^j}{1-z_0^je^{itj}}\Big|^2\\ &= \prod_{j=1}^{k-1}\frac{|(1-z_0^j\cos(tj))-iz_0^j\sin(tj)|^2}{|1-z_0^j|^2}\cdot\prod_{j=l}^{k+l-2}\frac{|1-z_0^j|^2}{|(1-z_0^j\cos(tj))-iz_0^j\sin(tj)|^2}\\ &= \prod_{j=1}^{k-1}\frac{(1-z_0^j\cos(tj))^2+(z_0^j\sin(tj))^2}{(1-z_0^j)^2}\cdot\prod_{j=l}^{k+l-2}\frac{(1-z_0^j)^2}{(1-z_0^j\cos(tj))^2+(z_0^j\sin(tj))^2}\\ &= \prod_{j=1}^{k-1}\frac{1-2z_0^j\cos(tj)+z_0^{2j}}{(1-z_0^j)^2}\cdot\prod_{j=l}^{k+l-2}\frac{(1-z_0^j)^2}{1-2z_0^j\cos(tj)+z_0^{2j}}\\ &= \prod_{j=1}^{k-1}\Big(1+\frac{2(1-\cos(tj))z_0^j}{(1-z_0^j)^2}\Big)\cdot\prod_{j=l}^{k+l-2}\Big(1+\frac{2(1-\cos(tj))z_0^j}{(1-z_0^j)^2}\Big)^{-1}\\ &\ge \prod_{\sqrt{n}\le j\le 2\sqrt{n}}\Big(1+2z_0^j(1-\cos(tj))\Big)\cdot\prod_{j=l}^{k+l-2}\Big(1+M(1-\cos(tj))z_0^j\Big)^{-1}\\ &\ge \prod_{\sqrt{n}\le j\le 2\sqrt{n}}\Big(1+m(1-\cos(tj))\Big)\cdot\prod_{j=l}^{k+l-2}\Big(1+M(1-\cos(tj))z_0^j\Big)^{-1},\end{aligned}$$

wobei m, M von n unabhängige Konstanten sind, denn einerseits ist $z_0^{2\sqrt{n}} = e^{-2c+o(1)} = O(1)$, andererseits gilt $z_0^l \ll e^{-\frac{1}{4}\log n} = n^{-1/4}$, da $l > \frac{1}{4c}\sqrt{n}\log n$.

Für $n^{-5/7} \leq |t| \leq \frac{\pi}{2\sqrt{n}}$ gilt $n^{-3/14} \leq |tj| \leq \pi$ für $\sqrt{n} \leq j \leq 2\sqrt{n}$. Wir verwenden nun die Ungleichung $\frac{u^2}{2} \geq 1 - \cos u \geq \frac{2u^2}{\pi^2}$, die für $0 \leq u \leq \pi$ gilt. Aus ihr folgt

$$\prod_{\sqrt{n} \leq j \leq 2\sqrt{n}} \Big(1 + m(1 - \cos(tj))\Big) \geq \prod_{\sqrt{n} \leq j \leq 2\sqrt{n}} \Big(1 + C_1 t^2 j^2\Big) \geq (1 + C_1 t^2 n)^{\sqrt{n}+O(1)}$$

für eine gewisse Konstante C_1. Es gibt weiters eine Konstante K, für die die Ungleichung $1 + u \geq e^{Ku}$ erfüllt ist, falls $0 \leq u \leq C_1\pi^2/4$. Da $C_1 t^2 n \leq C_1 \pi^2/4$ gilt, folgt damit

$$\prod_{\sqrt{n} \leq j \leq 2\sqrt{n}} \Big(1 + m(1 - \cos(tj))\Big) \geq e^{C_1 K t^2 n^{3/2} + O(t^2 n)} \gg e^{C_2 t^2 n^{3/2}}$$

für eine Konstante C_2. Andererseits ist

$$\prod_{j=l}^{k+l-2} \Big(1 + M(1 - \cos(tj)) z_0^j\Big)^{-1} \geq \prod_{j=l}^{k+l-2} \Big(1 + C_3 t^2 j^2 z_0^j\Big)^{-1},$$

und da $j^2 z_0^j = j^2 e^{-\beta j}$ für $j \geq \frac{2}{\beta}$ monoton fällt, folgt

$$j^2 z_0^j \leq \begin{cases} l^2 z_0^l \leq \frac{1}{16c^2} n \log^2 n (n^{-1/4} + o(n^{-1/4})) = O(n^{3/4}\log^2 n) & \text{falls } j < \frac{1}{2c}\sqrt{n}\log n, \\ \frac{1}{4c^2} n \log^2 n (n^{-1/2} + o(n^{-1/2})) = O(\sqrt{n}\log^2 n) & \text{andernfalls,} \end{cases}$$

und damit schlussendlich

$$\begin{aligned} \prod_{j=l}^{k+l-2} \Big(1 + M(1 - \cos(tj)) z_0^j\Big)^{-1} &\geq \Big(1 + t^2 O(n^{3/4}\log^2 n)\Big)^{-\frac{1}{4c}\sqrt{n}\log n} \Big(1 + t^2 O(\sqrt{n}\log^2 n)\Big)^{-k} \\ &\geq \exp\Big(- t^2 O(n^{3/4}\log^2 n) \cdot \frac{1}{4c}\sqrt{n}\log n\Big) \\ &\quad \cdot \exp\Big(- t^2 O(\sqrt{n}\log^2 n) \cdot \frac{1}{c} n^{3/4}/\log n\Big) \\ &= \exp(-t^2 O(n^{5/4}\log^3 n)). \end{aligned}$$

D.h., im Fall $n^{-5/7} \leq |t| \leq \frac{\pi}{2\sqrt{n}}$ haben wir

$$\left| \frac{e^{g(z_0)}}{e^{g(z_0 e^{it})}} \right| \geq e^{\frac{C_2}{2} t^2 (n^{3/2} + O(n^{5/4}\log^3 n))} \geq e^{\frac{C_2}{2}(n^{1/14} + o(1))}, \tag{3.25}$$

was tatsächlich schneller als jede Potenz von n wächst.

Für $\frac{\pi}{2\sqrt{n}} \leq |t| \leq \pi$ bestimmen wir

$$\Big| \{\sqrt{n} \leq j \leq 2\sqrt{n} : |tj - 2k\pi| \leq n^{-1/12} \text{ für ein } k \in \mathbb{Z}\} \Big|.$$

Für k kommen nur höchstens $\frac{\sqrt{n}|t|}{2\pi}+1$ verschiedene Werte in Frage, da $|t|\sqrt{n} \le |tj| \le 2|t|\sqrt{n}$ gilt. Zu jedem $k \in \mathbb{N}$ wiederum können höchstens $\frac{2n^{-1/12}}{|t|} + 1$ verschiedene Werte von j gehören. Daher ist die Mächtigkeit obiger Menge höchtens

$$\Big(\frac{2n^{-1/12}}{|t|} + 1\Big)\Big(\frac{\sqrt{n}|t|}{2\pi} + 1\Big) = \frac{\sqrt{n}|t|}{2\pi} + O(n^{5/12}) \le \frac{\sqrt{n}}{2} + O(n^{5/12}).$$

Es ergibt sich für solches t damit

$$\begin{aligned}\prod_{\sqrt{n}\le j\le 2\sqrt{n}} \Big(1 + m(1-\cos(tj))\Big) &\ge (1 + m(1-\cos n^{-1/12}))^{\sqrt{n}/2+O(n^{5/12})} \\ &= (1 + mn^{-1/6}/2 + O(n^{-1/3}))^{\sqrt{n}/2+O(n^{1/3})} \\ &= \exp(mn^{1/3}/4 + O(n^{1/6})).\end{aligned}$$

Andererseits schätzen wir $1 - \cos(tj)$ durch 2 ab und erhalten wie im anderen Fall

$$\begin{aligned}\prod_{j=l}^{k+l-2} \Big(1 + M(1-\cos(tj))z_0^j\Big)^{-1} &\ge \Big(1 + O(n^{-1/4})\Big)^{-\frac{1}{4c}\sqrt{n}\log n}\Big(1 + O(n^{-1/2})\Big)^{-k} \\ &\ge \exp\Big(- O(-n^{-1/4}) \cdot \frac{1}{4c}\sqrt{n}\log n\Big) \\ &\quad \cdot \exp\Big(- O(n^{-1/2}) \cdot \frac{1}{c} n^{3/4}/\log n\Big) \\ &= \exp(-O(n^{1/4}\log n)),\end{aligned}$$

womit sich im Fall $\frac{\pi}{2\sqrt{n}} \le |t| \le \pi$

$$\Big|\frac{e^{g(z_0)}}{e^{g(z_0e^{it})}}\Big| \ge e^{mn^{1/3}/4+O(n^{1/4}\log n)} \tag{3.26}$$

ergibt. Schlussendlich folgern wir

$$\begin{aligned}\frac{1}{2\pi}\Big|\int_{n^{-5/7}}^{\pi} \exp g(z_0e^{it})\,dt\Big| &\le \frac{1}{2\pi}\int_{n^{-5/7}}^{\pi} |\exp g(z_0e^{it})|\,dt \\ &= \frac{1}{2\pi}\int_{n^{-5/7}}^{\pi} (\exp g(z_0))\Big|\frac{e^{g(z_0e^{it})}}{e^{g(z_0)}}\Big|\,dt \\ &\le \frac{1}{2}(\exp g(z_0))\exp(-C_2n^{1/14}/2 + o(1))\end{aligned}$$

und analog auch für das Integral zwischen $-\pi$ und $-n^{-5/7}$. Daher erhalten wir

$$p(n,k,l) = \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{\exp g(z)}{z}\,dz \sim \frac{1}{2\cdot 6^{1/4}n^{3/4}}\exp g(z_0). \tag{3.27}$$

Setzen wir nun noch (3.22) für $g(z_0)$ ein und verwenden $p(n) \sim \frac{e^{2c\sqrt{n}}}{4\sqrt{3}n}$ aus Satz 2.15, so ergibt sich zusammenfassend

$$\begin{aligned} p(n,k,l) \sim p(n) \exp\Big(&- (\lambda+\mu) - \frac{\sqrt{n}}{c}(e^{-\lambda} + e^{-\mu}) - \frac{\sqrt{n}}{4c}(e^{-2\lambda} + e^{-2\mu}) \\ &- \frac{\sqrt{n}}{4c^3}((\lambda+1)^2 e^{-2\lambda} + (\mu+1)^2 e^{-2\mu}) + \frac{\sqrt{n}}{c} e^{-(\lambda+\mu)} \\ &- \frac{\sqrt{n}}{2c^3}(\lambda+1)(\mu+1)e^{-(\lambda+\mu)} + \frac{1}{2c^2}\lambda\mu(e^{-\lambda} + e^{-\mu})\Big). \end{aligned} \tag{3.28}$$

□

Als unmittelbare Folgerung aus dem soeben bewiesenen Satz erhalten wir

Korollar 3.7 (Szekeres [31]) Es seien $\lambda = \frac{ck}{\sqrt{n}}$ und $\mu = \frac{cl}{\sqrt{n}}$ derart, dass $(\frac{1}{4} + \varepsilon)\log n < \lambda, \mu < \log n$ gilt. Dann folgt

$$p(n,k,l) \sim p(n) \exp\Big(- (\lambda+\mu) - \frac{\sqrt{n}}{c}(e^{-\lambda} + e^{-\mu})\Big). \tag{3.29}$$

Beweis: Zum Beweis genügt die Beobachtung, dass alle übrigen Terme aus Satz 3.1 in diesem Fall nur einen Beitrag $o(1)$ liefern und daher vernachlässigt werden können. □

Bemerkung: Die übrigen Terme, die in diesem Fall wegfallen, zeigen die Korrelation der k- und l-Verteilungen, falls entweder λ oder μ nahe bei $\frac{1}{4}\log n$ liegt.

Bemerkung: Dieses Korollar bedeutet, dass die Verteilungen der Partitionen bezüglich Länge und Maximum im Bereich ihrer Mediane voneinander im Wesentlichen unabhängig sind und die Dichte $\exp\Big(- \frac{cx}{\sqrt{n}} - \frac{\sqrt{n}}{c} e^{-\frac{cx}{\sqrt{n}}}\Big)$ haben. Die Abbildung, in der Länge und Maximum von 1000 Zufallspartitionen von 1000 eingetragen sind, illustriert die Verteilung. Verwenden wir nun

$$\begin{aligned} \sum_{\frac{\sqrt{n}}{4c}\log n < k < \frac{\sqrt{n}}{c}\log n} \exp\Big(- \lambda - \frac{\sqrt{n}}{c} e^{-\lambda}\Big) &\sim \int_{\frac{\sqrt{n}}{4c}\log n}^{\frac{\sqrt{n}}{c}\log n} \exp\Big(- \lambda - \frac{\sqrt{n}}{c} e^{-\lambda}\Big)\, dk \\ &= \frac{\sqrt{n}}{c} \int_{\frac{1}{4}\log n}^{\log n} \exp\Big(- \lambda - \frac{\sqrt{n}}{c} e^{-\lambda}\Big)\, d\lambda \\ &\overset{t=\frac{\sqrt{n}}{c}e^{-\lambda}}{=} \int_{\frac{n^{-1/2}}{c}}^{\frac{n^{1/4}}{c}} e^{-t}\, dt \sim 1, \end{aligned}$$

so folgt unmittelbar, dass

$$\sum_{\frac{\sqrt{n}}{4c}\log n < k,l < \frac{\sqrt{n}}{c}\log n} p(n,k,l) \sim p(n).$$

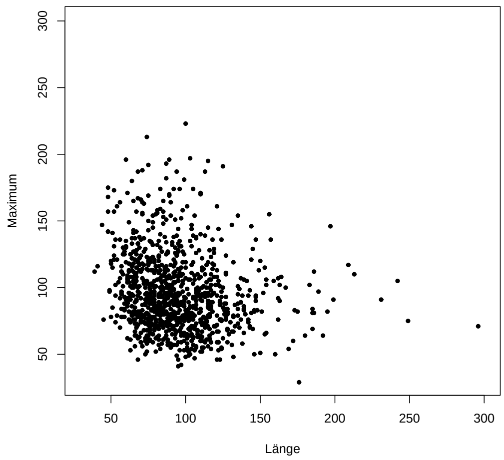

Abbildung 3.1: Länge und Maximum von Zufallspartitionen

D.h., fast alle Partitionen liegen im Bereich $\frac{\sqrt{n}}{4c}\log n < k, l < \frac{\sqrt{n}}{c}\log n$ und genauer noch (wie man auf dieselbe Art zeigt und wie bereits in der Einleitung erwähnt wurde) im Bereich $\frac{1}{2c}\sqrt{n}\log n \pm \omega(n)\sqrt{n}$ für eine beliebige Funktion $\omega(n)$, die mit n nach ∞ strebt. Es ergibt sich damit sofort

Satz 3.8 (Szekeres [31]) *Falls $\lambda = \frac{ck}{\sqrt{n}}$ die Ungleichung $(\frac{1}{4}+\varepsilon)\log n < \lambda < \log n$ erfüllt, dann gilt*

$$p(n,k) \sim p(n)\exp\Big(-\lambda - \frac{\sqrt{n}}{c}e^{-\lambda}\Big). \tag{3.30}$$

Beweis: Wir schreiben

$$p(n,k) = \sum_{l=1}^{n} p(n,k,l)$$

und erhalten aus obigen Bemerkungen

$$\begin{aligned} p(n,k) &\sim \sum_{\frac{\sqrt{n}}{4c}\log n < l < \frac{\sqrt{n}}{c}\log n} p(n,k,l) \\ &\sim p(n)\exp\Big(-\lambda - \frac{\sqrt{n}}{c}e^{-\lambda}\Big). \end{aligned}$$

□

Die Abbildung zeigt das Histogramm der Längen von 10000 Zufallspartitionen von 1000. Zusätzlich ist die theoretische Dichte eingezeichnet. Angemerkt sei schließlich noch, dass Szekeres auch analoge Fragen für die Verteilung von Länge und Maximum von Partitionen mit lauter ungleichen Teilen behandelt hat (s. [30]).

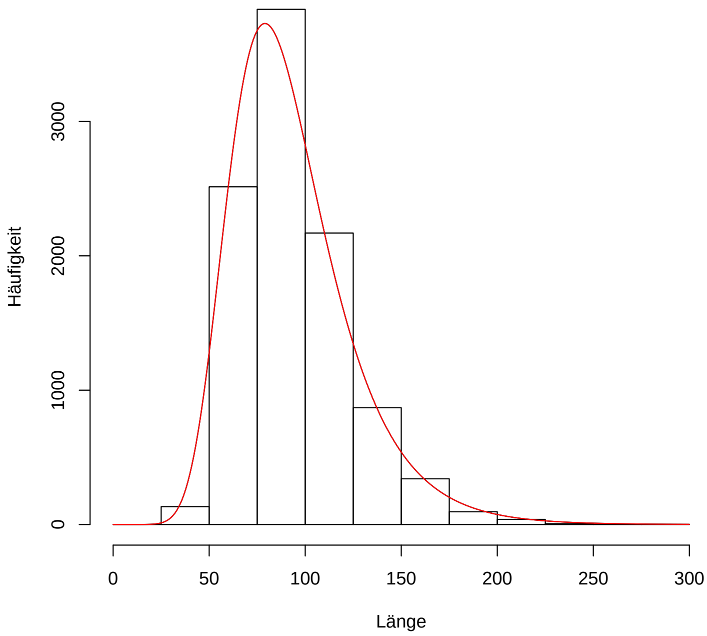

Abbildung 3.2: Länge von Zufallspartitionen

Kapitel 4

Partitionen und Bäume

4.1 Sternartige Bäume und Fibonaccizahlen

Partitionen besitzen auch gewisse Anwendungen in der Graphentheorie: es ist möglich, jeder Partition von n in wohldefinierter Weise einen Baum mit n Kanten zuzuordnen. Bis auf geringe Ausnahmen ist diese Zuordnung injektiv. Wir beginnen mit der Definition eines sternartigen Baumes:

Definition 4.1 Ein Baum heißt *sternartig*, wenn er Durchmesser ≤ 4 hat.

Definition 4.2 Es sei $(c_1, \ldots, c_d)$ eine Partition von n. Der sternartige Baum, der dieser Partition zugeordnet wird, hat die folgende Form:

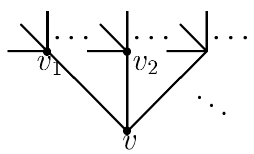

Abbildung 4.1: Der sternartige Baum $S(c_1, \ldots, c_d)$

wobei $v_1, \ldots, v_d$ Grad $c_1, \ldots, c_d$ haben. Offensichtlich hat dieser Baum genau n Kanten (zu jedem v_i gehören genau c_i Kanten, und es gibt keine weiteren). Dieser Baum sei im Folgenden mit $S(c_1, \ldots, c_d)$ bezeichnet.

Proposition 4.1 Jeder sternartige Baum T hat die Form $S(c_1, \ldots, c_d)$ für eine gewisse Partition von n; Diese Partition ist auch eindeutig, falls der Baum Durchmesser 4 hat. Andernfalls gibt es genau zwei verschiedene solche Partitionen, ausgenommen den Baum $S(\frac{n+1}{2}, 1, \ldots, 1)$, falls n ungerade ist.

Beweis: Sei zunächst der Durchmesser von T gleich 4. Wir wählen einen beliebigen Durchmesser v_0, v_1, v_2, v_3, v_4. Dann muss der Abstand von v_2 zu jedem anderen Knoten w kleiner oder gleich 2 sein (ansonsten gäbe es einen Weg der Länge ≥ 5 von v_0 nach w oder von v_4 nach w). Daher sind alle Teilbäume, die an v_2 hängen, Sterne, die mit v_2 an ihrem Mittelpunkt verbunden sind. Es folgt, dass der Baum die gewünschte Form hat, wobei v_2 das eindeutige Zentrum ist (von jedem anderen Punkt aus gibt es Wege der Länge ≥ 3).

Ist der Durchmesser dagegen 3, dann sehen wir analog, dass T ein Doppelstern sein muss. Dabei gibt es zwei Möglichkeiten für die Wahl des Zentrums, die auf die beiden möglichen Repräsentationen $S(k, \underbrace{1, \ldots, 1}_{l-1})$ und $S(l, \underbrace{1, \ldots, 1}_{k-1})$ führen, wobei k und l die Grade der beiden möglichen Zentren sind. Diese beiden Repräsentationen fallen genau dann zusammen, wenn $k = l = \frac{n+1}{2}$.

Schlussendlich bleibt der Stern S_n, der als einziger Durchmesser 2 hat. Er hat die beiden Darstellungen $S(1, \ldots, 1)$ and $S(n)$. Damit ist die Behauptung bewiesen. □

Satz 4.2 *Es gibt $p(n) - \lfloor \frac{n}{2} \rfloor$ nichtisomorphe sternartige Bäume mit n Kanten, wobei $p(n)$ wie in den vergangenen Kapiteln die Zahl der Partitionen von n bezeichnet.*

Beweis: Nach Proposition 4.1 gehört jede Partition zu genau einem sternartigen Baum und vice versa. Die einzigen Ausnahmen sind die Partitionen der Form $(k, 1, \ldots, 1)$. Hier repräsentieren zwei Partitionen genau dann den gleichen Baum, wenn sie von der Form $(k, 1, \ldots, 1)$ und $(l, 1, \ldots, 1)$ $(k, l \geq 1,\ k \neq l)$ mit $k + l = n + 1$ sind. Es gibt genau $\lfloor \frac{n}{2} \rfloor$ Paare (k, l) mit $1 \leq k < l$ und $k + l = n + 1$. Damit ist bereits alles bewiesen. □

Sternartige Graphen sind von Bedeutung in der Behandlung der sogenannten Fibonaccizahlen von Graphen. Die Fibonaccizahl $F(G)$ eines Graphen G ist definiert durch die Anzahl der unabhängigen Knotenteilmengen von G. Eine Menge von Knoten heißt dabei unabhängig, wenn keine zwei Knoten durch eine Kante verbunden sind. Der Name erklärt sich dadurch, dass sich für einfache Wege genau die Fibonaccizahlen ergeben.

Fibonaccizahlen von Graphen finden Anwendung in der Chemie, wo die Fibonaccizahl eines zu einem Molekül gehörigen Graphen Merrifield-Simmons-Index oder σ-Index genannt wird. Merrifield und Simmons [20] konnten Zusammenhänge zwischen dem σ-Index und gewissen physikalischen Eigenschaften von Molekülen, wie etwa dem Siedepunkt, aufzeigen. Das inverse Problem, bei dem zu einem gegebenen σ-Index ein Graph mit diesem Index gesucht wird, kann zur Erstellung von Datenbanken zur Entwicklung von Medikamenten dienen (s. [18]).

Für einen sternartigen Baum $S(c_1, \ldots, c_d)$ erhält man eine einfache Formel für die Fibonaccizahl.

Lemma 4.3 Es sei T ein Baum und $v \in V(T)$. $T \setminus \{v\}$ zerfalle in Teilbäume $T_1, \ldots, T_k$; weiters sei $\{v_i\} = N(v) \cap T_i$, wobei $N(v)$ die Nachbarn von v bezeichnet. $T_i \setminus \{v_i\}$ zerfalle seinerseits in Teilbäume $T_{i1}, \ldots, T_{il_i}$. Dann gilt

$$F(T) = \prod_{i=1}^{k} F(T_i) + \prod_{i=1}^{k} \prod_{j=1}^{l_i} F(T_{ij}). \tag{4.1}$$

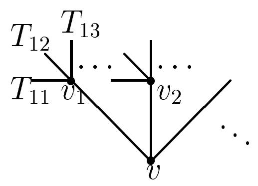

Abbildung 4.2: Zu Lemma 4.3

Beweis: Es gibt genau $\prod_{i=1}^{k} F(T_i)$ unabhängige Knotenteilmengen von T, die v nicht enthalten, und $\prod_{i=1}^{k} \prod_{j=1}^{l_i} F(T_{ij})$ unabhängige Knotenteilmengen von T, die v enthalten (und damit $v_1, \ldots, v_k$ nicht enthalten). Daraus ergibt sich das Lemma unmittelbar. □

Lemma 4.4

$$F(S(c_1, \ldots, c_d)) = \prod_{i=1}^{d} (2^{c_i - 1} + 1) + 2^{n-d} \tag{4.2}$$

Beweis: Unter Verwendung der Tatsache, dass die Fibonaccizahl des Sterns $S_{c_i - 1}$ gleich $2^{c_i-1} + 1$ (2^{c_i-1} verschiedene Teilmengen der äußeren Knoten sowie die eine Menge, die nur aus dem Mittelpunkt besteht) ist, folgt diese Formel trivialerweise aus Lemma 4.3. □

Der Baum maximaler Fibonaccizahl ist, wie bereits Prodinger und Tichy bewiesen (s. [22]), der Sterngraph $S_n \simeq S(1, \ldots, 1)$. Wie sich herausstellt, sind die Bäume größtmöglicher Fibonaccizahl bis zu einem gewissen Punkt allesamt sternartig (s. [16]). Daraus ergibt sich etwa die natürliche Frage, welche Fibonaccizahl ein sternartiger Graph im Durchschnitt hat. Dieser Frage wird in der späteren Folge nachgegangen, wobei sich wiederum eine interessante Anwendung der "circle method" auftut. Zunächst wenden wir uns einer speziellen Subklasse von Bäumen zu, deren Interpretation auch vom Standpunkt der Theorie der Partitionen her interessant ist.

4.2 Eine spezielle Klasse sternartiger Bäume

Wir betrachten die Menge aller sternartigen Bäume mit der Eigenschaft, dass der Mittelpunkt gleichzeitig auch der Knoten maximalen Grades ist. Wir bezeichnen die Mächtigkeit der Menge mit $t(n)$, wobei n wiederum die Zahl der Kanten darstellt.

Man sieht leicht ein, dass diese Klasse von Bäumen genau jenen Partitionen entspricht, für die die Länge (die dem Grad des Mittelpunktes entspricht) größer oder gleich dem maximalen Element der Partition ist (die dem Maximum der Grade aller Nachbarn des Mittelpunktes entspricht). Wir können $t(n)$ daher auch in folgender Art und Weise anschreiben:

$$t(n) = |\{c_d \leq \ldots \leq c_1 \leq d : \ c_1 + \ldots + c_d = n, \ c_i \geq 1\}| \tag{4.3}$$

Nun interessieren wir uns für das asymptotische Verhalten von $t(n)$ für $n \to \infty$. Unter Zuhilfenahme der Ergebnisse aus dem vorangegangenen Kapitel ist die Bestimmung dieser Asymptotik leicht möglich.

Bei der Konjugation einer Partition werden, wie bereits bemerkt, die Rollen von Länge und Maximum vertauscht. Mit diesem Argument sehen wir, dass $t(n)$ gleichzeitig die Zahl der Partitionen mit Maximum $\geq$ Länge ist. Es folgt unmittelbar, dass

$$p(n) = 2t(n) - r(n), \tag{4.4}$$

wobei $r(n)$ die Zahl jener Partitionen ist, für die Länge und Maximum gleich sind. $r(n)$ kann auch in folgender Form geschrieben werden:

$$r(n) = \sum_d p(n, d, d). \tag{4.5}$$

Ist $c = (c_1, \ldots, c_d)$ eine Partition, so bezeichnet man $c_1 - d$ (also die Differenz aus Maximum und Länge) als *Rang* der Partition. $t(n)$ entspricht daher der Zahl aller Partitionen mit Rang ≥ 0 (oder ≤ 0), $r(n)$ ist die Zahl aller Partitionen mit Rang 0.

Satz 4.5 *Es gilt die folgende Asymptotik:*

$$t(n) = p(n)\Big(\frac{1}{2} + \frac{\pi}{8\sqrt{6n}} + o(n^{-1/2})\Big). \tag{4.6}$$

Beweis: Es sei während dieses Beweises $c = \frac{\pi}{\sqrt{6}}$ wie im vorigen Kapitel. Wir teilen die Summe für $r(n)$ in drei Teile:

- Zunächst sei $D_1 = \frac{1}{3}\frac{\sqrt{n}}{c}\log n$ und $S_1 = \sum_{d<D_1} p(n, d, d)$. Aus Satz 3.8 wissen wir, dass
$$S_1 \leq P(n, D_1) = p(n + D_1, D_1) \sim p(n + D_1) \exp\Big(-\lambda - \frac{\sqrt{n}}{c}e^{-\lambda}\Big),$$
wobei $\lambda = \frac{c}{\sqrt{n+D_1}}D_1 = \frac{1}{3}\sqrt{\frac{n}{n+D_1}}\log n = \frac{1}{3}\log n(1 + O(n^{-1/2}\log n))$. Es folgt
$$\lambda + \frac{\sqrt{n}}{c}e^{-\lambda} \sim \frac{1}{3}\log n + \frac{\sqrt{n}}{c}n^{-1/3}(1 + o(1)) \gg n^{1/6}.$$

Damit wissen wir, dass $\exp(-\lambda - \frac{\sqrt{n}}{c}e^{-\lambda})$ schneller fällt als jede Potenz von n. Zusätzlich gilt nach Satz 2.15

$$\begin{aligned}\frac{p(n+D_1)}{p(n)} &\sim \exp\Big(2c(\sqrt{n+D_1}-\sqrt{n})\Big)\\ &= \exp\Big(2c\frac{D_1}{\sqrt{n+D_1}+\sqrt{n}}\Big)\\ &\le \exp\Big(\frac{1}{3}\log n\Big) = n^{1/3}\end{aligned}$$

und damit $S_1 \ll p(n)n^{-\alpha}$ für beliebiges $\alpha > 0$.

- Sei $D_2 = n^{2/3}$ und $S_2 = \sum_{d>D_2} p(n,d,d)$. Offensichtlich gilt $p(n,d,d) \le p(n-2d+1)$, wie ein simples kombinatorisches Argument zeigt: wenn man im Ferrer-Diagramm einer Partition mit Länge und Maximum $= d$ die erste Zeile und Spalte entfernt, bleibt eine Partition von $n-2d+1$ übrig. Nach Satz 2.15 gilt für jedes $d > D_2$

$$\begin{aligned}\frac{p(n-2d+1)}{p(n)} &\sim \exp\Big(2c(\sqrt{n-2d+1}-\sqrt{n})\Big)\\ &= \exp\Big(2c\frac{1-2d}{\sqrt{n-2d+1}+\sqrt{n}}\Big)\\ &\le \exp\Big(2c\frac{-2n^{2/3}}{\sqrt{n}}\Big)\\ &= \exp(-4cn^{1/6}).\end{aligned}$$

Es folgt, dass $S_2 \le \frac{n}{2}p(n)\exp(-4cn^{1/6}) \ll p(n)n^{-\alpha}$ für beliebiges $\alpha > 0$.

- Für $D_1 \le d \le D_2$ erhalten wir aus Satz 3.1

$$\begin{aligned}p(n,d,d) \sim{}& p(n)\exp\Big(-2\lambda - 2\frac{\sqrt{n}}{c}e^{-\lambda} - \frac{\sqrt{n}}{2c}e^{-2\lambda} - \frac{\sqrt{n}}{2c^3}(\lambda+1)^2e^{-2\lambda}\\ &+ \frac{\sqrt{n}}{c}e^{-2\lambda} - \frac{\sqrt{n}}{2c^3}(\lambda+1)^2e^{-2\lambda} + \frac{1}{c^2}\lambda^2e^{-\lambda}\Big),\end{aligned}$$

wobei $\lambda = \frac{cd}{\sqrt{n}}$. Nun sei $S_2 = \sum_{D_1\le d\le D_2} p(n,d,d)$. Dann gilt

$$\begin{aligned}\frac{S_3}{p(n)} &\sim \int_{D_1}^{D_2} \exp\Big(-2\lambda - \frac{2\sqrt{n}}{c}e^{-\lambda} + \frac{1}{c^2}\lambda^2e^{-\lambda} + \frac{\sqrt{n}}{2c}e^{-2\lambda} - \frac{\sqrt{n}}{c^3}(\lambda+1)^2e^{-2\lambda}\Big)\,dd\\ &= \frac{\sqrt{n}}{c}\int_{\frac{cD_1}{\sqrt{n}}}^{\frac{cD_2}{\sqrt{n}}} \exp\Big(-2\lambda - \frac{2\sqrt{n}}{c}e^{-\lambda} + \frac{1}{c^2}\lambda^2e^{-\lambda} + \frac{\sqrt{n}}{2c}e^{-2\lambda} - \frac{\sqrt{n}}{c^3}(\lambda+1)^2e^{-2\lambda}\Big)\,d\lambda.\end{aligned}$$

$\frac{c}{\sqrt{n}}D_1 = \frac{1}{3}\log n$ und $\frac{c}{\sqrt{n}}D_2 = cn^{1/6}$; innerhalb dieser Grenzen gilt $\lambda^2e^{-\lambda} \ll (\log n)^2n^{-1/3}$,

$\sqrt{n}e^{-2\lambda} \ll n^{-1/6}$ und $\sqrt{n}(\lambda+1)^2 e^{-2\lambda} \ll (\log n)^2 n^{-1/6}$. Daraus folgt

$$\begin{aligned}
\frac{S_3}{p(n)} &\sim \frac{\sqrt{n}}{c}\int_{\frac{cD_1}{\sqrt{n}}}^{\frac{cD_2}{\sqrt{n}}} \exp\Big(-2\lambda - \frac{2\sqrt{n}}{c}e^{-\lambda} + o(1)\Big)\,d\lambda \\
&= \frac{\sqrt{n}}{c}(1+o(1))\int_{\frac{1}{3}\log n}^{cn^{1/6}} \exp\Big(-2\lambda - \frac{2\sqrt{n}}{c}e^{-\lambda}\Big)\,d\lambda \\
&\overset{t=\frac{2\sqrt{n}}{c}e^{-\lambda}}{=} \frac{\sqrt{n}}{c}(1+o(1))\int_{\frac{2\sqrt{n}}{c}e^{-cn^{1/6}}}^{\frac{2}{c}n^{1/6}} \Big(\frac{ct}{2\sqrt{n}}\Big)^2 e^{-t}\frac{1}{t}\,dt \\
&= \frac{c}{4\sqrt{n}}(1+o(1))\int_{\frac{2\sqrt{n}}{c}e^{-cn^{1/6}}}^{\frac{2}{c}n^{1/6}} te^{-t}\,dt \\
&= \frac{c}{4\sqrt{n}}(1+o(1))\Big(-(t+1)e^{-t}\Big|_{\frac{2\sqrt{n}}{c}e^{-cn^{1/6}}}^{\frac{2}{c}n^{1/6}}\Big) \\
&= \frac{c}{4\sqrt{n}}(1+o(1))(1+o(1)) \\
&= \frac{c}{4\sqrt{n}} + o(n^{-1/2}).
\end{aligned}$$

Zusammenfassend sehen wir, dass $r(n) = S_1 + S_2 + S_3 = p(n)\Big(\frac{c}{4\sqrt{n}} + o(n^{-1/2})\Big)$ und daher

$$t(n) = p(n)\Big(\frac{1}{2} + \frac{\pi}{8\sqrt{6n}} + o(n^{-1/2})\Big),$$

womit die Behauptung bewiesen ist. □

$t(n)$ und $r(n)$ sind auch in "Sloanes On-Line Encyclopedia of Integer Sequences" [27] verzeichnet: $t(n)$ ist Folge Nr. A064174, $r(n)$ ist Folge Nr. A047993. Die Tabelle zeigt einige Werte von $t(n)$ und $r(n)$.

4.3 Die durchschnittliche Fibonaccizahl sternartiger Bäume

Nun wenden wir uns der bereits erwähnten Fragestellung zu, welche durchschnittliche Fibonaccizahl ein sternartiger Baum mit n Kanten hat. Die Berechnung erfolgt im Wesentlichen analog zum Beweis von Satz 3.1. Zur Bestimmung der darin auftretenden Konstanten werden wieder einige uneigentliche Integrale benötigt, die im folgenden Lemma kurz behandelt werden.

n	$t(n)$	$r(n)$
1	1	1
2	1	0
3	2	1
4	3	1
5	4	1
6	6	1
7	9	3
8	12	2
9	17	4
10	23	4
20	336	45
50	106864	9502
100	98399016	6228740
200	2032145505447	91291981506

Tabelle 4.1: Einige Werte von $t(n)$ und $r(n)$.

Lemma 4.6

$$\int_0^\infty -\log(1-e^{-u}/2)\,du = \frac{\pi^2}{12} - \frac{(\log 2)^2}{2}\,, \tag{4.7}$$

$$\int_0^\infty \frac{u}{2e^u-1}\,du = \frac{\pi^2}{12} - \frac{(\log 2)^2}{2}\,, \tag{4.8}$$

$$\int_0^\infty \frac{u^2e^u}{(2e^u-1)^2}\,du = \frac{\pi^2}{12} - \frac{(\log 2)^2}{2}\,. \tag{4.9}$$

Beweis: Man entwickelt die Integrale jeweils in Reihen und integriert gliedweise. Dies ist hier in allen Fällen jedenfalls möglich, da die Reihen gleichmäßig konvergieren (sie lassen sich in trivialer Weise durch geometrische Reihen abschätzen):

$$\begin{aligned}
\int_0^\infty -\log(1-e^{-u}/2)\,du &= \int_0^\infty \sum_{k=1}^\infty \frac{1}{k2^k}e^{-uk}\,du \\
&= \sum_{k=1}^\infty \frac{1}{k^2 2^k}, \\
\int_0^\infty \frac{u}{2e^u-1}\,du &= \int_0^\infty \sum_{k=1}^\infty 2^{-k}ue^{-uk}\,du \\
&= \sum_{k=1}^\infty \frac{1}{k^2 2^k},
\end{aligned}$$

$$\begin{aligned}\int_0^\infty \frac{u^2e^u}{(2e^u-1)^2}\,du &= \int_0^\infty \sum_{k=1}^\infty k2^{-k-1}u^2e^{-uk}\,du \\ &= \sum_{k=1}^\infty \frac{1}{k^2 2^k}.\end{aligned}$$

Also haben tatsächlich all diese Integrale denselben Wert, der durch die unendliche Summe $\sum_{k=1}^\infty \frac{1}{k^2 2^k}$ gegeben ist. Um den genauen Wert zu bestimmen, betrachten wir allgemein die Funktion $\mathrm{Li}_2(x) = \sum_{k=1}^\infty \frac{1}{k^2}x^k$. Sie wird auch als Dilogarithmus bezeichnet (siehe z.B. [17], wo allgemeiner Polylogarithmen von der Form $\sum_{k=1}^\infty \frac{1}{k^r}x^k$ behandelt werden). Ihr Wert an der Stelle $\frac{1}{2}$ kann explizit bestimmt werden, da Li_2 eine Funktionalgleichung erfüllt.
Man beachte dazu zunächst, dass $\frac{d}{dx}\,\mathrm{Li}_2(x) = -\frac{\log(1-x)}{x}$ gilt. Daher gilt auch

$$\frac{d}{dx}\Big(\,\mathrm{Li}_2(x)+\mathrm{Li}_2(1-x)\Big) = -\frac{\log(1-x)}{x} + \frac{\log x}{1-x} = \frac{d}{dx}\Big(-\log x\log(1-x)\Big).$$

Daher ist $\mathrm{Li}_2(x)+\mathrm{Li}_2(1-x)+\log x\log(1-x) =: C$ konstant. Lässt man x gegen 0 gehen, so erhält man

$$\begin{aligned}C &= \mathrm{Li}_2(0)+\mathrm{Li}_2(1)+\lim_{x\to 0}\log x\log(1-x) \\ &= 0+\sum_{k=1}^\infty \frac{1}{k^2} + \lim_{x\to 0}(-x\log x + O(x^2\log x)) \\ &= \zeta(2) = \frac{\pi^2}{6}.\end{aligned}$$

Durch Einsetzen von $x=\frac{1}{2}$ ergibt sich schließlich

$$2\,\mathrm{Li}_2\left(\frac{1}{2}\right) = \frac{\pi^2}{6} - \log^2\frac{1}{2}$$

oder

$$\mathrm{Li}_2\left(\frac{1}{2}\right) = \frac{\pi^2}{12} - \frac{(\log 2)^2}{2},$$

was zu beweisen war. □

Damit lassen sich nun folgende asymptotische Entwicklungen für Summenausdrücke bestimmen:

Lemma 4.7 Es strebe β nach 0. Dann gilt

1.
$$-\sum_{j=1}^\infty \log(1-e^{-\beta j}/2) = \frac{b^2}{\beta} - \frac{1}{2}\log 2 + O(\beta), \tag{4.10}$$

2.
$$\sum_{j=1}^{\infty} \frac{j}{2e^{\beta j}-1} = \frac{b^2}{\beta^2} + O(1), \tag{4.11}$$

3.
$$\sum_{j=1}^{\infty} \frac{j^2 e^{\beta j}}{(1-2e^{\beta j})^2} = \frac{b^2}{\beta^3} + O(\beta^{-2}), \tag{4.12}$$

wobei $b^2 = \frac{\pi^2}{12} - \frac{(\log 2)^2}{2}$.

Beweis: Alle drei folgen direkt aus der Euler-Maclaurinschen Summenformel (Lemma 3.2):

$$\begin{aligned}
-\sum_{j=1}^{N-1} \log(1-e^{-\beta j}/2) &= -\sum_{j=0}^{N-1} \log(1-e^{-\beta j}/2) - \log 2 \\
&= -\int_0^N \log(1-e^{-\beta x}/2)\,dx - \frac{1}{2}\Big(-\log(1-e^{-\beta N}/2) + \log\frac{1}{2}\Big) \\
&\quad + \frac{1}{12}\Big(-\frac{\beta}{2e^{\beta N}-1} + \beta\Big) + \int_0^N \frac{B_2(\{x\})}{2}\frac{2\beta^2 e^{\beta x}}{(1-2e^{\beta x})^2}\,dx - \log 2
\end{aligned}$$

und daher für $N \to \infty$

$$\begin{aligned}
-\sum_{j=1}^{\infty} \log(1-e^{-\beta j}/2) &= -\int_0^\infty \log(1-e^{-\beta x}/2)\,dx - \frac{1}{2}\log 2 + \frac{\beta}{12} \\
&\quad + \int_0^\infty \frac{B_2(\{x\})\beta^2 e^{\beta x}}{(1-2e^{\beta x})^2}\,dx \\
&= -\frac{1}{\beta}\int_0^\infty \log(1-e^{-u}/2)\,du - \frac{1}{2}\log 2 + \frac{\beta}{12} \\
&\quad + \beta\int_0^\infty \frac{B_2(\{u/\beta\})e^u}{(1-2e^u)^2}\,du \\
&= \frac{b^2}{\beta} - \frac{1}{2}\log 2 + O(\beta).
\end{aligned}$$

Damit ist 1. bewiesen. 2. folgt ebenso:

$$\begin{aligned}
\sum_{j=1}^{N-1} \frac{j}{2e^{\beta j}-1} &= \sum_{j=0}^{N-1} \frac{j}{2e^{\beta j}-1} \\
&= \int_0^N \frac{x}{2e^{\beta x}-1}\,dx - \frac{1}{2}\Big(\frac{N}{2e^{\beta N}-1}\Big) + \frac{1}{12}\Big(\frac{-1+2e^{\beta N}(1-\beta N)}{(2e^{\beta N}-1)^2} - 1\Big) \\
&\quad + \int_0^N \frac{B_2(\{x\})}{2}\frac{2\beta e^{\beta x}(2+\beta x+2e^{\beta x}(\beta x-2))}{(2e^{\beta x}-1)^3}\,dx
\end{aligned}$$

und daher für $N \to \infty$

$$\begin{aligned}\sum_{j=1}^{\infty} \frac{j}{2e^{\beta j}-1} &= \int_0^\infty \frac{x}{2e^{\beta x}-1}\,dx - \frac{1}{12} + \int_0^\infty \frac{B_2(\{x\})\beta e^{\beta x}(2+\beta x+2e^{\beta x}(\beta x-2))}{(2e^{\beta x}-1)^3}\,dx \\ &= \frac{1}{\beta^2}\int_0^\infty \frac{u}{2e^u-1}\,du - \frac{1}{12} - \int_0^\infty \frac{B_2(\{u/\beta\})e^u(2+u+2e^u(u-2))}{(2e^u-1)^3}\,du \\ &= \frac{b^2}{\beta^2} + O(1).\end{aligned}$$

Schließlich ergibt sich auch 3. in gleicher Weise:

$$\begin{aligned}\sum_{j=1}^{N-1} \frac{j^2e^{\beta j}}{(1-2e^{\beta j})^2} &= \sum_{j=0}^{N-1} \frac{j^2e^{\beta j}}{(1-2e^{\beta j})^2} \\ &= \int_0^N \frac{x^2e^{\beta x}}{(1-2e^{\beta x})^2}\,dx - \frac{1}{2}\Big(\frac{N^2e^{\beta N}}{(1-2e^{\beta N})^2}\Big) \\ &\quad + \int_0^N B_1(\{x\})\frac{xe^{\beta x}(2+\beta x+2e^{\beta x}(\beta x-2))}{(1-2e^{\beta x})^3}\,dx\end{aligned}$$

und daher für $N \to \infty$

$$\begin{aligned}\sum_{j=1}^{\infty} \frac{j^2e^{\beta j}}{(1-2e^{\beta j})^2} &= \int_0^\infty \frac{x^2e^{\beta x}}{(1-2e^{\beta x})^2}\,dx + \int_0^\infty B_1(\{x\})\frac{xe^{\beta x}(2+\beta x+2e^{\beta x}(\beta x-2))}{(1-2e^{\beta x})^3}\,dx \\ &= \frac{1}{\beta^3}\int_0^\infty \frac{u^2e^u}{(1-2e^u)^2}\,du + \frac{1}{\beta^2}\int_0^\infty B_1(\{u/\beta\})\frac{ue^u(2+u+2e^u(u-2))}{(1-2e^u)^3}\,du \\ &= \frac{b^2}{\beta^3} + O(\beta^{-2}).\end{aligned}$$

□

Aus diesen Resultaten lassen sich wiederum auch die Entwicklungen für leicht modifizierte Summenausdrücke bestimmen:

Lemma 4.8

1.

$$-\sum_{j=1}^{\infty} \log\Big(1-\Big(\frac{1}{2}+2^{-j}\Big)e^{-\beta j}\Big) = \frac{b^2}{\beta} - \log\beta - \frac{3}{2}\log 2 - \log\rho + O(\beta), \tag{4.13}$$

2.

$$\sum_{j=1}^{\infty} \frac{j}{(1/2+2^{-j})^{-1}e^{\beta j}-1} = \frac{b^2}{\beta^2} + \frac{1}{\beta} + O(1), \tag{4.14}$$

3.
$$\sum_{j=1}^{\infty} \frac{j^2(1/2+2^{-j})^{-1}e^{\beta j}}{(1-(1/2+2^{-j})^{-1}e^{\beta j})^2} = \frac{2b^2}{\beta^3} + O(\beta^{-2}), \tag{4.15}$$

wobei $\rho = \prod_{j=1}^{\infty}(1-2^{-j}) = 0.288788095\ldots$.

Beweis: Ad 1.:

$$\sum_{j=1}^{\infty} \log\left(1-\left(\frac{1}{2}+2^{-j}\right)e^{-\beta j}\right) = \sum_{j=1}^{\infty} \log\left(1-\frac{1}{2}e^{-\beta j}\right) + \sum_{j=1}^{\infty} \log\left(\frac{1-\left(\frac{1}{2}+2^{-j}\right)e^{-\beta j}}{1-\frac{1}{2}e^{-\beta j}}\right).$$

In der zweiten Summe haben wir für $j=1$

$$\begin{aligned}
\log\left(\frac{1-e^{-\beta}}{1-e^{-\beta}/2}\right) &= \log\left(\frac{\beta-\beta^2/2+O(\beta^3)}{1/2+\beta/2+O(\beta^2)}\right)\\
&= \log((\beta-\beta^2/2+O(\beta^3))(2-2\beta+O(\beta^2)))\\
&= \log(2\beta(1+O(\beta)))\\
&= \log(2\beta)+O(\beta)
\end{aligned}$$

sowie für alle anderen j die folgenden Abschätzungen:

- $$\begin{aligned}
&\frac{1-\left(\frac{1}{2}+2^{-j}\right)e^{-\beta j}}{1-\frac{1}{2}e^{-\beta j}} \geq 1-2^{1-j}\\
\iff & 1-\left(\frac{1}{2}+2^{-j}\right)e^{-\beta j} \geq 1-2^{1-j}-\frac{1}{2}e^{-\beta j}+2^{-j}e^{-\beta j}\\
\iff & 2^{1-j} \geq 2^{1-j}e^{-\beta j},
\end{aligned}$$

- $$\begin{aligned}
&\frac{1-\left(\frac{1}{2}+2^{-j}\right)e^{-\beta j}}{1-\frac{1}{2}e^{-\beta j}} \leq (1-2^{1-j})(1+8j2^{-j}\beta)\\
\iff & 1-\left(\frac{1}{2}+2^{-j}\right)e^{-\beta j} \leq 1-2^{1-j}-\frac{1}{2}e^{-\beta j}+2^{-j}e^{-\beta j}\\
&\qquad\qquad +8j2^{-j}\beta\left(1-\frac{1}{2}e^{-\beta j}\right)(1-2^{1-j})\\
\iff & 2^{1-j}\underbrace{(1-e^{-\beta j})}_{\leq \beta j} \leq 8j2^{-j}\beta\underbrace{(1-e^{-\beta j}/2)}_{\geq 1/2}\underbrace{(1-2^{1-j})}_{\geq 1/2}.
\end{aligned}$$

Damit folgt aber

$$\sum_{j=1}^{\infty} \log \left(\frac{1 - \left(\frac{1}{2} + 2^{-j}\right) e^{-\beta j}}{1 - \frac{1}{2} e^{-\beta j}} \right) = \log(2\beta) + O(\beta) + \sum_{j=2}^{\infty} \log(1 - 2^{1-j}) + R(\beta),$$

wobei

$$0 \leq R(\beta) \leq \sum_{j=2}^{\infty} \log(1 + 8j2^{-j}\beta) \leq \sum_{j=2}^{\infty} 8j2^{-j}\beta = 12\beta$$

und damit

$$\sum_{j=1}^{\infty} \log \left(\frac{1 - \left(\frac{1}{2} + 2^{-j}\right) e^{-\beta j}}{1 - \frac{1}{2} e^{-\beta j}} \right) = \log(2\beta) + \log \rho + O(\beta).$$

Daraus ergibt sich nun das behauptete Resultat:

$$\begin{aligned} -\sum_{j=1}^{\infty} \log \left(1 - \left(\frac{1}{2} + 2^{-j} \right) e^{-\beta j} \right) &= \frac{b^2}{\beta} - \frac{1}{2} \log 2 + O(\beta) - \log(2\beta) - \log \rho + O(\beta) \\ &= \frac{b^2}{\beta} - \frac{3}{2} \log 2 - \log \beta - \log \rho + O(\beta). \end{aligned}$$

In 2. und 3. schreiben wir als Abkürzung $a(j) := (1/2 + 2^{-j})^{-1}$ und verwenden, dass

$$(2 - 2^{2-j}) a(j)^{-1} = (2 - 2^{2-j})(1/2 + 2^{-j}) = 1 - 2^{2-2j} \leq 1$$

gilt und somit $0 \leq 2 - a(j) \leq 2^{2-j}$. Weiters ist $2 \geq a(j) \geq 4/3$ für $j \geq 2$. Nun zu 2.:

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{j}{2e^{\beta j} - 1} - \sum_{j=1}^{\infty} \frac{j}{a(j)e^{\beta j} - 1} &= \frac{1}{2e^{\beta} - 1} - \frac{1}{e^{\beta} - 1} + \sum_{j=2}^{\infty} \left(\frac{j}{2e^{\beta j} - 1} - \frac{j}{a(j)e^{\beta j} - 1} \right) \\ &= -\frac{1}{\beta} + O(1) + \sum_{j=2}^{\infty} \frac{j(a(j) - 2)e^{\beta j}}{(2e^{\beta j} - 1)(a(j)e^{\beta j} - 1)}. \end{aligned}$$

Nun ist aber

$$\left| \sum_{j=2}^{\infty} \frac{j(a(j) - 2)e^{\beta j}}{(2e^{\beta j} - 1)(a(j)e^{\beta j} - 1)} \right| \leq \sum_{j=2}^{\infty} \frac{j2^{2-j}e^{\beta j}}{e^{\beta j}e^{\beta j}/3} \leq 12 \sum_{j=2}^{\infty} j2^{-j} = 18$$

und daher

$$\sum_{j=1}^{\infty} \frac{j}{a(j)e^{\beta j} - 1} = \frac{b^2}{\beta^2} + \frac{1}{\beta} + O(1).$$

Analog haben wir für 3.:

$$
\begin{aligned}
\sum_{j=1}^{\infty} \frac{2j^2 e^{\beta j}}{(2e^{\beta j}-1)^2} - \sum_{j=1}^{\infty} \frac{a(j) j^2 e^{\beta j}}{(a(j)e^{\beta j}-1)^2} \\
&= \frac{2e^{\beta}}{(2e^{\beta}-1)^2} - \frac{e^{\beta}}{(e^{\beta}-1)^2} + \sum_{j=2}^{\infty} \left(\frac{2j^2 e^{\beta j}}{(2e^{\beta j}-1)^2} - \frac{a(j) j^2 e^{\beta j}}{(a(j)e^{\beta j}-1)^2} \right) \\
&= -\frac{1}{\beta^2} + O(\beta^{-1}) + \sum_{j=2}^{\infty} j^2 e^{\beta j} \frac{2(a(j)e^{\beta j}-1)^2 - a(j)(2e^{\beta t}-1)^2}{(2e^{\beta j}-1)^2 (a(j)e^{\beta j}-1)^2} \\
&= -\frac{1}{\beta^2} + O(\beta^{-1}) + \sum_{j=2}^{\infty} j^2 e^{\beta j} \frac{(a(j)-2)(2a(j)e^{2\beta j}-1)}{(2e^{\beta j}-1)^2 (a(j)e^{\beta j}-1)^2}.
\end{aligned}
$$

Nun ist aber

$$
\left| \sum_{j=2}^{\infty} j^2 e^{\beta j} \frac{(a(j)-2)(2a(j)e^{2\beta j}-1)}{(2e^{\beta j}-1)^2 (a(j)e^{\beta j}-1)^2} \right| \le \sum_{j=2}^{\infty} j^2 e^{\beta j} \frac{2^{2-j} \cdot 4e^{2\beta j}}{(e^{\beta j} e^{\beta j}/3)^2} \le 144 \sum_{j=2}^{\infty} j^2 2^{-j} = 792
$$

und daher

$$
\sum_{j=1}^{\infty} \frac{j^2 a(j) e^{\beta j}}{(1-a(j)e^{\beta j})^2} = \frac{2b^2}{\beta^3} + O(\beta^{-2}).
$$

□

Satz 4.9 *Die durchschnittliche Fibonaccizahl eines sternartigen Baumes mit n Kanten ist*

$$
\mathrm{av}(n) \sim A \cdot 2^n \cdot \exp(B\sqrt{n}) \cdot n^{3/4}, \tag{4.16}
$$

wobei $A = 2.739149898\ldots$ eine numerisch berechenbare Konstante und

$$
B = \sqrt{\pi^2/3 - 2(\log 2)^2} - \sqrt{2\pi^2/3} = -1.039005919\ldots
$$

ist.

Beweis: Die einer Partition $c = (c_1, \ldots, c_d)$ zugeordnete Fibonaccizahl ist nach Lemma 4.4

$$
(2^{c_1-1}+1)\ldots(2^{c_d-1}+1) + 2^{n-d}.
$$

Zu bestimmen ist daher zunächst

$$
\sum_{c \vdash n} (2^{c_1-1}+1)\ldots(2^{c_d-1}+1) + 2^{n-d} =: s(n). \tag{4.17}
$$

Die erzeugende Funktion dieses Ausdrucks ist (mit derselben Argumentation wie in Lemma 1.2, wobei an die Stelle von 1 ein Faktor $2^{c_i-1}+1$ bzw. 2^{c_i-1} für jedes Element einer Partition tritt)

$$\prod_{j=1}^{\infty}(1-(2^{j-1}+1)x^j)^{-1}+\prod_{j=1}^{\infty}(1-2^{j-1}x^j)^{-1}. \tag{4.18}$$

Ersetzt man x durch $z/2$, so erhält man die erzeugende Funktion für $2^{-n}s(n)$:

$$\prod_{j=1}^{\infty}\left(1-\left(\frac{1}{2}+2^{-j}\right)z^j\right)^{-1}+\prod_{j=1}^{\infty}\left(1-\frac{1}{2}z^j\right)^{-1}.$$

Wie sich herausstellt, liefert der erste Teil den asymptotisch relevanten Beitrag. Der Beweis wird im Wesentlichen geführt wie jener von Satz 3.1. Sei $G(z) := \prod_{j=1}^{\infty}(1-(\frac{1}{2}+2^{-j})z^j)^{-1}$ und $F(z) := \prod_{j=1}^{\infty}(1-\frac{1}{2}z^j)^{-1}$. Diese Funktionen sind holomorph auf jeder kompakten Kreisscheibe vom Radius < 1 um 0. Daher erhalten wir

$$2^{-n}s(n)=\frac{1}{2\pi i}\Big(\int_{\mathcal{C}_1} z^{-n-1}G(z)\,dz+\int_{\mathcal{C}_2} z^{-n-1}F(z)\,dz\Big),$$

wobei $\mathcal{C}_1$, $\mathcal{C}_2$ geeignete Kurven um 0 sind. Beide Integrale werden nun mit Hilfe der Sattelpunktsmethode abgeschätzt:

$$z^{-n}G(z)=\exp g(z), \tag{4.19}$$

wobei $g(z) = -\sum_{j=1}^{\infty}\log(1-(1/2+2^{-j})z^j)-n\log z$. Die Ableitungen von g sind

$$\begin{aligned}
g'(z) &= \sum_{j=1}^{\infty}\frac{ja(j)^{-1}z^{j-1}}{1-a(j)^{-1}z^j}-\frac{n}{z}\\
&= \sum_{j=1}^{\infty}\frac{j}{z}\frac{1}{a(j)z^{-j}-1}-\frac{n}{z},\\
g''(z) &= \sum_{j=1}^{\infty}\frac{j(1+(j-1)a(j)z^{-j})}{z^2(a(j)z^{-j}-1)^2}+\frac{n}{z^2}\\
&= \sum_{j=1}^{\infty}\frac{j}{z^2(1-a(j)z^{-j})}+\sum_{j=1}^{\infty}\frac{j^2a(j)z^{-j}}{z^2(1-a(j)z^{-j})^2}+\frac{n}{z^2},\\
g'''(z) &= \sum_{j=1}^{\infty}\frac{jz^{-2j}(j^2a(j)(a(j)+z^j)+3ja(j)(-a(j)+z^j)+2(-a(j)+z^j)^2)}{z^3(a(j)z^{-j}-1)^3}-\frac{2n}{z^3},
\end{aligned}$$

wobei wieder die Abkürzung $a(j) := (1/2+2^{-j})^{-1}$ verwendet wird. Sei nun $z_0 = e^{-\beta}$ der Sattelpunkt, der somit die Bedingung

$$g'(e^{-\beta})=0 \iff \sum_{j=1}^{\infty}\frac{j}{(1/2+2^{-j})^{-1}e^{\beta j}-1}=n$$

erfüllen muss. Es ergibt sich nach dem vorigen Lemma

$$\frac{b^2}{\beta^2} + \frac{1}{\beta} + O(1) = n$$

oder

$$\Big(\frac{b}{\beta} + \frac{1}{2b}\Big)^2 = (\sqrt{n} + O(n^{-1/2}))^2,$$

also $\frac{1}{\beta} = \frac{\sqrt{n}}{b} - \frac{1}{2b^2} + O(n^{-1/2})$ bzw. $\beta = \frac{b}{\sqrt{n}} + \frac{1}{2n} + O(n^{-3/2})$. Unter Verwendung des vorigen Lemmas erhält man für den Sattelpunkt

$$\begin{aligned}
g(e^{-\beta}) &= -\sum_{j=1}^{\infty} \log(1-(1/2+2^{-j})e^{-\beta j}) + n\beta \\
&= \frac{b^2}{\beta} - \log\beta - \frac{3}{2}\log 2 - \log\rho + n\beta + O(\beta) \\
&= b^2\Big(\frac{\sqrt{n}}{b} - \frac{1}{2b^2} + O(n^{-1/2})\Big) - \log\Big(\frac{b}{\sqrt{n}} + \frac{1}{2n} + O(n^{-3/2})\Big) - \frac{3}{2}\log 2 \\
&\quad - \log\rho + n\Big(\frac{b}{\sqrt{n}} + \frac{1}{2n} + O(n^{-3/2})\Big) + O(n^{-1/2}) \\
&= b\sqrt{n} - \frac{1}{2} + O(n^{-1/2}) - \log\frac{b}{\sqrt{n}} - \log(1+O(n^{-1/2})) - \frac{3}{2}\log 2 \\
&\quad - \log\rho + b\sqrt{n} + \frac{1}{2} + O(n^{-1/2}) \\
&= 2b\sqrt{n} + \log\frac{\sqrt{n}}{2\sqrt{2}b\rho} + O(n^{-1/2})
\end{aligned}$$

sowie

$$\begin{aligned}
g''(e^{-\beta}) &= \sum_{j=1}^{\infty} \frac{je^{2\beta}}{(1-(1/2+2^{-j})^{-1}e^{\beta j})} + \sum_{j=1}^{\infty} \frac{j^2e^{2\beta}(1/2+2^{-j})^{-1}e^{\beta j}}{(1-(1/2+2^{-j})^{-1}e^{\beta j})^2} + ne^{2\beta} \\
&= e^{2\beta}\Big(-\Big(\frac{b^2}{\beta^2} + \frac{1}{\beta} + O(1)\Big) + \Big(\frac{2b^2}{\beta^3} + O(\beta^{-2})\Big) + n\Big) \\
&= (1+O(\beta))\Big(2b^2\frac{n^{3/2}}{b^3} + O(n)\Big) \\
&= \frac{2}{b}n^{3/2} + O(n).
\end{aligned}$$

Schließlich ergibt sich für die dritte Ableitung

$$g'''(e^{-\beta}) = \sum_{j=1}^{\infty} \frac{je^{2\beta j}(j^2a(j)(a(j)+e^{-\beta j}) + 3ja(j)(-a(j)+e^{-\beta j}) + 2(-a(j)+e^{-\beta j})^2)}{e^{-3\beta}(a(j)e^{\beta j}-1)^3} - 2ne^{3\beta}.$$

Der Zähler kann leicht abgeschätzt werden, indem man $2 \geq a(j) \geq 1$ und $1 \geq e^{-\beta j} \geq 0$ verwendet:

$$-12j \leq j^2 a(j)(a(j) + e^{-\beta j}) + 3ja(j)(-a(j) + e^{-\beta j}) + 2(-a(j) + e^{-\beta j})^2 \leq 6j^2 + 8 \leq 14j^2.$$

Es folgt (da $a(j)e^{\beta j} \geq 1 + e^{\beta j}/3$ für $j \geq 2$ gilt)

$$\begin{aligned} |g'''(e^{-\beta})| &\leq \sum_{j=1}^{\infty} \frac{14j^3 e^{2\beta j}}{e^{-3\beta}(a(j)e^{\beta j} - 1)^3} + 2ne^{3\beta} \\ &\leq \frac{14e^{5\beta}}{(e^{\beta} - 1)^3} + \sum_{j=2}^{\infty} \frac{14j^3 e^{2\beta j + 3\beta}}{(e^{\beta j}/3)^3} + 2ne^{3\beta} \\ &= O(\beta^{-3}) + 378(1 + O(\beta)) \sum_{j=2}^{\infty} j^3 e^{-\beta j} + O(n) \\ &= \frac{378e^{-\beta}(-1 + 4e^{\beta} - 5e^{2\beta} + 8e^{3\beta})}{(e^{\beta} - 1)^4} + O(n^{3/2}) \\ &= O(\beta^{-4}) = O(n^2) \end{aligned}$$

und daher

$$g(z) = g(z_0) + \frac{1}{b} n^{3/2} (z - z_0)^2 + O(n(z - z_0)^2) + O(n^2 (z - z_0)^3).$$

Wir wählen nun $\mathcal{C}_1 = \{z_0 e^{it} \mid -\pi < t \leq \pi\}$. Dann ergibt sich

$$\begin{aligned} \frac{1}{2\pi i} \int_{\mathcal{C}_1} z^{-n-1} G(z)\, dz &= \frac{1}{2\pi i} \int_{\mathcal{C}_1} \frac{\exp g(z)}{z}\, dz \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp g(z_0 e^{it})\, dt. \end{aligned}$$

Für $-n^{-5/7} \leq t \leq n^{-5/7}$ gilt $z - z_0 = z_0(it + O(t^2)) = (1 + O(n^{-1/2}))(it + O(n^{-10/7})) = it + O(n^{-17/14})$ und damit

$$g(z) = g(z_0) - \frac{1}{b} n^{3/2} t^2 + O(n^{-1/7}). \tag{4.20}$$

Daher ist ein Teil des Integrals

$$\frac{1}{2\pi} \int_{-n^{-5/7}}^{n^{-5/7}} \exp g(z_0 e^{it})\, dt = \frac{1}{2\pi} (\exp g(z_0))(1 + O(n^{-1/7})) \int_{-n^{-5/7}}^{n^{-5/7}} e^{-\frac{1}{b} n^{3/2} t^2}\, dt,$$

wobei

$$\int_{n^{-5/7}}^{\infty} e^{-\frac{1}{b} n^{3/2} t^2}\, dt \leq \int_{n^{-5/7}}^{\infty} e^{-\frac{1}{b} n^{11/14} t}\, dt = b n^{-11/14} e^{-\frac{1}{b} n^{1/14}},$$

und daher

$$\int_{-n^{-5/7}}^{n^{-5/7}} e^{-\frac{1}{b}n^{3/2}t^2}\,dt = \int_{-\infty}^{\infty} e^{-\frac{1}{b}n^{3/2}t^2}\,dt - 2\int_{n^{-5/7}}^{\infty} e^{-\frac{1}{b}n^{3/2}t^2}\,dt$$
$$= \sqrt{\frac{\pi b}{n^{3/2}}} + O(n^{-11/14}e^{-\frac{1}{b}n^{1/14}}).$$

Für diesen Teil des Integrals erhalten wir somit

$$\frac{1}{2\pi}\int_{-n^{-5/7}}^{n^{-5/7}} \exp g(z_0e^{it})\,dt = \sqrt{\frac{b}{4\pi n^{3/2}}}(\exp g(z_0))(1+O(n^{-1/7}))$$
$$= \frac{1}{4\rho\sqrt{2\pi b}}e^{2b\sqrt{n}}n^{-1/4}(1+O(n^{-1/7})).$$

Der Rest des Integrals ist wieder vernachlässigbar: dies wird erneut gezeigt, indem wir zunächst beweisen, dass $\left|\frac{G(z_0)}{G(z_0e^{it})}\right|$ für $n^{-5/7} \le |t| \le \pi$ schneller als jede Potenz von n wächst. Es gilt

$$\begin{aligned}
\left|\frac{G(z_0)}{G(z_0e^{it})}\right|^2 &= \left|\prod_{j=1}^{\infty}\frac{(1-a(j)^{-1}z_0^j)^{-1}}{(1-a(j)^{-1}z_0^je^{itj})^{-1}}\right|^2\\
&= \left|\prod_{j=1}^{\infty}\frac{a(j)-z_0^je^{itj}}{a(j)-z_0^j}\right|^2\\
&= \prod_{j=1}^{\infty}\frac{|(a(j)-z_0^j\cos(tj))-iz_0^j\sin(tj)|^2}{|a(j)-z_0^j|^2}\\
&= \prod_{j=1}^{\infty}\frac{(a(j)-z_0^j\cos(tj))^2+(z_0^j\sin(tj))^2}{(a(j)-z_0^j)^2}\\
&= \prod_{j=1}^{\infty}\frac{a(j)^2-2a(j)z_0^j\cos(tj)+z_0^{2j}}{(a(j)-z_0^j)^2}\\
&= \prod_{j=1}^{\infty}\Big(1+\frac{2a(j)(1-\cos(tj))z_0^j}{(a(j)-z_0^j)^2}\Big)\\
&\ge \prod_{j=1}^{\infty}\Big(1+z_0^j(1-\cos(tj))\Big)\\
&\ge \prod_{\sqrt{n}\le j\le 2\sqrt{n}}\Big(1+z_0^j(1-\cos(tj))\Big)\\
&\ge \prod_{\sqrt{n}\le j\le 2\sqrt{n}}\Big(1+z_0^{2\sqrt{n}}(1-\cos(tj))\Big)\\
&\ge \prod_{\sqrt{n}\le j\le 2\sqrt{n}}\Big(1+m(1-\cos(tj))\Big),
\end{aligned}$$

wobei m eine Konstante ist, denn $z_0^{2\sqrt{n}} = \exp(-2b+O(n^{-1/2})) = O(1)$. Wie im Beweis von Satz 3.1 erhalten wir

$$\Big|\frac{G(z_0)}{G(z_0e^{it})}\Big|^2 \geq \exp(C_2 n^{1/14} + o(1)). \tag{4.21}$$

Daraus folgern wir

$$\begin{aligned}\frac{1}{2\pi}\Big|\int_{n^{-5/7}}^{\pi} \exp g(z_0e^{it})\,dt\Big| &\leq \frac{1}{2\pi}\int_{n^{-5/7}}^{\pi} |z_0e^{it}|^{-n}|G(z_0e^{it})|\,dt\\ &= \frac{1}{2\pi}\int_{n^{-5/7}}^{\pi} z_0^{-n}G(z_0)\Big|\frac{G(z_0e^{it})}{G(z_0)}\Big|\,dt\\ &\leq \frac{1}{2}(\exp g(z_0))\exp(-C_2n^{1/14}/2 + o(1))\end{aligned}$$

und analog auch wieder für das Integral zwischen $-\pi$ und $-n^{-5/7}$. Daher erhalten wir

$$\frac{1}{2\pi i}\int_{\mathcal{C}_1} z^{-n-1}G(z)\,dz \sim \frac{1}{4\rho\sqrt{2\pi b}}e^{2b\sqrt{n}}n^{-1/4}. \tag{4.22}$$

Analog wird nun mit $F(z)$ verfahren, wobei sich herausstellt, dass der Beitrag dieser Funktion kleiner ist:

$$z^{-n}F(z) = \exp f(z), \tag{4.23}$$

wobei $f(z) = -\sum_{j=1}^{\infty}\log(1-z^j/2) - n\log z$. Die Ableitungen von f sind

$$\begin{aligned}f'(z) &= \sum_{j=1}^{\infty}\frac{jz^{j-1}/2}{1-z^j/2} - \frac{n}{z}\\ &= \sum_{j=1}^{\infty}\frac{j}{z}\frac{1}{2z^{-j}-1} - \frac{n}{z},\\ f''(z) &= \sum_{j=1}^{\infty}\frac{j(1+2(j-1)z^{-j})}{z^2(2z^{-j}-1)^2} + \frac{n}{z^2}\\ &= \sum_{j=1}^{\infty}\frac{j}{z^2(1-2z^{-j})} + \sum_{j=1}^{\infty}\frac{2j^2z^{-j}}{z^2(1-2z^{-j})^2} + \frac{n}{z^2},\\ f'''(z) &= \sum_{j=1}^{\infty}\frac{jz^{-2j}(2j^2(2+z^j)+6j(-2+z^j)+2(-2+z^j)^2)}{z^3(2z^{-j}-1)^3} - \frac{2n}{z^3}.\end{aligned}$$

Sei wieder $z_0 = e^{-\beta}$ der Sattelpunkt, der somit die Bedingung

$$f'(e^{-\beta}) = 0 \iff \sum_{j=1}^{\infty}\frac{j}{2e^{\beta j}-1} = n$$

erfüllen muss. Es ergibt sich nach Lemma 4.7

$$\frac{b^2}{\beta^2} + O(1) = n$$

oder

$$\left(\frac{b}{\beta}\right)^2 = (\sqrt{n} + O(n^{-1/2}))^2,$$

also $\frac{1}{\beta} = \frac{\sqrt{n}}{b} + O(n^{-1/2})$ bzw. $\beta = \frac{b}{\sqrt{n}} + O(n^{-3/2})$. Unter Verwendung desselben Lemmas erhält man für den Sattelpunkt

$$\begin{aligned}
f(e^{-\beta}) &= -\sum_{j=1}^{\infty} \log(1 - e^{-\beta j}/2) + n\beta \\
&= \frac{b^2}{\beta} - \frac{1}{2}\log 2 + n\beta + O(\beta) \\
&= b^2\Big(\frac{\sqrt{n}}{b} + O(n^{-1/2})\Big) - \frac{1}{2}\log 2 + n\Big(\frac{b}{\sqrt{n}} + O(n^{-3/2})\Big) + O(n^{-1/2}) \\
&= b\sqrt{n} + O(n^{-1/2}) - \frac{1}{2}\log 2 + b\sqrt{n} + O(n^{-1/2}) \\
&= 2b\sqrt{n} - \frac{1}{2}\log 2 + O(n^{-1/2})
\end{aligned}$$

sowie

$$\begin{aligned}
f''(e^{-\beta}) &= \sum_{j=1}^{\infty} \frac{je^{2\beta}}{(1-2e^{\beta j})} + \sum_{j=1}^{\infty} \frac{2j^2 e^{2\beta} e^{\beta j}}{(1-2e^{\beta j})^2} + ne^{2\beta} \\
&= e^{2\beta}\Big(-\Big(\frac{b^2}{\beta^2} + O(1)\Big) + \Big(\frac{2b^2}{\beta^3} + O(\beta^{-2})\Big) + n\Big) \\
&= (1 + O(\beta))\Big(2b^2 \frac{n^{3/2}}{b^3} + O(n)\Big) \\
&= \frac{2}{b} n^{3/2} + O(n).
\end{aligned}$$

Schließlich ergibt sich für die dritte Ableitung analog zu g'''

$$f'''(e^{-\beta}) = O(n^2)$$

und daher

$$f(z) = f(z_0) + \frac{1}{b} n^{3/2}(z - z_0)^2 + O(n(z - z_0)^2) + O(n^2(z - z_0)^3)).$$

Dieselbe Argumentation wie zuvor führt auf

$$\frac{1}{2\pi i}\int_{\mathcal{C}_2} z^{-n-1}F(z)\,dz \sim \sqrt{\frac{b}{8\pi}} e^{2b\sqrt{n}} n^{-3/4}. \tag{4.24}$$

Zusammenfassend ergibt sich damit

$$s(n) \sim 2^n \frac{1}{4\rho\sqrt{2\pi b}} e^{2b\sqrt{n}} n^{-1/4}. \tag{4.25}$$

Jeder Partition ist genau ein sternartiger Baum zugeordnet und umgekehrt, wobei es nur $\lfloor \frac{n}{2} \rfloor$ Ausnahmebäume gibt, die zwei Partitionen zugeordnet sind. Deren Fibonaccizahl ist jedoch sicher jeweils $\leq 2^n + 1$, der maximalen Fibonaccizahl (s. [16]), die Summe aller ihrer Fibonaccizahlen also klein im Vergleich zu $s(n)$. Also erhält man als durchschnittliche Fibonaccizahl eines sternartigen Baumes

$$\begin{aligned} \operatorname{av}(n) \sim s(n)/p(n) &\sim 2^n \frac{1}{4\rho\sqrt{2\pi b}} e^{2b\sqrt{n}} n^{-1/4} e^{-\pi\sqrt{2n/3}} \cdot 4\sqrt{3}n \\ &= A \cdot 2^n \cdot \exp(B\sqrt{n}) \cdot n^{3/4}, \end{aligned}$$

wobei, wie behauptet, $B = 2b - \pi\sqrt{2/3} = \sqrt{\pi^2/3 - 2(\log 2)^2} - \sqrt{2\pi^2/3}$ und $A = \frac{1}{\rho\sqrt{2\pi b/3}} = 2.739149898\ldots$. □

Bemerkung: In [16] wird gezeigt, dass die maximale Fibonaccizahl eines sternartigen Graphen asymptotisch 2^n ist, die minimale $5^{n/3}$. Das vorliegende Ergebnis zeigt, dass die logarithmierte Fibonaccizahl eines durchschnittlichen sternartigen Graphen asymptotisch gleich groß ist wie die des maximalen. Die Abschätzung $\operatorname{av}(n) = 2^{n+O(\sqrt{n})}$ lässt sich mit bedeutend einfacheren Mitteln gewinnen und wird auch bereits in [16] erwähnt.

Teil II

Das Waring'sche Problem

Kapitel 5

Das Problem und einige Spezialfälle

1770 konnte Lagrange den sogenannten "Vier-Quadrate-Satz" beweisen, demzufolge sich jede natürliche Zahl als Summe von vier Quadratzahlen schreiben lässt. Dies war bereits von Bachet (1621) und von Fermat (1640) vermutet worden. In seinen *Meditationes algebraicae* [37] stellte Waring 1770 die allgemeinere Vermutung auf, dass jede natürliche Zahl Summe von neun Kuben, neunzehn vierten Potenzen und allgemein Summe einer begrenzten Anzahl von k-ten Potenzen ist. Dies konnte 1909 von Hilbert [10] bewiesen werden, wobei ein induktives Argument zusammen mit polynomiellen Identitäten wie

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)^2 = \frac{1}{6} \sum_{1 \leq i < j \leq 4} (x_i + x_j)^4 + \frac{1}{6} \sum_{1 \leq i < j \leq 4} (x_i - x_j)^4$$

verwendet wird.

Nun ist es aber auch von Interesse, möglichst gute Schranken für die maximale Anzahl benötigter k-ter Potenzen sowie auch asymptotische Formeln für die Anzahl verschiedener Lösungen zu finden. Hier spielt die "circle method" wieder eine entscheidende Rolle. Mit ihrer Hilfe konnten Hardy und Littlewood [7] zehn Jahre nach Hilbert Antwort auf diese Probleme geben. Ihre Technik wurde später u.a. von Vinogradov (s. [34, 35, 36]) und Wooley [41] weiter verbessert.

Zunächst seien die wesentlichsten Begriffe definiert:

Definition 5.1 Eine Teilmenge B der natürlichen Zahlen heißt *Basis* bzw. *asymptotische Basis* der Ordnung s, wenn jede natürliche Zahl bzw. jede hinreichend große natürliche Zahl sich als Summe von s Summanden aus B schreiben lässt.

Die Vermutung von Waring besagt also, dass die Menge der k-ten Potenzen für beliebiges k eine Basis endlicher Ordnung darstellt. Für die entsprechenden Ordnungen sind die Bezeichnungen $g(n)$ bzw. $G(n)$ üblich:

Definition 5.2 Die Ordnung der Menge der k-ten Potenzen als Basis bzw. asymptotische Basis wird mit $g(n)$ bzw. $G(n)$ bezeichnet.

Es ist seit Lagrange bekannt, dass $g(2) = G(2) = 4$ ist. Wieferich [40] und Kempner [15] konnten zeigen, dass $g(3) = 9$ ist (es lässt sich zumindest leicht einsehen, dass dies minimal ist, da sich 23 nicht als Summe von acht Kuben schreiben lässt; neben 239 ist dies das einzige solche Beispiel). Es ist bekannt, dass $4 \leq G(4) \leq 7$ gilt, die genaue Ordnung der Menge der Kuben als asymptotische Basis ist jedoch nach wie vor ein ungelöstes Problem. Für $k = 4$ sind dagegen $g(4) = 19$ und $G(4) = 16$ wiederum bekannt.

Für die Größe $g(k)$ besteht eine allgemeine Vermutung:

$$g(k) = 2^k + \left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 2 \tag{5.1}$$

für beliebiges $k \geq 1$. Dies wurde bereits von Euler als untere Schranke angegeben. Zum Beweis betrachte man die Zahl

$$N = 2^k \left\lfloor \left(\frac{3}{2}\right)^k \right\rfloor - 1;$$

sie lässt sich nicht als Summe von weniger k-ten Potenzen schreiben, da $N < 3^k$ und daher nur Summen von 2^k und 1 gebildet werden können. Man benötigt $\lfloor (\frac{3}{2})^k \rfloor - 1$ Summanden 2^k und $2^k - 1$ Summanden 1.

5.1 Der Spezialfall $k = 1$

Offensichtlich ist $g(1) = G(1) = 1$. Auch für die Anzahl verschiedener Repräsentationen durch s Summanden gibt es eine einfache Formel, die hier kurz angeführt sei – wir werden sie in Kapitel 7 zu Vergleichszwecken heranziehen:

Proposition 5.1 Die Anzahl von Darstellungen von $n > 0$ als Summe von s positiven ganzen Zahlen (diesmal ohne Rücksicht auf die Reihenfolge) ist $\binom{n-1}{s-1}$.

Beweis: Betrachten wir n als eine Streckenlänge auf der Zahlengeraden, so entspricht eine Darstellung als Summe von s positiven ganzen Zahlen einer Art, diese Strecke durch $s - 1$ Trennstriche in Teile zu zerlegen. Da es $n - 1$ verschiedene Positionen für die Trennstriche gibt, erhalten wir genau $\binom{n-1}{s-1}$ Möglichkeiten. □

5.2 Der Spezialfall $k = 2$

Wie bereits erwähnt, stammt der Beweis für diesen Spezialfall des Waring'schen Problems von Lagrange. Dieser verwendet ein induktives Argument, das auf der Vier-Quadrate-Gleichung von Euler beruht:

Lemma 5.2 Es gilt die polynomielle Identität

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2, \tag{5.2}$$

wobei

$$
\begin{aligned}
z_1 &= x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 \\
z_2 &= x_1y_2 - x_2y_1 - x_3y_4 + x_4y_3 \\
z_3 &= x_1y_3 + x_2y_4 - x_3y_1 - x_4y_2 \\
z_4 &= x_1y_4 - x_2y_3 + x_3y_2 - x_4y_1.
\end{aligned}
\tag{5.3}
$$

Beweis: Die Richtigkeit der Gleichung ergibt sich durch simples Ausmultiplizieren. Es sei lediglich bemerkt, dass $x_1^2 + x_2^2 + x_3^2 + x_4^2$ als quadrierter Absolutbetrag des Elements $x = x_1 - x_2 i - x_3 j - x_4 k$ aus der Menge der Hamiltonschen Quaternionen interpretiert werden kann. Mit $y = y_1 + y_2 i + y_3 j + y_4 k$ und $z = xy = z_1 + z_2 i + z_3 j + z_4 k$ entspricht die obige polynomielle Identität dann der Gleichung $|x||y| = |xy|$ in der Menge der Quaternionen. □

Korollar 5.3 Können n und m als Summe von vier Quadraten geschrieben werden, dann auch nm.

Beweis: Ist $n = x_1^2 + x_2^2 + x_3^2 + x_4^2$ und $m = y_1^2 + y_2^2 + y_3^2 + y_4^2$, so liefert obige Formel unmittelbar eine Darstellung $nm = z_1^2 + z_2^2 + z_3^2 + z_4^2$. □

Satz 5.4 (Vier-Quadrate-Satz, Lagrange) *Jede natürliche Zahl ist Summe von vier Quadratzahlen.*

Nach dem Korollar genügt es, sich für den Beweis, sich auf Primzahlen p einzuschränken. Wegen $2 = 1^2 + 1^2 + 0^2 + 0^2$ reicht es, ungerade Primzahlen zu berücksichtigen.

Die Mengen

$$\{a^2 \mid a = 0, \ldots, (p-1)/2\} \quad \text{und} \quad \{-b^2 - 1 \mid b = 0, \ldots, (p-1)/2\}$$

enthalten jeweils $(p+1)/2$ verschiedene Kongruenzklassen modulo p. Nach dem Schubfachschlussprinzip gibt es daher a und b, sodass a^2 und $-b^2 - 1$ in derselben Kongruenzklasse liegen, d.h. $a^2 = -b^2 - 1 + np$ für ein $n \in \mathbb{Z}$. Dann gilt

$$p \leq np = a^2 + b^2 + 1^2 + 0^2 \leq 2\left(\frac{p-1}{2}\right)^2 + 1 < \frac{p^2}{2} + 1 < p^2.$$

Daher ist $1 \leq n < p$. Es sei nun m die kleinste positive ganze Zahl, für die mp Summe von vier Quadraten ist, es sei also

$$mp = x_1^2 + x_2^2 + x_3^2 + x_4^2.$$

Wenn gezeigt werden kann, dass $m = 1$ sein muss, ist der Beweis erbracht. Dies wird durch Widerspruch gezeigt, wir nehmen also $m > 1$ an. Wegen der Minimalitätseigenschaft gilt $m \leq n$ und daher $1 < m < p$. Man kann nun ganze Zahlen y_i derart wählen, dass

$$y_i \equiv x_i \mod m$$

und $-m/2 < y_i \leq m/2$ für alle i. Dann folgt

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 \equiv x_1^2 + x_2^2 + x_3^2 + x_4^2 \equiv 0 \mod m$$

und daher $mr = y_1^2 + y_2^2 + y_3^2 + y_4^2$ für eine ganze Zahl $r \geq 0$. Falls $r = 0$, so sind alle y_i gleich 0 und damit jedes x_i durch m teilbar, womit $m^2|mp$ im Widerspruch zu $1 < m < p$ folgen würde. Also ist $r \geq 1$ und

$$mr = y_1^2 + y_2^2 + y_3^2 + y_4^2 \leq 4(m/2)^2 = m^2.$$

$r = m$ gilt nur, falls alle y_i gleich $m/2$ sind. Dann wäre aber wieder mp durch m^2 teilbar, was unmöglich ist. Daher gilt $1 \leq r < m$. Nun verwenden wir wieder das Lemma 5.2 und erhalten

$$\begin{aligned} m^2rp = (mp)(mr) &= (x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) \\ &= z_1^2 + z_2^2 + z_3^2 + z_4^2, \end{aligned}$$

wobei sich die z_i wie in der Formulierung des Lemmas ergeben. Wegen $y_i \equiv x_i \mod m$ folgt aus diesen Gleichungen $z_i \equiv 0 \mod m$ für alle i. Daher kann man $w_i = z_i/m$ setzen und erhält

$$rp = w_1^2 + w_2^2 + w_3^2 + w_4^2$$

mit $1 \leq r < m$, im Widerspruch zur Minimalität von m. □

Bemerkung: Legendre konnte 1798 zeigen, dass jede natürliche Zahl, die nicht von der Form $4^a(8b+7)$ ist, sogar als Summe dreier Quadrate geschrieben werden kann. Der Beweis wird meist mit Hilfe quadratischer Formen geführt (siehe etwa [21], S.7ff.). Schließlich lässt sich eine natürliche Zahl n genau dann als Summe von zwei Quadraten schreiben, wenn jeder Primfaktor der Form $4m + 3$ in der Primfaktorzerlegung von n mit einer geraden Potenz vorkommt (siehe z.B. [14], S.45 f.).

Kapitel 6

Die Ungleichung von Weyl und das Lemma von Hua

Die wesentlichen beiden Hilfsmittel im Beweis der asymptotischen Formel von Hardy und Littlewood sind die Ungleichung von Weyl und das Lemma von Hua, die zur Abschätzung von Exponentialsummen dienen und auch in anderen Bereichen der analytischen Zahlentheorie Anwendung finden. Diese beiden Hilfsmittel werden nun in diesem Kapitel am Ende einer Folge von Lemmata bewiesen, bevor auf den eigentlichen Beweis im folgenden Kapitel eingegangen wird. Die Beweisschritte folgen dabei im Wesentlichen jenen in [21]. Zunächst folgen einige Resultate über den Differenzoperator Δ_d.

6.1 Der Differenzoperator

Definition 6.1 Der (Vorwärts-)*Differenzoperator* Δ_d ist durch

$$\Delta_d(f)(x) = f(x+d) - f(x)$$

definiert. Durch wiederholte Anwendung erhält man den *iterierten Differenzoperator* $\Delta_{d_l,\ldots,d_1}$:

$$\Delta_{d_l,\ldots,d_1} = \Delta_{d_l} \circ \ldots \circ \Delta_{d_1}.$$

Weiters sei $\Delta^{(l)}$ der l-fach iterierte Differenzoperator Δ_1, d.h. $\Delta^{(l)} = \Delta_{1,\ldots,1}$.

Lemma 6.1 Es sei $l \geq 1$. Dann gilt

$$\Delta^{(l)}(f)(x) = \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} f(x+j). \tag{6.1}$$

Beweis: Durch vollständige Induktion. Für $l = 1$ ist die Gültigkeit der Formel offensichtlich. Der Induktionsschritt ergibt sich in folgender Weise:

$$\begin{aligned}
\Delta^{(l+1)}(f)(x) &= \Delta^{(l)}(f)(x+1) - \Delta^{(l)}(f)(x) \\
&= \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} f(x+j+1) - \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} f(x+j) \\
&= \sum_{j=1}^{l+1} (-1)^{l+1-j} \binom{l}{j-1} f(x+j) - \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} f(x+j) \\
&= f(x+l+1) + \sum_{j=1}^{l} (-1)^{l+1-j} \left(\binom{l}{j-1} + \binom{l}{j} \right) f(x+j) - (-1)^l f(x) \\
&= \sum_{j=0}^{l+1} (-1)^{l+1-j} \binom{l+1}{j} f(x+j).
\end{aligned}$$

□

Lemma 6.2 Es sei $1 \leq l \leq k$. Dann gilt für den iterierten Differenzoperator

$$\Delta_{d_l,\ldots,d_1}(x^k) = \sum_{j,j_1,\ldots,j_l} \frac{k!}{j! j_1! \ldots j_l!} d_1^{j_1} \ldots d_l^{j_l} x^j = d_1 \ldots d_l p_{k-l}(x), \tag{6.2}$$

wobei die Summe über alle $j, j_1, \ldots, j_l$ mit $j + j_1 + \ldots + j_l = k$ und $j \geq 0$, $j_i \geq 1$ geht und $p_{k-l}(x)$ ein Polynom vom Grad $k-l$ mit Leitkoeffizient $k(k-1)\ldots(k-l+1)$ ist. Sind die d_i ganze Zahlen, dann ist $p_{k-l}(x)$ ein ganzzahliges Polynom.

Beweis: Wieder mittels Induktion nach l. Für $l = 1$ folgt das Ergebnis unmittelbar aus dem binomischen Lehrsatz. Für den Induktionsschritt erhalten wir

$$\begin{aligned}
\Delta_{d_{l+1},\ldots,d_1}(x^k) = \Delta_{d_{l+1}}(\Delta_{d_l,\ldots,d_1}(x^k)) &= \sum_{\substack{m+j_1+\ldots+j_l=k \\ m\geq 0, j_i \geq 1}} \frac{k!}{m! j_1! \ldots j_l!} d_1^{j_1} \ldots d_l^{j_l} \Delta_{j_{l+1}}(x^m) \\
&= \sum_{\substack{m+j_1+\ldots+j_l=k \\ m\geq 0, j_i \geq 1}} \frac{k!}{m! j_1! \ldots j_l!} d_1^{j_1} \ldots d_l^{j_l} \sum_{\substack{j_{l+1}+j=m \\ j\geq 0, j_{l+1} \geq 1}} \frac{m!}{j! j_{l+1}!} d_{l+1}^{j_{l+1}} x^j \\
&= \sum_{\substack{j+j_1+\ldots+j_{l+1}=k \\ j\geq 0, j_i \geq 1}} \frac{k!}{j! j_1! \ldots j_l! j_{l+1}!} d_1^{j_1} \ldots d_l^{j_l} d_{l+1}^{j_{l+1}} x^j.
\end{aligned}$$

Korollar 6.3 Für $k \geq 2$ gilt

$$\Delta_{d_{k-1},\ldots,d_1}(x^k) = d_1 \ldots d_{k-1} k! \Big(x + \frac{d_1 + \ldots + d_{k-1}}{2} \Big). \tag{6.3}$$

Lemma 6.4 Es sei $l \geq 1$ und $f(x) = \alpha x^k + \ldots$ ein Polynom vom Grad k. Dann gilt

$$\Delta_{d_l,\ldots,d_1}(f)(x) = (d_1 \ldots d_l)\,(k(k-1)\ldots(k-l+1))\,\alpha x^{k-l} + \ldots \tag{6.4}$$

Beweis: Folgt aus der Linearität des Differenzoperators und den vorangegangenen Lemmata. □

Lemma 6.5 Es sei $1 \leq l \leq k$. Wenn $-P \leq d_1, \ldots, d_l, x \leq P$, dann gilt die asymptotische Abschätzung

$$\Delta_{d_l,\ldots,d_1}(x^k) \ll P^k, \tag{6.5}$$

wobei die implizierte Konstante nur von k abhängt.

Beweis: Aus dem Lemma 6.2 folgt

$$\begin{aligned}
|\Delta_{d_l,\ldots,d_1}(x^k)| &\leq \sum_{\substack{j+j_1+\ldots+j_l=k \\ j\geq 0, j_i \geq 1}} \frac{k!}{j!j_1!\ldots j_l!} P^{j_1+\ldots+j_l+j} \\
&\leq \sum_{\substack{j+j_1+\ldots+j_l=k \\ j\geq 0, j_i \geq 0}} \frac{k!}{j!j_1!\ldots j_l!} P^{k} \\
&= (l+1)^k P^k \leq (k+1)^k P^k \ll P^k.
\end{aligned}$$

□

6.2 Norm einer reellen Zahl und Exponentialsummen

Ich beginne mit einem kurzen Beweis des berühmten Dirichlet'schen Approximationssatzes:

Lemma 6.6 (Dirichlet) Sind α und Q reelle Zahlen mit $Q \geq 1$, so gibt es teilerfremde ganze Zahlen a und $q \leq Q$, sodass

$$\left|\alpha - \frac{a}{q}\right| < \frac{1}{qQ}. \tag{6.6}$$

Beweis: Es sei $N = \lfloor Q \rfloor$. Angenommen, es wäre $\{q\alpha\} \in [0, 1/(N+1))$ für eine positive ganze Zahl $q \leq N$. Setzt man dann $a = \lfloor q\alpha \rfloor$, dann ergibt sich

$$0 \leq \{q\alpha\} = q\alpha - \lfloor q\alpha \rfloor = q\alpha - a < \frac{1}{N+1}$$

und damit

$$\left|\alpha - \frac{a}{q}\right| < \frac{1}{q(N+1)} < \frac{1}{qQ} \leq \frac{1}{q^2}.$$

Analoges folgt, wenn $\{q\alpha\} \in [N/(N+1), 1)$ gilt. Wir nehmen nun an, es wäre $\{q\alpha\} \in [1/(N+1), N/(N+1))$ für alle $q \leq N$ (andernfalls sind a, q bereits gefunden; sind sie nicht teilerfremd, so dividiert man einfach durch den größten gemeinsamen Teiler). Dann liegt jede dieser Zahlen in einem Intervall der Form $[i/(N+1),(i+1)/(N+1))$ für $1 \le i \le N-1$. Nach dem Schubfachschlussprinzip gibt es damit $q_1, q_2 \le N$, sodass $\{q_1\alpha\}$ und $\{q_2\alpha\}$ im gleichen Intervall liegen. Damit folgt aber $|\{q_1\alpha\} - \{q_2\alpha\}| < 1/(N+1)$ und daher mit $q = q_2 - q_1$ und $a = \lfloor q_2\alpha \rfloor - \lfloor q_1\alpha \rfloor$:

$$|q\alpha - a| = |(q_2\alpha - \lfloor q_2\alpha\rfloor) - (q_1\alpha - \lfloor q_1\alpha\rfloor)| = |\{q_2\alpha\} - \{q_1\alpha\}| < \frac{1}{N+1} < \frac{1}{Q}.$$

□

Es folgt die Definition der Norm sowie einige ihrer Eigenschaften:

Definition 6.2 Die Norm einer reellen Zahl α ist der Abstand zur nächsten ganzen Zahl, d.h.

$$\|\alpha\| = \min(\{\alpha\}, 1 - \{\alpha\}). \tag{6.7}$$

Bemerkung: Man kann damit α in der Form $\alpha = n \pm \|\alpha\|$ für ganzzahliges n schreiben, womit

$$|\sin \pi\alpha| = \sin \pi\|\alpha\|$$

folgt. Die Norm erfüllt weiters die Dreiecksungleichung, d.h.

$$\|\alpha + \beta\| \le \|\alpha\| + \|\beta\|.$$

Lemma 6.7 Für $0 < \alpha < \frac{1}{2}$ gilt die Ungleichung

$$2\alpha < \sin \pi\alpha < \pi\alpha. \tag{6.8}$$

Beweis: Die zweite Ableitung der Funktion $f(x) = \sin \pi x$ ist $-\pi^2 \sin \pi x$, was auf dem Intervall $(0, \frac{1}{2})$ strikt negativ ist. f ist daher auf diesem Intervall konkav und liegt somit zwischen der Sehne zwischen 0 und $\frac{1}{2}$ und der Tangente in 0. □

Lemma 6.8 Es sei (wie auch im Folgenden) $e(t) = e^{2\pi i t}$. Für $\alpha \in \mathbb{R}$ und natürliche Zahlen $N_1 < N_2$ gilt

$$\sum_{n=N_1+1}^{N_2} e(\alpha n) \ll \min(N_2 - N_1, \|\alpha\|^{-1}). \tag{6.9}$$

Beweis: Da für alle reellen Zahlen $|e(t)| = 1$ gilt, folgt die Abschätzung durch $N_2 - N_1$ in trivialer Weise. Andererseits bilden die $e(\alpha n)$ eine geometrische Reihe, und daher gilt auch

$$\begin{aligned}
\Big|\sum_{n=N_1+1}^{N_2} e(\alpha n)\Big| &= \Big|e(\alpha(N_1+1)) \sum_{n=0}^{N_2-N_1-1} e(\alpha)^n\Big| \\
&= \Big|\frac{e(\alpha(N_2-N_1)) - 1}{e(\alpha) - 1}\Big| \\
&\le \frac{2}{|e(\alpha) - 1|} = \frac{2}{|e(\alpha/2) - e(-\alpha/2)|}
\end{aligned}$$

$$\begin{aligned}
&= \frac{2}{|2i \sin \pi\alpha|} = \frac{1}{|\sin \pi\alpha|} \\
&= \frac{1}{\sin \pi\|\alpha\|} \le \frac{1}{2\|\alpha\|}.
\end{aligned}$$

□

Lemma 6.9 Es seien $\alpha \in \mathbb{R}$ und a, q teilerfremde ganze Zahlen mit $q \ge 1$. Wenn $|\alpha - \frac{a}{q}| \le \frac{1}{q^2}$, dann gilt

$$\sum_{1 \le r \le q/2} \frac{1}{\|\alpha r\|} \ll q \log q. \tag{6.10}$$

Beweis: Für $q = 1$ ist das Lemma offensichtlich erfüllt, da beide Seiten gleich 0 sind. Daher können wir $q \ge 2$ annehmen. Nun gibt es zu jeder ganzen Zahl r ein $s(r) \in \mathbb{Z}$ mit $0 \le s(r) \le q/2$ und ein $m(r) \in \mathbb{Z}$, sodass

$$\frac{s(r)}{q} = \left\|\frac{ar}{q}\right\| = \pm\Big(\frac{ar}{q} - m(r)\Big).$$

Nachdem a und q teilerfremd sind, gilt $s(r) = 0$ genau für $r \equiv 0 \bmod q$, also ist $s(r) \ge 1$ für $1 \le r \le q/2$. Es sei

$$\alpha - \frac{a}{q} = \frac{\theta}{q^2}$$

für ein $\theta \in [-1, 1]$. Dann folgt

$$\alpha r = \frac{ar}{q} + \frac{\theta r}{q^2} = \frac{ar}{q} + \frac{\theta'}{2q},$$

wobei $|\theta'| = |2\theta r/q| \le |\theta| \le 1$. Damit folgt nach der Dreiecksungleichung für die Norm

$$\begin{aligned}
\|\alpha r\| &= \left\|\frac{ar}{q} + \frac{\theta'}{2q}\right\| = \left\|m(r) \pm \frac{s(r)}{q} + \frac{\theta'}{2q}\right\| \\
&= \left\|\frac{s(r)}{q} \pm \frac{\theta'}{2q}\right\| \ge \left\|\frac{s(r)}{q}\right\| - \left\|\frac{\theta'}{2q}\right\| \\
&\ge \frac{s(r)}{q} - \frac{1}{2q} \ge \frac{1}{2q}.
\end{aligned}$$

Sei nun $1 \le r_1 \le r_2 \le q/2$. Es zeigt sich, dass $s(r_1) = s(r_2)$ bereits $r_1 = r_2$ impliziert: aus $s(r_1) = s(r_2)$ folgt nämlich

$$\left\|\frac{ar_1}{q}\right\| = \pm\Big(\frac{ar_1}{q} - m(r_1)\Big) = \pm\Big(\frac{ar_2}{q} - m(r_2)\Big) = \left\|\frac{ar_2}{q}\right\|$$

und damit $ar_1 \equiv \pm ar_2 \bmod q$. Da a und q teilerfremd sind, muss damit $r_1 \equiv \pm r_2 \bmod q$ gelten, was nur für $r_1 = r_2$ möglich ist. Damit folgt, dass

$$\left\{\left\|\frac{ar}{q}\right\| : 1 \le r \le \frac{q}{2}\right\} = \left\{\frac{s(r)}{q} : 1 \le r \le \frac{q}{2}\right\} = \left\{\frac{s}{q} : 1 \le s \le \frac{q}{2}\right\}$$

und somit

$$\sum_{1\le r\le q/2}\frac{1}{\|\alpha r\|}\le\sum_{1\le r\le q/2}\frac{1}{\frac{s(r)}{q}-\frac{1}{2q}}=\sum_{1\le s\le q/2}\frac{1}{\frac{s}{q}-\frac{1}{2q}}$$
$$=2q\sum_{1\le s\le q/2}\frac{1}{2s-1}\le 2q\sum_{1\le s\le q/2}\frac{1}{s}\ll q\log q.$$

□

Lemma 6.10 Es seien $\alpha\in\mathbb{R}$ und a,q teilerfremde ganze Zahlen mit $q\ge 1$. Wenn $|\alpha-\frac{a}{q}|\le\frac{1}{q^2}$, dann gilt für alle reellen Zahlen $V\ge 0$ und alle ganzen Zahlen $h\ge 0$

$$\sum_{r=1}^{q}\min\Big(V,\frac{1}{\|\alpha(hq+r)\|}\Big)\ll V+q\log q. \tag{6.11}$$

Beweis: Es sei wieder $\alpha=\frac{a}{q}+\frac{\theta}{q^2}$ für ein $\theta\in[-1,1]$. Dann folgt

$$\begin{aligned}\alpha(hq+r)&=ah+\frac{ar}{q}+\frac{\theta h}{q}+\frac{\theta r}{q^2}\\&=ah+\frac{ar}{q}+\frac{\lfloor\theta h\rfloor+\{\theta h\}}{q}+\frac{\theta r}{q^2}\\&=ah+\frac{ar+\lfloor\theta h\rfloor+\delta(r)}{q},\end{aligned}$$

wobei $-1\le\delta(r)=\{\theta h\}+\frac{\theta r}{q}<2$. Zu jedem ganzzahligen $r\in[1,q]$ gibt es ein eindeutiges ganzzahliges r', sodass

$$\{\alpha(hq+r)\}=\frac{ar+\lfloor\theta h\rfloor+\delta(r)}{q}-r'.$$

Es sei $t\in[0,1-\frac{1}{q}]$ beliebig. Falls

$$t\le\{\alpha(hq+r)\}\le t+\frac{1}{q},$$

dann folgt

$$qt\le ar-qr'+\lfloor\theta h\rfloor+\delta(r)\le qt+1.$$

Damit erhält man dann einerseits

$$ar-qr'\le qt-\lfloor\theta h\rfloor+1-\delta(r)\le qt-\lfloor\theta h\rfloor+2$$

und andererseits

$$ar-qr'\ge qt-\lfloor\theta h\rfloor-\delta(r)>qt-\lfloor\theta h\rfloor-2.$$

Also liegt $ar - qr'$ in einem halboffenen Intervall J der Länge 4, nämlich

$$J = (qt - \lfloor \theta h \rfloor - 2, qt - \lfloor \theta h \rfloor + 2].$$

Dieses Intervall enthält genau 4 verschiedene ganze Zahlen. Falls $1 \leq r_1 \leq r_2 \leq q$ und $ar_1 - qr_1' = ar_2 - qr_2'$, dann folgt $ar_1 \equiv ar_2 \bmod q$, damit wie zuvor $r_1 \equiv r_2 \bmod q$ und schließlich $r_1 = r_2$.

Also sind alle $ar - qr'$ paarweise verschieden, und somit gibt es für beliebiges $t \in [0, 1 - \frac{1}{q}]$ höchstens vier verschiedene $r \in [1, q]$, sodass $\{\alpha(hq + r)\} \in [t, t + \frac{1}{q}]$.

Als nächstes bemerken wir, dass $\|aq + r\| \in [t, t + \frac{1}{q}]$ genau dann gilt, wenn entweder $\{aq + r\} \in [t, t + \frac{1}{q}]$ oder $\{aq + r\} \in [t', t' + \frac{1}{q}]$ für $t' = 1 - t - \frac{1}{q}$. Daher gibt es höchstens acht verschiedene $r \in [1, q]$, sodass $\|aq + r\| \in [t, t + \frac{1}{q}]$.

Setzen wir insbesondere $J(s) = [\frac{s}{q}, \frac{s+1}{q}]$, dann gilt $\|aq + r\| \in J(s)$ für höchstens acht verschiedene r.

Falls $\|aq + r\| \in J(0)$, so verwenden wir die Ungleichung

$$\min\left(V, \frac{1}{\|\alpha(hq + r)\|}\right) \leq V,$$

ansonsten (wenn also $\|aq + r\| \in J(s)$ für ein $1 \leq s < q/2$) die Ungleichung

$$\min\left(V, \frac{1}{\|\alpha(hq + r)\|}\right) \leq \frac{1}{\|\alpha(hq + r)\|} \leq \frac{q}{s}.$$

Fassen wir diese Ungleichungen zusammen, so erhalten wir

$$\sum_{r=1}^{q} \min\left(V, \frac{1}{\|\alpha(hq + r)\|}\right) \leq 8V + 8 \sum_{1 \leq s < q/2} \frac{q}{s} \ll V + q \log q.$$

□

Lemma 6.11 Es seien $\alpha \in \mathbb{R}$ und a, q teilerfremde ganze Zahlen mit $q \geq 1$. Wenn $|\alpha - \frac{a}{q}| \leq \frac{1}{q^2}$, dann gilt für alle reellen Zahlen $U \geq 1$ und alle ganzen Zahlen $n \geq 1$

$$\sum_{1 \leq k \leq U} \min\left(\frac{n}{k}, \frac{1}{\|\alpha k\|}\right) \ll \left(\frac{n}{q} + U + q\right) \log(2qU). \tag{6.12}$$

Beweis: Wir schreiben k in der Form $k = hq + r$, wobei $1 \leq r \leq q$ und $0 \leq h < \frac{U}{q}$. Dann gilt

$$\begin{aligned} S &= \sum_{1 \leq k \leq U} \min\left(\frac{n}{k}, \frac{1}{\|\alpha k\|}\right) \\ &\leq \sum_{0 \leq h < U/q} \sum_{1 \leq r \leq q} \min\left(\frac{n}{hq + r}, \frac{1}{\|\alpha(hq + r)\|}\right). \end{aligned}$$

Falls $h = 0$ und $1 \leq r \leq q/2$, dann erhalten wir aus Lemma 6.9

$$\sum_{1 \leq r \leq q/2} \min\Big(\frac{n}{r}, \frac{1}{\|\alpha r\|}\Big) \leq \sum_{1 \leq r \leq q/2} \frac{1}{\|\alpha r\|} \ll q \log q.$$

Die übrigen Terme werden mittels

$$\frac{1}{hq + r} < \frac{2}{(h+1)q}$$

abgeschätzt, was für $h \geq 1$ oder $h = 0$, $r > q/2$ in offensichtlicher Weise folgt. Damit ergibt sich nun

$$S \ll q \log q + \sum_{0 \leq h < U/q} \sum_{1 \leq r \leq q} \min\Big(\frac{n}{(h+1)q}, \frac{1}{\|\alpha(hq + r)\|}\Big).$$

Die innere Summe kann mittels Lemma 6.10 abgeschätzt werden, wobei $V = \frac{n}{(h+1)q}$ gesetzt wird. Dann folgt

$$\begin{aligned} S &\ll q \log q + \sum_{0 \leq h < U/q} \Big(\frac{n}{(h+1)q} + q \log q\Big) \\ &\ll q \log q + \Big(\frac{U}{q} + 1\Big) q \log q + \frac{n}{q} \sum_{0 \leq h < U/q} \frac{1}{h+1} \\ &\ll q \log q + U \log q + q \log q + \frac{n}{q} \log\Big(\frac{U}{q} + 1\Big) \\ &\ll \Big(\frac{n}{q} + U + q\Big) \log(2qU), \end{aligned}$$

da offenbar $\frac{U}{q} + 1 \leq U + q \leq 2qU$ ist. □

Lemma 6.12 Es seien $\alpha \in \mathbb{R}$ und a, q teilerfremde ganze Zahlen mit $q \geq 1$. Wenn $|\alpha - \frac{a}{q}| \leq \frac{1}{q^2}$, dann gilt für alle reellen Zahlen U und n

$$\sum_{1 \leq k \leq U} \min\Big(n, \frac{1}{\|\alpha k\|}\Big) \ll \Big(q + U + n + \frac{Un}{q}\Big) \max(1, \log q). \tag{6.13}$$

Beweis: Der Beweis erfolgt im Wesentlichen wie der vorherige. Es gilt

$$\begin{aligned} S = \sum_{1 \leq k \leq U} \min\Big(n, \frac{1}{\|\alpha k\|}\Big) &\leq \sum_{0 \leq h < U/q} \sum_{1 \leq r \leq q} \min\Big(n, \frac{1}{\|\alpha(hq + r)\|}\Big) \\ &\ll q \log q + \sum_{0 \leq h < U/q} \Big(n + \sum_{1 \leq s < q/2} \frac{q}{s}\Big) \\ &\ll q \log q + \sum_{0 \leq h < U/q} (n + q \log q) \\ &\ll q \log q + \Big(\frac{U}{q} + 1\Big)(n + q \log q) \end{aligned}$$

$$\begin{aligned}
&\ll q\log q + U\log q + n + \frac{Un}{q}\\
&\ll \Big(q + U + n + \frac{Un}{q}\Big)\max(1,\log q).
\end{aligned}$$

□

Im Folgenden wird nun mit $[M, N]$ das Intervall ganzer Zahlen m bezeichnet, für die $M \leq m \leq N$ gilt.

Lemma 6.13 Es seien N_1, N_2, N ganze Zahlen mit $N_1 < N_2$ und $0 \leq N_2 - N_1 \leq N$. Ist $f(n)$ eine reellwertige arithmetische Funktion und $S(f)$ definiert durch

$$S(f) = \sum_{n=N_1+1}^{N_2} e(f(n)), \tag{6.14}$$

so gilt

$$|S(f)|^2 = \sum_{|d|<N} S_d(f), \text{ wobei } S_d(f) = \sum_{n\in I(d)} e(\Delta_d(f)(n)). \tag{6.15}$$

Hierbei ist $I(d)$ ein Intervall aufeinanderfolgender ganzer Zahlen, das in $[N_1 + 1, N_2]$ enthalten ist.

Beweis: Für $d \in \mathbb{Z}$ setzen wir $I(d) = [N_1 + 1 - d, N_2 - d] \cap [N_1 + 1, N_2]$. Dann erhalten wir die folgende Gleichungskette:

$$\begin{aligned}
|S(f)|^2 = S(f)\overline{S(f)} &= \sum_{n=N_1+1}^{N_2} e(f(n)) \sum_{m=N_1+1}^{N_2} \overline{e(f(m))}\\
&= \sum_{n=N_1+1}^{N_2}\sum_{m=N_1+1}^{N_2} e(f(n) - f(m))\\
&= \sum_{n=N_1+1}^{N_2}\sum_{d=N_1+1-n}^{N_2-n} e(f(n+d) - f(n))\\
&= \sum_{n=N_1+1}^{N_2}\sum_{d=N_1+1-n}^{N_2-n} e(\Delta_d(f)(n))\\
&= \sum_{d=-(N_2-N_1-1)}^{N_2-N_1-1}\sum_{n\in I(d)} e(\Delta_d(f)(n))\\
&= \sum_{|d|<N}\sum_{n\in I(d)} e(\Delta_d(f)(n))\\
&= \sum_{|d|<N} S_d(f).
\end{aligned}$$

□

Lemma 6.14 Es seien N_1, N_2, N ganze Zahlen mit $N_1 < N_2$ und $0 \leq N_2 - N_1 \leq N$. Ist $l \geq 1$ eine ganze Zahl, $f(n)$ wieder eine reellwertige arithmetische Funktion und $S(f)$ definiert wie zuvor, so gilt

$$|S(f)|^{2^l} \leq (2N)^{2^l-l-1} \sum_{|d_1|<N} \cdots \sum_{|d_l|<N} S_{d_l,\dots,d_1}(f), \tag{6.16}$$

wobei

$$S_{d_l,\dots,d_1}(f) = \sum_{n \in I(d_l,\dots,d_1)} e(\Delta_{d_l,\dots,d_1}(f)(n)). \tag{6.17}$$

Hierbei ist $I(d_l, \dots, d_1)$ ein Intervall aufeinanderfolgender ganzer Zahlen, das in $[N_1+1, N_2]$ enthalten ist.

Beweis: Durch Induktion nach l. Für $l = 1$ ergibt sich die Behauptung aus dem vorigen Lemma. Für den Induktionsschritt verwenden wir die Cauchy-Schwarz-Ungleichung:

$$\begin{aligned}
|S(f)|^{2^{l+1}} = \left(|S(f)|^{2^l}\right)^2 &\leq \left((2N)^{2^l-l-1} \sum_{|d_1|<N} \cdots \sum_{|d_l|<N} |S_{d_l,\dots,d_1}(f)|\right)^2 \\
&= (2N)^{2^{l+1}-2l-2}\Big(\sum_{|d_1|<N} \cdots \sum_{|d_l|<N} |S_{d_l,\dots,d_1}(f)|\Big)^2 \\
&\leq (2N)^{2^{l+1}-2l-2}(2N)^l \sum_{|d_1|<N} \cdots \sum_{|d_l|<N} |S_{d_l,\dots,d_1}(f)|^2.
\end{aligned}$$

Nach dem vorigen Lemma gibt es ein Intervall

$$I(d_{l+1}, d_l, \dots, d_1) \subseteq I(d_l, \dots, d_1) \subseteq [N_1 + 1, N_2],$$

sodass

$$\begin{aligned}
|S_{d_l,\dots,d_1}(f)|^2 &= \Big|\sum_{n \in I(d_l,\dots,d_1)} e(\Delta_{d_l,\dots,d_1}(f)(n))\Big|^2 \\
&= \sum_{|d_{l+1}|<N} \sum_{n \in I(d_{l+1},d_l,\dots,d_1)} e(\Delta_{d_{l+1},d_l,\dots,d_1}(f)(n)) \\
&= \sum_{|d_{l+1}|<N} S_{d_{l+1},d_l,\dots,d_1}(f).
\end{aligned}$$

Durch Einsetzen in obigen Ausdruck ergibt sich das Gewünschte. $\square$

Lemma 6.15 Es sei $k \geq 1$, $K = 2^{k-1}$ und $\varepsilon > 0$. Weiters sei $f(x) = \alpha x^k + \dots$ ein Polynom vom Grad k mir reellen Koeffizienten. Für den Audruck

$$S(f) = \sum_{n=1}^{N} e(f(n)) \tag{6.18}$$

gilt die Abschätzung

$$|S(f)|^K \ll N^{K-1} + N^{K-k+\varepsilon} \sum_{m=1}^{k!N^{k-1}} \min(N, \|m\alpha\|^{-1}), \tag{6.19}$$

wobei die implizierte Konstante nur von k und ε abhängt.

Beweis: Wenden wir das vorherige Lemma mit $l = k-1$ an, so erhalten wir

$$|S(f)|^K \leq (2N)^{K-k} \sum_{|d_1|<N} \cdots \sum_{|d_{k-1}|<N} |S_{d_{k-1},\ldots,d_1}(f)|.$$

Dabei ist

$$S_{d_{k-1},\ldots,d_1}(f) = \sum_{n \in I(d_{k-1},\ldots,d_1)} e(\Delta_{d_{k-1},\ldots,d_1}(f)(n)),$$

und $I(d_{k-1},\ldots,d_1)$ ein Intervall ganzer Zahlen, das in $[1,N]$ enthalten ist. Wegen $|e(t)| = 1$ für alle reellen t folgt unmittelbar die obere Schranke

$$|S_{d_{k-1},\ldots,d_1}(f)| \leq \sum_{n \in I(d_{k-1},\ldots,d_1)} |e(\Delta_{d_{k-1},\ldots,d_1}(f)(n))| \leq N.$$

Nun wissen wir aus Lemma 6.4, dass $(k-1)$-malige Anwendung des Differenzoperators auf f ein lineares Polynom erzeugt, das die Form

$$\Delta_{d_{k-1},\ldots,d_1}(f)(x) = \lambda x + \beta$$

mit $\lambda = d_{k-1}\ldots d_1 k!\alpha$ hat. Es sei nun $I(d_{k-1},\ldots,d_1) = [N_1+1, N_2]$ für gewisse $0 \leq N_1 \leq N_2 \leq N$. Nach Lemma 6.8 haben wir dann

$$\begin{aligned}
|S_{d_{k-1},\ldots,d_1}(f)| &= \Big| \sum_{n \in I(d_{k-1},\ldots,d_1)} e(\Delta_{d_{k-1},\ldots,d_1}(f)(n)) \Big| \\
&= \Big| \sum_{n=N_1+1}^{N_2} e(\lambda n + \beta) \Big| = \Big| \sum_{n=N_1+1}^{N_2} e(\lambda n) \Big| \\
&\leq \frac{1}{\|\lambda\|} = \frac{1}{\|d_{k-1}\ldots d_1 k!\alpha\|}.
\end{aligned}$$

Damit folgt zunächst

$$|S_{d_{k-1},\ldots,d_1}(f)| \leq \min(N, \|d_{k-1}\ldots d_1 k!\alpha\|^{-1})$$

und durch Einsetzen

$$\begin{aligned}
|S(f)|^K &\leq (2N)^{K-k} \sum_{|d_1|<N} \cdots \sum_{|d_{k-1}|<N} |S_{d_{k-1},\ldots,d_1}(f)| \\
&\leq (2N)^{K-k} \sum_{|d_1|<N} \cdots \sum_{|d_{k-1}|<N} \min(N, \|d_{k-1}\ldots d_1 k!\alpha\|^{-1}).
\end{aligned}$$

Es gibt höchstens $(k-1)(2N)^{k-2}$ verschiedene $(k-1)$-tupel $d_{k-1}, \ldots, d_1$, für die das Produkt $d_{k-1} \ldots d_1 k! \alpha$ gleich 0 ist. Für diese verwenden wir die Abschätzung durch N und erhalten weiter

$$\begin{aligned} |S(f)|^K &\leq (2N)^{K-k}(k-1)(2N)^{k-2}N + (2N)^{K-k} \sum_{1 \leq |d_1| < N} \cdots \sum_{1 \leq |d_{k-1}| < N} \min(N, \|d_{k-1} \ldots d_1 k! \alpha\|^{-1}) \\ &\leq k(2N)^{K-1} + 2^{K-1}N^{K-k} \sum_{1 \leq |d_1| < N} \cdots \sum_{1 \leq |d_{k-1}| < N} \min(N, \|d_{k-1} \ldots d_1 k! \alpha\|^{-1}) \\ &\ll N^{K-1} + N^{K-k} \sum_{1 \leq |d_1| < N} \cdots \sum_{1 \leq |d_{k-1}| < N} \min(N, \|d_{k-1} \ldots d_1 k! \alpha\|^{-1}), \end{aligned}$$

wobei die Konstante nur von k abhängt. Darüber hinaus gilt stets $1 \leq |d_{k-1} \ldots d_1 k!| \leq k! N^{k-1}$. Wenn gezeigt werden kann, dass die Anzahl der Darstellungen von $1 \leq m \leq k! N^{k-1}$ in der Form $|d_{k-1} \ldots d_1 k!|$ asymptotisch durch $m^{\varepsilon/(k-1)} \ll N^{\varepsilon}$ abgeschätzt werden kann, dann folgt

$$\begin{aligned} |S(f)|^K &\ll N^{K-1} + N^{K-k} \sum_{1 \leq |d_1| < N} \cdots \sum_{1 \leq |d_{k-1}| < N} \min(N, \|d_{k-1} \ldots d_1 k! \alpha\|^{-1}) \\ &\ll N^{K-1} + N^{K-k+\varepsilon} \sum_{m=1}^{k! N^{k-1}} \min(N, \|m\alpha\|^{-1}), \end{aligned}$$

was zu beweisen war. Dazu verwenden wir, dass die Anzahl der Teiler $\tau(m)$ einer Zahl m durch $\tau(m) \ll m^{\delta}$ für beliebiges $\delta > 0$ abgeschätzt werden kann (siehe z.B. [13], S.111). Wendet man dies auf $\delta = \varepsilon/(k-1)^2$ an, so erhält man, dass es für jedes d_i nur $\ll m^{\delta}$ Möglichkeiten und damit insgesamt nur höchstens $\ll m^{\delta(k-1)} = m^{\varepsilon/(k-1)}$ mögliche Darstellungen von m in der Form $d_{k-1} \ldots d_1 k!$ gibt. Die implizierte Konstante hängt tatsächlich nur von ε und k ab. $\square$

Satz 6.16 (Ungleichung von Weyl [39]) *Es seien $f(x) = \alpha x^k + \ldots$ ein reelles Polynom vom Grad $k \geq 2$ und a, q teilerfremde ganze Zahlen mit $q \geq 1$, sodass $|\alpha - \frac{a}{q}| \leq \frac{1}{q^2}$. Wenn wie zuvor*

$$S(f) = \sum_{n=1}^{N} e(f(n))$$

ist, $K = 2^{k-1}$ und $\varepsilon > 0$, dann gilt

$$S(f) \ll N^{1+\varepsilon}(N^{-1} + q^{-1} + N^{-k}q)^{1/K}, \tag{6.20}$$

wobei die implizierte Konstante nur von k und ε abhängt.

Beweis: Wegen $|S(f)| \leq N$ ist die Abschätzung trivial, falls $q \geq N^k$. Daher können wir $1 \leq q \leq N^k$ annehmen. Dann folgt $\log q \ll \log N \ll N^{\varepsilon}$. Nach dem vorigen Lemma ergibt sich

$$|S(f)|^K \ll N^{K-1} + N^{K-k+\varepsilon} \sum_{m=1}^{k! N^{k-1}} \min(N, \|m\alpha\|^{-1}).$$

Nach Lemma 6.12 gilt aber

$$\begin{aligned}\sum_{m=1}^{k!N^{k-1}} \min(N, \|m\alpha\|^{-1}) &\ll \Big(q + k!N^{k-1} + N + \frac{k!N^k}{q}\Big) \max(1, \log q) \\ &\ll \Big(q + N^{k-1} + \frac{N^k}{q}\Big) \log N \\ &\ll N^k \Big(qN^{-k} + N^{-1} + q^{-1}\Big) N^{\varepsilon}.\end{aligned}$$

Kombiniert ergibt dies

$$\begin{aligned}|S(f)|^K &\ll N^{K-1} + N^{K+\varepsilon}\Big(qN^{-k} + N^{-1} + q^{-1}\Big) \\ &\ll N^{K+\varepsilon}\Big(qN^{-k} + N^{-1} + q^{-1}\Big).\end{aligned}$$

Durch Ziehen der K-ten Wurzel und Ersetzen von $\frac{\varepsilon}{K}$ durch ε folgt der Satz. □

Korollar 6.17

$$S(q, a) = \sum_{x=1}^{q} e(ax^k/q) \ll q^{1-1/K+\varepsilon}. \tag{6.21}$$

Beweis: Wende die Ungleichung von Weyl auf das Polynom $f(x) = ax^k/q$ und $N = q$ an. Dann folgt unmittelbar

$$S(q, a) \ll q^{1+\varepsilon}(q^{-1} + q^{-k+1})^{1/K} \ll q^{1-1/K+\varepsilon}.$$

□

Korollar 6.18 Es sei $k \geq 2$. Dann gibt es ein $\delta > 0$, sodass für beliebiges $N \geq 2$ und beliebige teilerfremde ganze Zahlen a, q mit $N^{1/2} \leq q \leq N^{k-1/2}$ gilt:

$$\sum_{n=1}^{N} e(\alpha n^k/q) \ll N^{1-\delta}. \tag{6.22}$$

Beweis: Wende die Ungleichung von Weyl auf das Polynom $f(x) = ax^k/q$ an. Dann folgt unmittelbar

$$\begin{aligned}S(f) &\ll N^{1+\varepsilon}(N^{-1} + q^{-1} + N^{-k}q)^{1/K} \\ &\leq N^{1+\varepsilon}(N^{-1} + N^{-1/2} + N^{-1/2})^{1/K} \\ &\ll N^{1-1/(2K)+\varepsilon} \leq N^{1-\delta}\end{aligned}$$

für beliebig gewähltes $\delta < 1/(2K)$. □

Satz 6.19 (Lemma von Hua [11]) *Für $k \geq 2$ sei*

$$T(\alpha) = \sum_{n=1}^{N} e(\alpha n^k). \tag{6.23}$$

Dann gilt

$$\int_0^1 |T(\alpha)|^{2^k} \, d\alpha \ll N^{2^k - k + \varepsilon}. \tag{6.24}$$

Beweis: Wir zeigen mittels Induktion, dass sogar für $1 \leq j \leq k$ die Abschätzung

$$\int_0^1 |T(\alpha)|^{2^j} \, d\alpha \ll N^{2^j - j + \varepsilon}$$

gilt. Zunächst behandeln wir den Fall $j = 1$. Hier gilt

$$\int_0^1 |T(\alpha)|^2 \, d\alpha = \sum_{m=1}^{N} \sum_{n=1}^{N} \int_0^1 e(\alpha(m^k - n^k)) \, d\alpha = N,$$

da nur die Glieder mit $m = n$ etwas (jeweils 1) zur Summe beitragen, denn $\int_0^1 e(\alpha t) \, dt$ ist 0, falls $t \neq 0$, und 1, falls $t = 0$. Für den Induktionsschritt verwenden wir Lemma 6.2:

$$\Delta_{d_j, \ldots, d_1}(f)(x) = \alpha d_j \ldots d_1 p_{k-j}(x)$$

für ein ganzzahliges Polynom p_{k-j} vom Grad $k - j$, sofern die d_i ganze Zahlen sind. Wenden wir weiters Lemma 6.14 mit $N_1 = 0$, $N_2 = N$ und $f(x) = \alpha x^k$ (und somit $S(f) = T(\alpha)$) an, so erhalten wir

$$\begin{aligned} |T(\alpha)|^{2^j} &\leq (2N)^{2^j - j - 1} \sum_{|d_1| < N} \cdots \sum_{|d_l| < N} \sum_{n \in I(d_l, \ldots, d_1)} e(\Delta_{d_l, \ldots, d_1}(f)(n)) \\ &= (2N)^{2^j - j - 1} \sum_{|d_1| < N} \cdots \sum_{|d_l| < N} \sum_{n \in I(d_l, \ldots, d_1)} e(\alpha d_j \ldots d_1 p_{k-j}(n)), \end{aligned}$$

wobei $I(d_l, \ldots, d_1)$ ein Intervall aufeinanderfolgender ganzer Zahlen aus $[1, N]$ ist. Damit folgt dann

$$|T(\alpha)|^{2^j} \leq (2N)^{2^j - j - 1} \sum_d r(d) e(\alpha d),$$

wobei $r(d)$ die Anzahl der Faktorisierungen von d in der Form

$$d_j \ldots d_1 p_{k-j}(n)$$

mit $|d_i| < N$ und $n \in I(d_j, \ldots, d_1)$ ist. Wegen $d \ll N^k$ (Lemma 6.5) und der Tatsache, dass $p_{k-j}(x)$ für jedes y höchstens $k - j$ y-Stellen hat, ergibt sich wie im Beweis von Lemma 6.15

$$r(d) \ll |d|^{\varepsilon/k} \ll N^{\varepsilon}$$

für beliebiges ε, sofern $d \neq 0$. Da p_{n-k} nur höchstens $k-j$ Nullstellen hat, kann man $r(0)$ durch

$$r(0) \ll N^j$$

abschätzen.

Andererseits schreiben wir $|T(\alpha)|^{2^j}$ in folgender Weise um:

$$\begin{aligned}
|T(\alpha)|^{2^j} &= T(\alpha)^{2^{j-1}} T(-\alpha)^{2^{j-1}} \\
&= \Big(\sum_{x=1}^{N} e(\alpha x^k)\Big)^{2^j-1} \Big(\sum_{x=1}^{N} e(-\alpha x^k)\Big)^{2^j-1} \\
&= \sum_{x_1=1}^{N} \cdots \sum_{x_{2^{j-1}}=1}^{N} \sum_{y_1=1}^{N} \cdots \sum_{y_{2^{j-1}}=1}^{N} e\Big(\alpha\Big(\sum_{i=1}^{2^{j-1}} x_i^k - \sum_{i=1}^{2^{j-1}} y_i^k\Big)\Big) \\
&= \sum_d s(d) e(-\alpha d),
\end{aligned}$$

wobei $s(d)$ die Zahl der Darstellungen von d in der Form

$$\sum_{i=1}^{2^{j-1}} y_i^k - \sum_{i=1}^{2^{j-1}} x_i^k$$

ist. Dann gilt zunächst

$$\sum_d s(d) = |T(0)|^{2^j} = N^{2^j}$$

und nach Induktionsvoraussetzung

$$s(0) = \int_0^1 |T(\alpha)|^{2^j} \, d\alpha \ll N^{2^j - j + \varepsilon}.$$

Fügt man die beiden Abschätzungen für $|T(\alpha)|^{2^j}$ zusammen, so erhält man

$$\begin{aligned}
\int_0^1 |T(\alpha)|^{2^{j+1}} \, d\alpha &= \int_0^1 |T(\alpha)|^{2^j} |T(\alpha)|^{2^j} \, d\alpha \\
&\ll N^{2^j - j - 1} \int_0^1 \sum_{d'} r(d') e(\alpha d') \sum_d s(d) e(-\alpha d) \, d\alpha \\
&= N^{2^j - j - 1} \sum_d r(d) s(d) \\
&= N^{2^j - j - 1} \Big(r(0)s(0) + \sum_{d \neq 0} r(d) s(d) \Big) \\
&\ll N^{2^j - j - 1} N^j N^{2^j - j + \varepsilon} + N^{2^j - j - 1} N^{\varepsilon} \sum_{d \neq 0} s(d) \\
&\ll N^{2^{j+1} - (j+1) + \varepsilon} + N^{2^j - j - 1 + \varepsilon} N^{2^j} \ll N^{2^{j+1} - (j+1) + \varepsilon}.
\end{aligned}$$

□

Kapitel 7

Die asymptotische Formel von Hardy und Littlewood

Nun kommen wir zum eigentlichen Kern des Beweises von Hardy und Littlewood, in dem die "circle method" zur Anwendung kommt. Es sei $r_{k,s}(N)$ die Zahl der Darstellungen von N als Summe von s k-ten Potenzen. Im Gegensatz zum Partitionsproblem soll es hierbei auf die Reihenfolge ankommen. Wenn gezeigt werden kann, dass $r_{k,s}(N)$ für hinreichend großes N stets positiv ist, wenn s geeignet gewählt wird, dann folgt, dass jede natürliche Zahl als Summe einer beschränkten Anzahl von k-ten Potenzen geschrieben werden kann.

Im ursprünglichen Beweis verwendeten Hardy und Littlewood die erzeugende Funktion

$$f(z) = \sum_{n \geq 1} z^{n^k} \tag{7.1}$$

für $r_{k,1}(N)$. Die erzeugende Funktion für $r_{k,s}(N)$ ist dann $f(z)^s$, und wie schon für die Partitionsfunktion erhält man dann $r_{k,s}(N)$ aus der Cauchy'schen Integralformel mittels

$$r_{k,s}(N) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{f(z)^s}{z^{N+1}}\, dz \tag{7.2}$$

für einen geeigneten Integrationspfad $\mathcal{C}$ um 0. Dieses Integral muss nun abgeschätzt werden.

Dieser ursprüngliche Ansatz wurde von Vinogradov vereinfacht, indem er anstelle der erzeugenden Funktion f nur die abgeschnittene Reihe

$$p(z) = \sum_{n \geq 1, n^k \leq N} z^{n^k} \tag{7.3}$$

verwendete. Für $m \leq N$ stimmen die m-ten Koeffizienten von $p(z)^s$ mit denen von $f(z)^s$ überein, da alle Summanden in einer Darstellung von $m \leq N$ ja $\leq N$ sein müssen.

Nunmehr ersetzt man z durch $e(\alpha)$ und erhält ein trigonometrisches Polynom $F(\alpha) = p(e(\alpha))$. Damit wird aus dem obigen Kurvenintegral aber ein Integral über dem Einheitsintervall. Ist $r_{k,s}^{(N)}(m)$ die Anzahl der Darstellungen von m durch s k-te Potenzen $\leq N$, so folgt

$$F(\alpha)^s = \sum_{m=0}^{sN} r_{k,s}^{(N)}(m)e(m\alpha) \tag{7.4}$$

und damit aufgrund der Orthogonalitätsrelation für $e(n\alpha)$, nämlich

$$\int_0^1 e(m\alpha)e(-n\alpha)\,d\alpha = \delta_{mn}, \tag{7.5}$$

wobei δ_{mn} das Kronecker-Delta bezeichnet,

$$r_{k,s}(m) = r_{k,s}^{(N)}(m) = \int_0^1 F(\alpha)^s e(-m\alpha)\,d\alpha \tag{7.6}$$

für $m \leq N$ und insbesondere für $m = N$. Das Integral über $[0,1]$ wird nun in zwei Teile geteilt, in die "major arcs" $\mathfrak{M}$ und in die "minor arcs" $\mathfrak{m}$. Letztere nehmen zwar einen weitaus größeren Anteil des Einheitsintervalls ein, das Integral über $\mathfrak{m}$ ist jedoch vernachlässigbar, wie sich herausstellt.

Definition 7.1 Es sei $k \geq 2$, $N > 2^k$ und damit $P := \lfloor N^{1/k} \rfloor \geq 2$. Wähle ein beliebiges $\nu \in (0, 1/5)$. Für $1 \leq q \leq P^\nu$ und $0 \leq a \leq q$ mit $(a,q) = 1$ setze

$$\mathfrak{M}(q,a) = \Big\{\alpha \in [0,1] : \Big|\alpha - \frac{a}{q}\Big| \leq \frac{1}{P^{k-\nu}}\Big\}. \tag{7.7}$$

Die Menge $\mathfrak{M}$ der "major arcs" ist dann definiert durch

$$\bigcup_{1\leq q\leq P^\nu} \bigcup_{\substack{a=0\\(a,q)=1}}^{q} \mathfrak{M}(q,a).$$

Die "minor arcs" bilden das Komplement in $[0,1]$, d.h. $\mathfrak{m} = [0,1] \setminus \mathfrak{M}$.

Die "major arcs" enthalten jene reellen Zahlen, die in gewissem Sinne nahe an rationalen Zahlen liegen. Die $\mathfrak{M}(q,a)$ sind paarweise disjunkt, denn aus $\alpha \in \mathfrak{M}(q,a) \cap \mathfrak{M}(q',a')$ mit $\frac{a}{q} \neq \frac{a'}{q'}$ folgt

$$\frac{1}{P^{2\nu}} \leq \frac{1}{qq'} \leq \Big|\frac{a}{q} - \frac{a'}{q'}\Big| \leq \Big|\alpha - \frac{a}{q}\Big| + \Big|\alpha - \frac{a'}{q'}\Big| \leq \frac{2}{P^{k-\nu}},$$

was wegen $P \geq 2$ und $k \geq 2$ ausgeschlossen ist.

Offenbar gilt

$$\mathfrak{M}(1,0) = \Big[0, \frac{1}{P^{k-\nu}}\Big],\ \mathfrak{M}(1,1) = \Big[1 - \frac{1}{P^{k-\nu}}, 1\Big] \text{ und } \mathfrak{M}(q,a) = \Big[\frac{a}{q} - \frac{1}{P^{k-\nu}}, \frac{a}{q} + \frac{1}{P^{k-\nu}}\Big]$$

für $q \geq 2$. Daher ist das Lebesgue-Maß von $\mathfrak{M}$

$$\lambda(\mathfrak{M}) = \frac{2}{P^{k-\nu}} \sum_{1 \leq q \leq P^\nu} \varphi(q) \leq \frac{2}{P^{k-\nu}} \sum_{1 \leq q \leq P^\nu} q = \frac{2}{P^{k-\nu}} \frac{P^\nu(P^\nu+1)}{2} \leq \frac{2}{P^{k-3\nu}}, \tag{7.8}$$

was für $P \to \infty$ nach 0 strebt.

7.1 Die "minor arcs"

Proposition 7.1 Es sei $k \geq 2$ und $s > 2^k$. Dann gibt es ein $\delta_1 > 0$, sodass

$$\int_{\mathfrak{m}} F(\alpha)^s e(-N\alpha)\, d\alpha = O(P^{s-k-\delta_1}), \tag{7.9}$$

wobei die implizierte Konstante nur von s und k abhängt.

Beweis: Nach dem Satz von Dirichlet (Lemma 6.6), angewandt auf $Q = P^{k-\nu}$, gibt es zu jedem reellen α eine rationale Zahl $\frac{a}{q}$ (a, q teilerfremd) mit $1 \leq q \leq P^{k-\nu}$ und

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{qP^{k-\nu}} \leq \min\Big(\frac{1}{P^{k-\nu}}, \frac{1}{q^2}\Big). \tag{7.10}$$

Für ein $\alpha \in \mathfrak{m}$ muss insbesondere $\alpha \notin \mathfrak{M}(1,0)$ und $\alpha \notin \mathfrak{M}(1,1)$ gelten, woraus

$$\frac{1}{P^{k-\nu}} < \alpha < 1 - \frac{1}{P^{k-\nu}}$$

folgt. Weiters muss wegen (7.10) $P^\nu < q$ gelten, da sonst α in $\mathfrak{M}(q,a)$ läge. Sei nun $K = 2^{k-1}$. Nach der Ungleichung von Weyl, angewandt auf $f(x) = \alpha x^k$, gilt

$$\begin{aligned} F(\alpha) &\ll P^{1+\varepsilon}\Big(P^{-1} + q^{-1} + P^{-k} q\Big)^{1/K} \\ &\ll P^{1+\varepsilon}\Big(P^{-1} + P^{-\nu} + P^{-k} P^{k-\nu}\Big)^{1/K} \\ &\ll P^{1+\varepsilon-\nu/K}. \end{aligned}$$

Durch Anwendung des Lemmas von Hua erhält man nun

$$\begin{aligned} \Big|\int_{\mathfrak{m}} F(\alpha)^s e(-N\alpha)\, d\alpha\Big| &= \Big|\int_{\mathfrak{m}} F(\alpha)^{s-2^k} F(\alpha)^{2^k} e(-N\alpha)\, d\alpha\Big| \\ &\leq \int_{\mathfrak{m}} |F(\alpha)|^{s-2^k} |F(\alpha)|^{2^k}\, d\alpha \\ &\leq \max_{\alpha \in \mathfrak{m}} |F(\alpha)|^{s-2^k} \int_0^1 |F(\alpha)|^{2^k}\, d\alpha \\ &\ll (P^{1+\varepsilon-\nu/K})^{s-2^k} P^{2^k-k+\varepsilon} \\ &= P^{s-k-\delta_1}, \end{aligned}$$

wobei

$$\delta_1 = \frac{\nu(s-2^k)}{K} - (s-2^k+1)\varepsilon$$

für hinreichend kleines ε positiv ist. □

7.2 Die "major arcs"

Um die Funktion $F(\alpha)$ in der Nähe rationaler Zahlen zu behandeln, werden Hilfsfunktionen eingeführt, die wir zunächst abschätzen wollen. Dazu wird partielle Summation benötigt:

Lemma 7.2 (Partielle (Abelsche) Summation) Es seien $u(n)$ und $f(n)$ arithmetische Funktionen und $U(t) := \sum_{n\le t} u(n)$ für alle reellen t definiert. Für $a, b \in \mathbb{N}$ mit $a < b$ gilt dann

$$\sum_{n=a+1}^{b} u(n)f(n) = U(b)f(b) - U(a)f(a+1) - \sum_{n=a+1}^{b-1} U(n)(f(n+1) - f(n)). \tag{7.11}$$

Ist $f(t)$ auf einem Intervall $[y, x]$ stetig differenzierbar, dann gilt zudem

$$\sum_{y<n\le x} u(n)f(n) = U(x)f(x) - U(y)f(y) - \int_y^x U(t)f'(t)\,dt \tag{7.12}$$

Zum Beweis siehe etwa [9], S.91 f. und [21], S.304 f.

Lemma 7.3 Es sei

$$v(\beta) = \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} e(\beta m) \tag{7.13}$$

und

$$S(q,a) = \sum_{r=1}^{q} e(ar^k/q). \tag{7.14}$$

Dann gilt $S(q,a) \ll q^{1-1/K+\varepsilon}$ sowie $v(\beta) \ll \min(P, |\beta|^{-1/k})$, falls $|\beta| \le \frac{1}{2}$.

Beweis: Die Abschätzung für $S(q,a)$ wurde in Korollar 6.17 gezeigt. Für $v(\beta)$ betrachten wir die Funktion $f(x) = x^{1/k-1}/k$. Sie ist für $x \ge 1$ offensichtlich positiv, stetig und streng monoton fallend. Daher gilt

$$\begin{aligned} v(\beta) &\le \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} = \sum_{m=1}^{N} f(m) \\ &\le \int_1^N f(x)\,dx + f(1) \\ &= N^{1/k} - 1 + \frac{1}{k} < N^{1/k} \ll P. \end{aligned}$$

Für $|\beta| \le 1/N$ gilt $P \le N^{1/k} \le |\beta|^{-1/k}$, und die Abschätzung ist bewiesen. Ist andererseits $1/N < |\beta| \le \frac{1}{2}$ (womit $|\beta|^{1/k} \ll P$ folgt), so sei $M = \lfloor |\beta|^{-1} \rfloor$. Für M haben wir dann die Ungleichung

$$M \le \frac{1}{|\beta|} < M + 1 \le N.$$

Nun setzen wir $U(t) = \sum_{m \le t} e(\beta m)$ und verwenden partielle Summation. Aus Lemma 6.8 wissen wir dabei, dass $U(t) \ll \|\beta\|^{-1} = |\beta|^{-1}$.

$$\begin{aligned}\sum_{m=M+1}^{N} \frac{1}{k} m^{1/k-1} e(\beta m) &= f(N)U(N) - f(M)U(M) - \int_M^N U(t) f'(t)\, dt \\ &\ll \frac{M^{1/k-1}}{|\beta|} \ll |\beta|^{-1/k} \\ &\ll \min(P, |\beta|^{-1/k}).\end{aligned}$$

Damit ergibt sich schließlich auch in diesem Fall

$$\begin{aligned}v(\beta) &= \sum_{m=1}^{M} \frac{1}{k} m^{1/k-1} e(\beta m) + \sum_{m=M+1}^{N} \frac{1}{k} m^{1/k-1} e(\beta m) \\ &\ll M^{1/k} + \min(P, |\beta|^{-1/k}) \ll \min(P, |\beta|^{-1/k}).\end{aligned}$$

□

Lemma 7.4 Es seien q, a ganze Zahlen mit $1 \le q \le P^\nu$ und $0 \le a \le q$. Wenn $\alpha \in \mathfrak{M}(q, a)$, dann gilt

$$F(\alpha) = \frac{S(q, \alpha)}{q} v\Big(\alpha - \frac{a}{q}\Big) + O(P^{2\nu}). \tag{7.15}$$

Beweis: Es sei $\beta = \alpha - \frac{a}{q}$. Wegen $\alpha \in \mathfrak{M}(q, a)$ gilt dann $|\beta| \le P^{\nu - k}$ und damit

$$\begin{aligned}F(\alpha) - \frac{S(q, \alpha)}{q} v(\beta) &= \sum_{m=1}^{P} e(\alpha m^k) - \frac{S(q, \alpha)}{q} \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} e(\beta m) \\ &= \sum_{m=1}^{P} e\Big(\frac{am^k}{q}\Big) e(\beta m^k) - \frac{S(q, \alpha)}{q} \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} e(\beta m) \\ &= \sum_{m=1}^{N} u(m) e(\beta m),\end{aligned}$$

wobei $u(m)$ durch

$$u(m) = \begin{cases} e(am/q) - (S(q, a)/q) k^{-1} m^{1/k-1} & m \text{ ist } k\text{-te Potenz} \\ -(S(q, a)/q) k^{-1} m^{1/k-1} & \text{sonst} \end{cases}$$

definiert ist. Sei nun $y \geq 1$ beliebig. Wegen $|S(a,q)| \leq q$ gilt

$$\begin{aligned}
\sum_{1 \leq m \leq y} e(am^k/q) &= \sum_{r=1}^{q} e(ar^k/q) \sum_{\substack{1 \leq m \leq y \\ m \equiv r \mod q}} 1 \\
&= S(q,a)\Big(\frac{y}{q} + O(1)\Big) \\
&= \frac{yS(q,a)}{q} + O(q).
\end{aligned}$$

Für $t \geq 1$ folgt damit

$$\begin{aligned}
U(t) &= \sum_{1 \leq m \leq t} u(m) \\
&= \sum_{1 \leq m \leq t^{1/k}} e(am^k/q) - \frac{S(q,a)}{q} \sum_{1 \leq m \leq t} \frac{1}{k} m^{1/k-1} \\
&= t^{1/k} \frac{S(q,a)}{q} + O(q) - \frac{S(q,a)}{q}\Big(t^{1/k} + O(1)\Big) = O(q).
\end{aligned}$$

Nun wenden wir auf die Summe $\sum_{m=1}^{N} u(m)e(\beta m)$ partielle Summation an und erhalten

$$\begin{aligned}
\sum_{m=1}^{N} u(m)e(\beta m) &= e(\beta N)u(N) - 2\pi i \beta \int_1^N e(\beta t)U(t)\,dt \\
&= O(q) - 2\pi i\beta \int_1^N e(\beta t)O(q)\,dt \\
&\ll q + |\beta|Nq = q(1 + |\beta|N) \\
&\ll P^\nu(1 + P^{\nu-k}P^k) \ll P^{2\nu}.
\end{aligned}$$

□

Proposition 7.5 Es sei

$$\mathfrak{S}(N,Q) = \sum_{1 \leq q \leq Q} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} \Big(\frac{S(q,a)}{q}\Big)^s e(-Na/q) \tag{7.16}$$

und

$$J^*(N) = \int_{-P^{\nu-k}}^{P^{\nu-k}} v(\beta)^s e(-N\beta)\,d\beta. \tag{7.17}$$

Dann gilt

$$\int_{\mathfrak{M}} F(\alpha)^s e(-N\alpha)\,d\alpha = \mathfrak{S}(N,P^\nu)J^*(N) + O(P^{s-k-\delta_2}), \tag{7.18}$$

wobei $\delta_2 = 1 - 5\nu > 0$.

Beweis: Es sei $\alpha \in \mathfrak{M}(q,a)$ und β wie zuvor. Wir setzen zunächst

$$V = V(\alpha, q, a) = \frac{S(q,a)}{q} v(\beta). \tag{7.19}$$

Wegen $|S(q,a)| \leq q$ und $|v(\beta)| \ll P$ (Lemma 7.3) gilt $|V| \ll P$. Auch für $F = F(\alpha)$ gilt $|F| \leq P$, und da nach dem vorigen Lemma $F - V = O(P^{2\nu})$ ist, folgt

$$\begin{aligned} F^s - V^s &= (F - V)(F^{s-1} + F^{s-2}V + \ldots + V^{s-1}) \\ &\ll P^{2\nu} P^{s-1} = P^{s-1+2\nu}. \end{aligned}$$

Es wurde bereits gezeigt, dass das Lebesgue-Maß von $\mathfrak{M}$ durch $\lambda(\mathfrak{M}) \ll P^{3\nu-k}$ abgeschätzt werden kann. Daher folgt

$$\int_{\mathfrak{M}} |F^s - V^s| \, d\alpha \ll P^{3\nu-k} P^{s-1+2\nu} = P^{s-k-\delta_2}$$

mit $\delta_2 = 1 - 5\nu > 0$, da zu Beginn $\nu < \frac{1}{5}$ gewählt wurde. Wir ersetzen nun also F^s im Integranden durch V^s, wobei nur ein Fehlerterm der Größenordnung $P^{s-k-\delta_2}$ entsteht:

$$\begin{aligned} \int_{\mathfrak{M}} F(\alpha)^s e(-N\alpha) \, d\alpha &= \int_{\mathfrak{M}} V(\alpha, q, a)^s e(-N\alpha) \, d\alpha + O(P^{s-k-\delta_2}) \\ &= \sum_{1 \leq q \leq P^\nu} \sum_{\substack{a=0 \\ (a,q)=1}}^{q} \int_{\mathfrak{M}(q,a)} V(\alpha, q, a)^s e(-N\alpha) \, d\alpha + O(P^{s-k-\delta_2}). \end{aligned}$$

Für $q \geq 2$ haben wir

$$\begin{aligned} \int_{\mathfrak{M}(q,a)} V(\alpha, q, a)^s e(-N\alpha) \, d\alpha &= \int_{a/q-P^{\nu-k}}^{a/q+P^{\nu-k}} V(\alpha, q, a)^s e(-N\alpha) \, d\alpha \\ &= \int_{-P^{\nu-k}}^{P^{\nu-k}} V(\beta + a/q, q, a)^s e(-N(\beta + a/q)) \, d\beta \\ &= \Big(\frac{S(q,a)}{q}\Big)^s e(-Na/q) \int_{-P^{\nu-k}}^{P^{\nu-k}} v(\beta)^s e(-N\beta) \, d\beta \\ &= \Big(\frac{S(q,a)}{q}\Big)^s e(-Na/q) J^*(N). \end{aligned}$$

Für $q = 1$ dagegen ist $V(\alpha, 1, 0) = v(\alpha)$ und $V(\alpha, 1, 1) = v(\alpha - 1)$. Daher ergibt sich hier

$$\begin{aligned} &\int_{\mathfrak{M}(1,0)} V(\alpha, q, a)^s e(-N\alpha) \, d\alpha + \int_{\mathfrak{M}(1,1)} V(\alpha, q, a)^s e(-N\alpha) \, d\alpha \\ &\qquad = \int_0^{P^{\nu-k}} v(\alpha)^s e(-N\alpha) \, d\alpha + \int_{1-P^{\nu-k}}^{1} v(\alpha - 1)^s e(-N\alpha) \, d\alpha \\ &\qquad = \int_0^{P^{\nu-k}} v(\beta)^s e(-N\beta) \, d\beta + \int_{-P^{\nu-k}}^{0} v(\beta)^s e(-N\beta) \, d\beta = J^*(N). \end{aligned}$$

Durch Zusammenfassen dieser Gleichungen erhält man das Gewünschte. □

7.3 Das singuläre Integral

Das Integral

$$J(N) := \int_{-1/2}^{1/2} v(\beta)^s e(-\beta N)\, d\beta \tag{7.20}$$

wird *singuläres Integral* (für das Problem von Waring) genannt.

Es unterscheidet sich nur im Integrationsbereich von $J^*(N)$, das in Proposition 7.5 vorgekommen ist. Wie sich herausstellt, führt dieser Unterschied aber nur auf einen vernachlässigbaren Fehlerterm, sodass man durch die Betrachtung von $J(N)$ auf die Asymptotik von $J^*(N)$ kommt.

Proposition 7.6 Es gibt ein $\delta_3 > 0$, sodass $J^*(N) = J(N) + O(P^{s-k-\delta_3})$. Darüber hinaus gilt $J(N) \ll P^{s-k}$.

Beweis: Nach Lemma 7.3 gilt

$$\begin{aligned}
J(N) &\ll \int_0^{1/2} \min(P, |\beta|^{-1/k})^s \, d\beta \\
&= \int_0^{1/N} \min(P, |\beta|^{-1/k})^s \, d\beta + \int_{1/N}^{1/2} \min(P, |\beta|^{-1/k})^s \, d\beta \\
&= \int_0^{1/N} P^s \, d\beta + \int_{1/N}^{1/2} \beta^{-s/k} \, d\beta \\
&= \frac{P^s}{N} + \frac{1}{1-\frac{s}{k}} \left(2^{s/k-1} - N^{s/k-1} \right) \\
&\ll P^{s-k}
\end{aligned}$$

sowie

$$\begin{aligned}
J(N) - J^*(N) &= \int_{P^{\nu-k} \le |\beta| \le 1/2} v(\beta)^s e(-N\beta)\, d\beta \\
&\ll \int_{P^{\nu-k}}^{1/2} \beta^{-s/k}\, d\beta \\
&\ll P^{(\nu-k)(1-s/k)} = P^{s-k-\delta_3}
\end{aligned}$$

mit $\delta_3 = \nu(s/k-1) > 0$. □

Lemma 7.7 Es seien α und β reelle Zahlen mit $0 < \beta < 1$ und $\alpha \ge \beta$. Dann gilt

$$\sum_{m=1}^{N-1} m^{\beta-1}(N-m)^{\alpha-1} = N^{\alpha+\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} + O(N^{\alpha-1}), \tag{7.21}$$

wobei die implizierte Konstante nur von β abhängt.

Beweis: Die Funktion $g(x) = x^{\beta-1}(N-x)^{\alpha-1}$ ist positiv und stetig auf dem Intervall $(0, N)$, und es gilt

$$\begin{aligned}\int_0^N g(x)\,dx &= \int_0^N x^{\beta-1}(N-x)^{\alpha-1}\,dx = N^{\alpha+\beta-1}\int_0^1 t^{\beta-1}(1-t)^{\alpha-1}\,dt \\ &= N^{\alpha+\beta-1}B(\alpha,\beta) = N^{\alpha+\beta-1}\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},\end{aligned}$$

wobei $B(\alpha, \beta)$ die Beta-Funktion bezeichnet. Für $\alpha \geq 1$ ist

$$g'(x) = g(x)\Big(\frac{\beta-1}{x} - \frac{\alpha-1}{N-x}\Big) < 0,$$

d.h. $g(x)$ ist streng monoton fallend. Damit gilt dann aber

$$\int_1^N g(x)\,dx < \sum_{m=1}^{N-1} g(m) < \int_0^{N-1} g(x)\,dx$$

und in weiterer Folge

$$\begin{aligned}0 &< \int_0^N g(x)\,dx - \sum_{m=1}^{N-1} g(m) < \int_0^1 g(x)\,dx \\ &= \int_0^1 x^{\beta-1}(N-x)^{\alpha-1}\,dx \leq N^{\alpha-1}\int_0^1 x^{\beta-1}\,dx = \frac{N^{\alpha-1}}{\beta}.\end{aligned}$$

Ist andererseits $0 < \beta \leq \alpha < 1$, dann folgt $0 < \alpha + \beta < 2$, und $g(x)$ hat ein lokales Minimum bei $c = \frac{(1-\beta)N}{2-\alpha-\beta} \in [N/2, N)$. $g(x)$ ist nun auf $(0, c)$ streng monoton fallend und auf (c, N) streng monoton wachsend. Es folgen die Ungleichungen

$$\sum_{m=1}^{\lfloor c\rfloor} g(m) < \int_0^c g(x)\,dx$$

und

$$\sum_{m=1}^{\lfloor c\rfloor} g(m) > \int_1^c g(x)\,dx > \int_0^c g(x)\,dx - \frac{N^{\alpha-1}}{\beta}$$

sowie

$$\sum_{m=\lfloor c\rfloor+1}^{N-1} g(m) < \int_c^N g(x)\,dx$$

und

$$\sum_{m=\lfloor c\rfloor+1}^{N-1} g(m) > \int_c^{N-1} g(x)\,dx > \int_c^N g(x)\,dx - \frac{N^{\beta-1}}{\alpha}.$$

Damit folgt aber

$$0 < \int_0^N g(x)\,dx - \sum_{m=1}^{N-1} g(m) < \frac{N^{\alpha-1}}{\beta} + \frac{N^{\beta-1}}{\alpha} \le \frac{2N^{\alpha-1}}{\beta}.$$

Durch Kombination der Resultate ergibt sich die Behauptung. □

Proposition 7.8 Für $s \ge 2$ gilt

$$J(N) = \Gamma\Big(1+\frac{1}{k}\Big)^s \Gamma\Big(\frac{s}{k}\Big)^{-1} N^{s/k-1} + O(N^{(s-1)/k-1}). \tag{7.22}$$

Beweis: Wir schreiben $J_s(N)$ für $J(N)$, um die Abhängigkeit von s zu verdeutlichen. Der Beweis folgt nun durch Induktion nach s. Aus

$$v(\beta) = \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} e(\beta m)$$

folgt

$$v(\beta)^s = k^{-s} \sum_{m_1=1}^{N} \dots \sum_{m_s=1}^{N} (m_1 \dots m_s)^{1/k-1} e((m_1 + \dots + m_s)\beta)$$

und damit

$$\begin{aligned} J_s(N) &= k^{-s} \sum_{m_1=1}^{N} \dots \sum_{m_s=1}^{N} (m_1 \dots m_s)^{1/k-1} \int_{-1/2}^{1/2} e((m_1 + \dots + m_s - N)\beta)\,d\beta \\ &= k^{-s} \sum_{\substack{1 \le m_i \le N \\ m_1+\dots+m_s=N}} (m_1 \dots m_s)^{1/k-1}. \end{aligned}$$

Für $s = 2$ verwenden wir Lemma 7.7 mit $\alpha = \beta = 1/k$ und erhalten

$$\begin{aligned} J_2(N) &= k^{-2} \sum_{m=1}^{N-1} m^{1/k-1} (N-m)^{1/k-1} \\ &= \frac{(1/k)^2 \Gamma(1/k)^2}{\Gamma(2/k)} N^{2/k-1} + O(N^{1/k-1}) \\ &= \frac{\Gamma(1+1/k)^2}{\Gamma(2/k)} N^{2/k-1} + O(N^{1/k-1}). \end{aligned}$$

Für den Induktionsschritt formt man folgendermaßen um:

$$\begin{aligned}
J_{s+1}(N) &= \int_{-1/2}^{1/2} v(\beta)^{s+1} e(-N\beta)\, d\beta \\
&= \int_{-1/2}^{1/2} v(\beta) v(\beta)^{s} e(-N\beta)\, d\beta \\
&= \int_{-1/2}^{1/2} \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} e(\beta m) v(\beta)^{s} e(-N\beta)\, d\beta \\
&= \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} \int_{-1/2}^{1/2} v(\beta)^{s} e(-(N-m)\beta)\, d\beta \\
&= \sum_{m=1}^{N} \frac{1}{k} m^{1/k-1} J_s(N-m) \\
&= \frac{\Gamma(1+1/k)^s}{\Gamma(s/k)} \sum_{m=1}^{N-1} \frac{1}{k} m^{1/k-1} (N-m)^{s/k-1} + O\Big(\sum_{m=1}^{N-1} \frac{1}{k} m^{1/k-1} (N-m)^{(s-1)/k-1} \Big).
\end{aligned}$$

Wendet man Lemma 7.7 auf beide Summenausdrücke an (mit $\alpha = s/k$, $\beta = 1/k$ bzw. $\alpha = (s-1)/k$, $\beta = 1/k$), dann ergibt sich

$$\sum_{m=1}^{N-1} \frac{1}{k} m^{1/k-1} (N-m)^{s/k-1} = \frac{(1/k)\Gamma(1/k)\Gamma(s/k)}{\Gamma((s+1)/k)} N^{(s+1)/k-1} + O(N^{s/k-1})$$

beziehungsweise

$$\sum_{m=1}^{N-1} \frac{1}{k} m^{1/k-1} (N-m)^{(s-1)/k-1} = O(N^{s/k-1}).$$

Zusammenfassend erhält man

$$\begin{aligned}
J_{s+1}(N) &= \frac{(1/k)\Gamma(1/k)\Gamma(s/k)}{\Gamma((s+1)/k)} \frac{\Gamma(1+1/k)^s}{\Gamma(s/k)} N^{(s+1)/k-1} + O(N^{s/k-1}) \\
&= \frac{\Gamma(1+1/k)^{s+1}}{\Gamma((s+1)/k)} N^{(s+1)/k-1} + O(N^{s/k-1}).
\end{aligned}$$

□

7.4 Die singuläre Reihe

Neben dem Integralteil, der im letzten Abschnitt näher behandelt wurde, kam in der Zerlegung von Proposition 7.5 die Funktion $\mathfrak{S}(N,Q)$ vor, die durch

$$\mathfrak{S}(N,Q) = \sum_{1 \le q \le Q} A_N(q) \tag{7.23}$$

mit

$$A_N(q) = \sum_{\substack{a=1\\(a,q)=1}}^{q} \Big(\frac{S(q,a)}{q}\Big)^s e(-Na/q) \tag{7.24}$$

definiert war. Wir ersetzen dies nun durch die Reihe

$$\mathfrak{S}(N) = \sum_{q=1}^{\infty} A_N(q), \tag{7.25}$$

die sogenannte *singuläre Reihe*. Wählt man $0 < \varepsilon < 1/(sK)$, dann gilt für $s \geq 2^k + 1 = 2K + 1$

$$\frac{s}{K} - 1 - s\varepsilon \geq 1 + \frac{1}{K} - s\varepsilon = 1 + \delta_4$$

für $\delta_4 = 1/K - s\varepsilon > 0$. Aus Lemma 7.3 wissen wir, dass $S(q,a) \ll q^{1-1/K+\varepsilon}$ und damit

$$|A_N(q)| \leq q\Big(\frac{S(q,a)}{q}\Big)^s \ll q^{1-s/K+s\varepsilon} \leq \frac{1}{q^{1+\delta_4}}$$

gilt. $\mathfrak{S}(N)$ kann somit mittels $\zeta(1+\delta_4)$ abgeschätzt werden und konvergiert daher absolut und gleichmäßig bezüglich N. Es gibt somit eine Konstante c_2 (die nur von k und s abhängt), sodass $|\mathfrak{S}(N)| < c_2$ für alle N. Weiters zeigt sich, dass der Fehler, der beim Ersetzen von $\mathfrak{S}(N, P^\nu)$ durch $\mathfrak{S}(N)$ entsteht, klein ist, denn es gilt

$$\mathfrak{S}(N) - \mathfrak{S}(N, P^\nu) = \sum_{q > P^\nu} A_n(q) \ll \sum_{q>P^\nu} \frac{1}{q^{1+\delta_4}} \ll P^{-\nu\delta_4}. \tag{7.26}$$

Der letzte Schritt in der Herleitung der Formel von Hardy und Littlewood besteht darin, die Existenz einer unteren Schranke $c_1 > 0$ für $\mathfrak{S}(N)$ herzuleiten. Dies geschieht mit Methoden der elementaren Zahlentheorie.

Lemma 7.9 Es seien q und r teilerfremd. Dann gilt

$$S(qr, ar + bq) = S(q,a)S(r,b). \tag{7.27}$$

Beweis: Da q und r als teilerfremd vorausgesetzt wurden, sind die Mengen $\{xr : 1 \leq x \leq q\}$ und $\{yq : 1 \leq y \leq r\}$ vollständige Restsysteme modulo q bzw. r. Jede Kongruenzklasse bezüglich qr kann nach dem Chinesischen Restsatz eindeutig in der Form $xr + yq$ geschrieben werden, wobei $1 \leq x \leq q$ und $1 \leq y \leq r$ gilt. Es folgt damit

$$\begin{aligned} S(qr, ar+bq) &= \sum_{m=1}^{qr} e\Big(\frac{(ar+bq)m^k}{qr}\Big) \\ &= \sum_{x=1}^{q}\sum_{y=1}^{r} e\Big(\frac{(ar+bq)(xr+yq)^k}{qr}\Big) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{x=1}^{q}\sum_{y=1}^{r} e\Big(\frac{ar+bq}{qr}\sum_{l=0}^{k}\binom{k}{l}(xr)^l(yq)^{k-l}\Big) \\
&= \sum_{x=1}^{q}\sum_{y=1}^{r} e\Big(\frac{ar+bq}{qr}((xr)^k+(yl)^k)\Big) \\
&= \sum_{x=1}^{q}\sum_{y=1}^{r} e\Big(\frac{a(xr)^k}{q}\Big)e\Big(\frac{b(yq)^k}{r}\Big) \\
&= \sum_{x=1}^{q} e\Big(\frac{ax^k}{q}\Big)\sum_{y=1}^{r} e\Big(\frac{by^k}{r}\Big) \\
&= S(q,a)S(r,b).
\end{aligned}
$$

□

Lemma 7.10 Sind q und r relativ prim, dann gilt

$$A_N(qr) = A_N(q)A_N(r), \tag{7.28}$$

d.h. A_N ist multiplikativ.

Beweis: Wenn c und qr teilerfremd sind, dann gilt $c \equiv ar + bq \mod qr$ für gewisse a, b mit $(a,q) = (b,r) = 1$. Unter Verwendung des vorigen Lemmas folgt damit

$$
\begin{aligned}
A_N(qr) &= \sum_{\substack{c=1\\(c,qr)=1}}^{qr} \Big(\frac{S(qr,c)}{qr}\Big)^s e\Big(-\frac{cN}{qr}\Big) \\
&= \sum_{\substack{a=1\\(a,q)=1}}^{q}\ \sum_{\substack{b=1\\(b,r)=1}}^{r} \Big(\frac{S(qr,ar+bq)}{qr}\Big)^s e\Big(-\frac{(ar+bq)N}{qr}\Big) \\
&= \sum_{\substack{a=1\\(a,q)=1}}^{q}\ \sum_{\substack{b=1\\(b,r)=1}}^{r} \Big(\frac{S(q,a)}{q}\Big)^s \Big(\frac{S(r,b)}{r}\Big)^s e\Big(-\frac{aN}{q}\Big)e\Big(-\frac{bN}{r}\Big) \\
&= \sum_{\substack{a=1\\(a,q)=1}}^{q} \Big(\frac{S(q,a)}{q}\Big)^s e\Big(-\frac{aN}{q}\Big) \sum_{\substack{b=1\\(b,r)=1}}^{r} \Big(\frac{S(r,b)}{r}\Big)^s e\Big(-\frac{bN}{r}\Big) \\
&= A_N(q)A_N(r).
\end{aligned}
$$

□

Lemma 7.11 Für $q \in \mathbb{N}$ bezeichne $M_N(q)$ die Zahl der Lösungen von

$$x_1^k + \ldots + x_s^k \equiv N \mod q \tag{7.29}$$

mit ganzen Zahlen x_i und $1 \leq x_i \leq q$. Für eine Primzahl p ist die Reihe $\chi_N(p)$ durch

$$\chi_N(p) = 1 + \sum_{h=1}^{\infty} A_N(p^h) \tag{7.30}$$

definiert. Gilt wieder $s > 2^k$, dann konvergiert χ_N, und es gilt

$$\chi_N(p) = \lim_{h\to\infty} \frac{M_N(p^h)}{p^{h(s-1)}}. \tag{7.31}$$

Beweis: Die Konvergenz von $\chi_N(p)$ ergibt sich ebenso wie die Konvergenz von $\mathfrak{S}(N)$ aus der Abschätzung von $A_N(p^h)$. Ist $(a,q) = d$, dann folgt

$$\begin{aligned} S(q,a) &= \sum_{x=1}^{q} e\Big(\frac{ax^k}{q}\Big) = \sum_{x=1}^{q} e\Big(\frac{(a/d)x^k}{q/d}\Big) \\ &= d\sum_{x=1}^{q/d} e\Big(\frac{(a/d)x^k}{q/d}\Big) = dS(q/d, a/d). \end{aligned}$$

Wegen

$$\frac{1}{q}\sum_{a=1}^{q} e\Big(\frac{am}{q}\Big) = \begin{cases} 1 & m \equiv 0 \mod q \\ 0 & m \not\equiv 0 \mod q \end{cases}$$

gilt

$$\frac{1}{q}\sum_{a=1}^{q} e\Big(\frac{a(x_1^k + \ldots + x_s^k - N)}{q}\Big) = \begin{cases} 1 & x_1^k + \ldots + x_s^k \equiv N \mod q \\ 0 & x_1^k + \ldots + x_s^k \not\equiv N \mod q \end{cases}$$

und damit

$$\begin{aligned} M_N(q) &= \sum_{x_1=1}^{q} \cdots \sum_{x_s=1}^{q} \frac{1}{q}\sum_{a=1}^{q} e\Big(\frac{a(x_1^k + \ldots + x_s^k - N)}{q}\Big) \\ &= \frac{1}{q}\sum_{a=1}^{q}\sum_{x_1=1}^{q} e\Big(\frac{ax_1^k}{q}\Big) \cdots \sum_{x_s=1}^{q} e\Big(\frac{ax_s^k}{q}\Big) e\Big(\frac{-aN}{q}\Big) \\ &= \frac{1}{q}\sum_{a=1}^{q} S(q,a)^s e\Big(\frac{-aN}{q}\Big) \\ &= \frac{1}{q}\sum_{d|q}\sum_{\substack{a=1 \\ (a,q)=d}}^{q} S(q,a)^s e\Big(\frac{-aN}{q}\Big) \\ &= \frac{1}{q}\sum_{d|q}\sum_{\substack{a=1 \\ (a,q)=d}}^{q} d^s S(q/d, a/d)^s e\Big(\frac{-(a/d)N}{(q/d)}\Big) \end{aligned}$$

$$= q^{s-1} \sum_{d|q} \sum_{\substack{a=1 \\ (a,q)=d}}^{q} \Big(\frac{S(q/d, a/d)}{q/d}\Big)^s e\Big(\frac{-(a/d)N}{(q/d)}\Big)$$
$$= q^{s-1} \sum_{d|q} A_N(q/d).$$

Es folgt insbesondere für $q = p^h$

$$p^{h(1-s)} M_N(p^h) = \sum_{d|p^h} A_N(p^h/d) = 1 + \sum_{j=1}^{h} A_N(p^j)$$

und damit im Grenzwert

$$\chi_N(p) = \lim_{h\to\infty} \Big(1 + \sum_{j=1}^{h} A_N(p^j)\Big) = \lim_{h\to\infty} p^{h(1-s)} M_N(p^h).$$

□

Lemma 7.12 Für $s > 2^k$ ist

$$\mathfrak{S}(N) = \prod_p \chi_N(p). \tag{7.32}$$

Darüber hinaus gibt es eine Konstante c_2, die nur von k und s abhängt, sodass

$$0 < \mathfrak{S}(N) < c_2$$

für alle N, und es gibt eine Primzahl p_0, die von k und s abhängt, sodass

$$1/2 \le \prod_{p>p_0} \chi_N(p) \le 3/2 \tag{7.33}$$

für alle $N \ge 1$.

Beweis: Es wurde bereits gezeigt, dass für $s > 2^k$ die Abschätzung $A_N(q) \ll \frac{1}{q^{1+\delta_4}}$ gilt, wobei δ_4 von k und s abhängt. Damit konvergiert die Reihe absolut und gleichmäßig, und auch das Eulerprodukt konvergiert und stimmt im Wert überein (vgl. [21], S.325 ff.), d.h.

$$\mathfrak{S}(N) = \prod_p \chi_N(p).$$

Insbesondere ist $\chi_N(p) \neq 0$ für alle N, p. Aufgrund der Charakterisierung von $\chi_N(p)$ im vorigen Lemma ist $\chi_N(p)$ nichtnegativ und damit sogar strikt positiv. Also gilt dies auch für $\mathfrak{S}(N)$. Auch die Abschätzung von oben wurde bereits erwähnt. Zudem gilt

$$|\chi_N(p) - 1| \le \sum_{h=1}^{\infty} |A_N(p^h)| \ll \sum_{h=1}^{\infty} \frac{1}{p^{h(1+\delta_4)}} \ll \frac{1}{p^{1+\delta_4}}.$$

Damit gibt es eine Konstante c, die nur von k und s abhängt, sodass

$$1 - \frac{c}{p^{1+\delta_4}} \leq \chi_N(p) \leq 1 + \frac{c}{p^{1+\delta_4}}.$$

Da die unendlichen Produkte $\prod_p (1 \pm cp^{-1-\delta_4})$ konvergieren, muss für eine gewisse Primzahl p_0 die Abschätzung (7.33) gelten. Damit ist alles bewiesen. □

Wir wollen zeigen, dass $\mathfrak{S}(N)$ auch von unten durch eine von N unabhängige Konstante $c_1 > 0$ abgeschätzt werden kann. Nach dem soeben bewiesenen Resultat (7.33) genügt es dafür, zu beweisen, dass $\chi_N(p)$ für alle Primzahlen p gleichmäßig in N von unten beschränkt werden kann.

Lemma 7.13 Es sei p eine Primzahl, m eine nicht durch p teilbare natürliche Zahl. Es seien $\tau \geq 0$ und k_0 derart, dass $k = p^\tau k_0$ und $(p, k_0) = 1$. Definiere γ durch

$$\gamma = \begin{cases} \tau + 1 & p > 2 \\ \tau + 2 & p = 2. \end{cases}$$

Ist die Kongruenz $x^k \equiv m \bmod p^\gamma$ lösbar, dann auch die Kongruenz $y^k \equiv m \bmod p^h$ für alle $h \geq \gamma$.

Beweis: Zunächst sei $p \neq 2$. Dann gilt für $h \geq \gamma = \tau + 1$

$$(k, \varphi(p^h)) = (k_0 p^\tau, (p-1)p^{h-1}) = (k_0, p-1)p^\tau = (k, \varphi(p^\gamma)).$$

Die Gruppe der relativ primen Restklassen modulo p^h ist zyklisch mit Ordnung $\varphi(p^h) = (p-1)p^{h-1}$ (siehe z.B. [13], S.49ff.) und hat daher einen Erzeuger g, eine sog. Primitivwurzel. Diese ist dann auch Primitivwurzel modulo p^γ. Erfüllt x die Kongruenz $x^k \equiv m \bmod p^\gamma$, dann sind x und p relativ prim. Man kann daher r und u derart wählen, dass $x \equiv g^u \bmod p^h$ und $m \equiv g^r \bmod p^h$ gilt. Die Kongruenz, die x erfüllt, entspricht nun der Kongruenz

$$ku \equiv r \mod \varphi(p^\gamma).$$

Daraus folgt $r \equiv 0 \bmod (k, \varphi(p^\gamma))$ oder $r \equiv 0 \bmod (k, \varphi(p^h))$. Somit gibt es auch eine natürliche Zahl v, sodass

$$kv \equiv r \mod \varphi(p^h),$$

woraus sich $y^k \equiv m \bmod p^h$ für $y = g^v$ ergibt.

Nun sei $p = 2$. Damit müssen m und x ungerade sein. Für $\tau = 0$ ist auch k ungerade. Dann durchläuft y^k mit y alle ungeraden Kongruenzklassen modulo 2^h, und damit gibt es auch eine Lösung der Kongruenz $y^k \equiv m \bmod 2^h$. Ist $\tau \geq 1$, dann ist k gerade und $m \equiv x^k \equiv 1 \bmod 4$. Wegen $(-x)^k = x^k$ kann man annehmen, dass auch $x \equiv 1 \bmod 4$ gilt. Es ist bekannt (siehe wiederum [13]), dass die Kongruenzklassen modulo 2^h, die $\equiv 1 \bmod 4$

sind, eine zyklische Gruppe der Ordnung 2^{h-2} mit Erzeuger 5 bilden. Wähle r und u so, dass $x \equiv 5^u \mod 2^h$ und $m \equiv 5^r \mod 2^h$ gilt. Die Kongruenz von x entspricht dann

$$ku \equiv r \mod 2^{\gamma-2}.$$

Wie zuvor ergibt sich wegen

$$(k, 2^{h-2}) = (k_0 2^\tau, 2^{h-2}) = 2^\tau = (k, 2^{\gamma-2}),$$

dass auch ein v mit

$$kv \equiv r \mod 2^{h-2}$$

existiert, woraus sich wiederum $y^k \equiv m \mod p^h$ für $y = 5^v$ ergibt. □

Lemma 7.14 Es sei p eine Primzahl, γ wie zuvor. Wenn ganze Zahlen $a_1, \ldots, a_s$ existieren, die nicht alle durch p teilbar sind, und für die

$$a_1^k + \ldots + a_s^k \equiv N \mod p^\gamma \tag{7.34}$$

gilt, dann ist

$$\chi_N(p) \geq p^{\gamma(1-s)} > 0. \tag{7.35}$$

Beweis: Es gelte o.B.d.A. $a_1 \not\equiv 0 \mod p$. Sei $h > \gamma$. Zu jedem $i \in \{2, \ldots, s\}$ gibt es $p^{h-\gamma}$ paarweise inkongruente (modulo p^h) x_i, für die $x_i \equiv a_i \mod p^\gamma$ gilt. Die Kongruenz

$$x_1^k \equiv N - x_2^k - \ldots - x_s^k \mod p^\gamma$$

ist nach Voraussetzung lösbar (mit $x_1 = a_1$). Nach dem vorigen Lemma gibt es daher zu jedem $(s-1)$-tupel $(x_2, \ldots, x_s)$ ein x_1, sodass

$$x_1^k \equiv N - x_2^k - \ldots - x_s^k \mod p^h.$$

Daraus folgt unmittelbar $M_N(p^h) \geq p^{(h-\gamma)(s-1)}$ und folglich

$$\chi_N(p) = \lim_{h\to\infty} \frac{M_N(p^h)}{p^{h(s-1)}} \geq p^{\gamma(1-s)} > 0.$$

□

Lemma 7.15 Ist $s \geq 2k$ (k ungerade) bzw. $s \geq 4k$ (k gerade), dann gilt

$$\chi_N(p) \geq p^{\gamma(1-s)} > 0. \tag{7.36}$$

Beweis: Nach dem vorigen Lemma genügt es, die Lösbarkeit der Kongruenz

$$a_1^k + \ldots + a_s^k \equiv N \mod p^\gamma \tag{7.37}$$

nachzuweisen, wobei nicht alle a_i durch p teilbar sein sollen. Ist N nicht durch p teilbar, so gilt dies automatisch, ansonsten setze $a_s = 1$. Es genügt daher, die Existenz einer Lösung von (7.37) für $(N,p)=1$ und $s \geq 2k-1$ (k ungerade) bzw. $s \geq 4k-1$ (k gerade) zu zeigen. Zunächst sei p ungerade.

Sei g eine Primitivwurzel modulo p^γ. Die Ordnung von g ist dann $\varphi(p^\gamma) = (p-1)p^{\gamma-1} = (p-1)p^\tau$. Eine zu p teilerfremde ganze Zahl m mit $m \equiv g^r \mod p^\gamma$ ist genau dann ein k-ter Potenzrest modulo p^γ, wenn es ein v mit

$$kv \equiv r \mod (p-1)p^\tau$$

gibt. Wegen $k = k_0 p^\tau$ mit $(k_0,p)=1$ ist diese Kongruenz genau dann lösbar, wenn

$$r \equiv 0 \mod (k_0, p-1)p^\tau.$$

Daher gibt es genau

$$\frac{\varphi(p^\gamma)}{(k_0,p-1)p^\tau} = \frac{p-1}{(k_0,p-1)}$$

verschiedene k-te Potenzreste modulo p^γ. Es bezeichne $s(N)$ die kleinste natürliche Zahl s, für die die Kongruenz (7.37) lösbar ist, und $C(j)$ die Menge aller Kongruenzklassen N modulo p^γ mit $(N,p)=1$, für die $s(N)=j$ ist. Insbesondere beinhaltet $C(1)$ genau die k-ten Potenzreste.

Gilt $(m,p)=1$ und $N' = m^k N$, so folgt $s(N') = s(N)$ (multipliziere alle a_i mit m bzw. m^{-1} mod p^γ, um eine Darstellung von N in eine von N' und umgekehrt umzuwandeln). $C(j)$ ist somit abgeschlossen unter Multiplikation mit k-ten Potenzresten, woraus $|C(j)| \geq \frac{p-1}{(k_0,p-1)}$ folgt, falls $C(j) \neq \emptyset$. Es sei n maximal mit $C(n) \neq \emptyset$. Für $j < n$ sei N minimal mit $(N,p)=1$ und $s(N) > j$. Wähle $i \in \{1,2\}$ derart, dass $N-i$ relativ prim zu p ist (dies ist möglich, da p eine Primzahl ist). Wegen der Minimalität von N muss $s(N-i) \leq j$ gelten. Aus $N = (N-1) + 1^k = (N-2) + 1^k + 1^k$ folgt

$$j+1 \leq s(N) \leq s(N-i) + 2 \leq j+2.$$

und damit $s(N-i) \in \{j-1, j\}$. Also gibt es keine zwei aufeinanderfolgenden leeren Mengen $C(j)$ $(1 \leq j \leq n)$ und daher zumindest $\frac{n+1}{2}$ nichtleere Mengen $C(j)$. Da die $C(j)$ paarweise disjunkt sind, folgt

$$(p-1)p^\tau = \varphi(p^\gamma) = \sum_{\substack{j=1 \\ C(j)\neq\emptyset}}^{n} |C(j)| \geq \frac{n+1}{2} \frac{p-1}{(k_0,p-1)}$$

und damit

$$n \leq 2(k_0,p-1)p^\tau - 1 \leq 2k-1.$$

Also gilt tatsächlich $s(N) \leq 2k-1$ für eine ungerade Primzahl p und $(N,p)=1$.

Ist andererseits $p=2$ und k ungerade, so ist jede ungerade Zahl k-ter Potenzrest modulo

2^γ und daher $s(N) = 1$ für alle ungeraden N. Ist k gerade, so ist $k = 2^\tau k_0$ mit $\tau \geq 1$. Wir können $1 \leq N \leq 2^\gamma - 1$ annehmen. Wird

$$s = 2^\gamma - 1 = 4 \cdot 2^\tau - 1 \leq 4k - 1$$

gewählt, so kann (7.37) einfach gelöst werden, indem man $a_1 = \ldots = a_N = 1$ und $a_{N+1} = \ldots = a_s = 0$ setzt. Es folgt $s(N) \leq 4k - 1$ für alle ungeraden N, womit der Beweis vollständig ist. □

Proposition 7.16 Es gibt positive Konstanten c_1, c_2, die nur von k und s abhängen, sodass

$$c_1 < \mathfrak{S}(N) < c_2$$

für alle N gilt. Für hinreichend großes N ist

$$\mathfrak{S}(N, P^\nu) = \mathfrak{S}(N) + O(P^{-\nu\delta_4}).$$

Beweis: Bis auf die Existenz von c_1 wurde bereits alles gezeigt. Bereits bekannt ist auch, dass ein p_0 existiert, sodass

$$1/2 \leq \prod_{p > p_0} \chi_N(p) \leq 3/2$$

für alle N. Aus dem soeben bewiesenen Lemma wissen wir aber auch, dass für alle Primzahlen p die Abschätzung

$$\chi_N(p) \geq p^{\gamma(1-s)} > 0$$

gilt, womit

$$\mathfrak{S}(N) = \prod_p \chi_N(p) \geq \frac{1}{2} \prod_{p \leq p_0} \chi_N(p) \geq \frac{1}{2} \prod_{p \leq p_0} p^{\gamma(1-s)} =: c_1 > 0$$

folgt. □

7.5 Zusammenfassung der Ergebnisse

Nach all diesen Vorarbeiten können wir durch Zusammenfügen aller Resultate die asymptotische Formel von Hardy und Littlewood gewinnen, die zeigt, dass die Menge der k-ten Potenzen für beliebiges k eine Basis der natürlichen Zahlen bildet:

Satz 7.17 (Hardy/Littlewood [7]) *Es sei $k \geq 2$ und $s > 2^k$. Wenn $r_{k,s}(N)$ die Zahl der Darstellungen von N als Summe von s k-ten Potenzen bezeichnet, dann gibt es ein $\delta > 0$, das nur von s und k abhängt, sodass*

$$r_{k,s}(N) = \mathfrak{S}(N)\Gamma\Big(1 + \frac{1}{k}\Big)^s \Gamma\Big(\frac{s}{k}\Big)^{-1} N^{s/k-1} + O(N^{s/k-1-\delta}). \tag{7.38}$$

Die implizierte Konstante hängt ebenfalls nur von s und k ab, und es gibt $c_1 = c_1(k,s) > 0$, $c_2 = c_2(k,s) > 0$, sodass

$$c_1 < \mathfrak{S}(N) < c_2$$

für alle N. Insbesondere ist $r_{k,s}$ für hinreichend großes N stets positiv.

Beweis: Wir fassen die Ergebnisse aus den Propositionen 7.1, 7.5, 7.6, 7.8 und 7.16 zusammen. Sei $\delta_0 := \min(1, \delta_1, \delta_2, \delta_3, \nu\delta_4)$ mit $\delta_1, \ldots, \delta_4$ wie in den angegebenen Propositionen. Aus diesen ergibt sich

$$\begin{aligned}
r_{k,s}(N) &= \int_0^1 F(\alpha)^s e(-\alpha N)\, d\alpha \\
&= \int_{\mathfrak{M}} F(\alpha)^s e(-\alpha N)\, d\alpha + \int_{\mathfrak{m}} F(\alpha)^s e(-\alpha N)\, d\alpha \\
&= \mathfrak{S}(N, P^\nu) J^*(N) + O(P^{s-k-\delta_2}) + O(P^{s-k-\delta_1}) \\
&= (\mathfrak{S}(N) + O(P^{-\nu\delta_4}))(J(N) + O(P^{s-k-\delta_3})) + O(P^{s-k-\delta_2}) + O(P^{s-k-\delta_1}) \\
&= \mathfrak{S}(N) J(N) + O(P^{s-k-\delta_0}) \\
&= \mathfrak{S}(N) \Gamma\Big(1 + \frac{1}{k}\Big)^s \Gamma\Big(\frac{s}{k}\Big)^{-1} N^{s/k-1} + O(N^{(s-1)/k-1}) + O(N^{s/k-1-\delta_0/k}) \\
&= \mathfrak{S}(N) \Gamma\Big(1 + \frac{1}{k}\Big)^s \Gamma\Big(\frac{s}{k}\Big)^{-1} N^{s/k-1} + O(N^{s/k-1-\delta})
\end{aligned}$$

mit $\delta = \delta_0/k > 0$. □

Bemerkung: Es zeigt sich, dass die Formel auch für $k = 1$ ihre Gültigkeit behält. Es ist dann nach Proposition 5.1

$$r_{1,s}(N) = \binom{N-1}{s-1} = \frac{N^{s-1}}{(s-1)!} + O(N^{s-2}) = \frac{N^{s-1}}{\Gamma(s)} + O(N^{s-2}), \tag{7.39}$$

was mit der Formel (7.38) konsistent ist.

Bemerkung: In der bewiesenen Form (mit einem Gültigkeitsbereich von $s \geq s_0 = 2^k + 1$) bietet die Formel keine gute Abschätzung für die Ordnung der Menge der k-ten Potenzen als (asymptotische) Basis von $\mathbb{N}$. Hier konnten zahlreiche Verbesserungen getroffen werden. Bezeichnet $\tilde{G}(k)$ die kleinste natürliche Zahl s_0, sodass (7.38) für alle $s \geq s_0$ gilt, so konnte Ford [4] zeigen, dass sogar

$$\tilde{G}(k) \leq k^2(\log k + \log\log k + O(1)) \tag{7.40}$$

gilt. Was die Ordnung als asymptotische Basis angeht, so konnte die Schranke $2^k + 1$ von Wooley [41] bis auf

$$G(k) < k(\log k + \log\log k + O(1)) \tag{7.41}$$

verbessert werden.

Kapitel 8

Das Waring'sche Problem mit Ziffernbedingungen

In diesem abschließenden Kapitel wird eine Variante des Waring'schen Problems behandelt, bei dem die Summanden zusätzlichen Beschränkungen durch ihre Zifferndarstellung unterworfen sind. Dazu sei $q > 1$ die Basis des Ziffernsystems. Jede natürliche Zahl n hat dann eine eindeutige q-adische Darstellung in der Form

$$n = \sum_{i=0}^{r} c_i q^i$$

mit $0 \leq c_i < q$. Die *q-adische Ziffernsumme* von n ist

$$s_q(n) = \sum_{i=0}^{r} c_i.$$

Ihre wesentlichste Eigenschaft ist die *q-Additivität*, d.h. für $a, b, h \in \mathbb{N}$ mit $b < q^h$ gilt

$$s_q(aq^h + b) = s_q(a) + s_q(b).$$

Wir betrachten nun Mengen von der Form

$$A_{k,h,m} := \{n^k \mid s_q(n) \equiv h(m)\}$$

und behandeln analog zum Problem von Waring die Fragestellung, ob $A_{k,h,m}$ eine asymptotische Basis endlicher Ordnung darstellt. In der Tat stellt sich heraus, dass unter gewissen Voraussetzungen sogar eine asymptotische Formel von der Art (7.38) erfüllt ist. Dazu benötigt man eine durch Ziffernbedingungen modifizierte Version von Weyls Ungleichung.

Mengen von natürlichen Zahlen, deren q-adische Ziffernsumme eine gewisse Kongruenz erfüllt, wurden u.a. von Gelfond [5] und Mauduit/Sárközy [19] behandelt. Die hier beschriebene Version von Warings Problem wurde in einer Arbeit von Thuswaldner und

Tichy [32] bearbeitet.

Die wesentliche Voraussetzung, die getroffen werden muss, damit $A_{k,h,m}$ eine asymptotische Basis bildet, ist $(m, q-1) = 1$. Wegen $q^i \equiv 1 \mod (q-1)$ gilt nämlich $s_q(n) \equiv n \mod (q-1)$. Damit folgt aus $s_q(n) \equiv h \mod m$ aber $n \equiv h \mod (m, q-1)$. Summen von s Summanden $a_i \in A_{k,h,m}$ müssen daher stets die Kongruenz $a_1 + \ldots + a_s \equiv sh^k \mod (m, q-1)$ erfüllen und können somit nicht alle natürlichen Zahlen abdecken, falls $(m, q-1) \neq 1$.

In der Tat kann man das Problem sogar weiter verallgemeinern: ist wie zuvor $s > 2^k$, so werden wir zeigen, dass jedes hinreichend große $N \in \mathbb{N}$ eine Darstellung in der Form

$$N = x_1^k + \ldots x_s^k \text{ mit } s_{q_i} \equiv h_i \mod m_i \quad (1 \leq i \leq s)$$

hat. Hier muss die Voraussetzung $(m_i, q_i - 1) = 1$ für alle i getroffen werden.

Bevor die "circle method" zur Anwendung kommen kann, muss wieder eine Serie technischer Zwischenergebnisse bewiesen werden, an deren Ende die wesentlichen Resultate, Abschätzungen für Exponentialsummen mit Ziffernfunktionen, stehen.

8.1 Eine Klasse diskreter Funktionen

Definition 8.1 Es sei $\mathcal{M} = \{1, 2, \ldots, k\}$ und $\mathcal{M}' = \{0, 1, \ldots, k+1\}$. Wir definieren die Funktionenklasse $\mathcal{F}$ durch

$$\mathcal{F} := \{f : 2^{\mathcal{M}} \to \mathcal{M}'\}.$$

Es wird also jede Teilmenge von $\mathcal{M}$ (was der Menge der 0-1-Funktionen auf $\mathcal{M}$ entspricht) auf $\mathcal{M}'$ abgebildet. Zwei spezielle Vertreter dieser Klasse sind

$$F_0(S) := 0 \quad \forall S \subseteq \mathcal{M}$$

und

$$F_1(S) := \begin{cases} 1 & S = \mathcal{M} \\ 0 & S \neq \mathcal{M}. \end{cases}$$

Zudem sei der Operator Ξ auf $\mathcal{F}$ durch

$$\Xi_{\mathbf{r},i}(f)(S) := \left\lfloor \frac{i + \sum_{j \in S} r_j + f(S)}{q} \right\rfloor \tag{8.1}$$

definiert, wobei $\mathbf{r} = (r_1, \ldots, r_k) \in \{0, \ldots, q-1\}^k$ und $0 \leq i < q$ beliebig gewählt werden können.

Lemma 8.1 Für jedes Paar $(\mathbf{r}, i)$ gilt

$$\Xi_{\mathbf{r},i}(\mathcal{F}) \subseteq \mathcal{F}.$$

Beweis: Es muss gezeigt werden, dass $\Xi_{\mathbf{r},i}(f)(S)$ stets in $\mathcal{M}'$ liegt. Aufgrund der Voraussetzungen über $\mathbf{r}$ und i gilt aber

$$0 \leq \left\lfloor \frac{i + \sum_{j \in S} r_j + f(S)}{q} \right\rfloor \leq \frac{(k+1)(q-1)+k+1}{q} = k+1.$$

□

Nunmehr werden die Iterierten des Operators Ξ durch

$$\Xi_{(\mathbf{r}_l, i_l)_{1 \leq l \leq L}} := \Xi_{\mathbf{r}_L, i_L} \circ \ldots \circ \Xi_{\mathbf{r}_1, i_1}$$

definiert.

Definition 8.2 Es sei $k = d(q-1) + \rho$ mit $0 \leq \rho < q-1$ und $L'' := \lfloor \frac{k-1}{q-1} \rfloor + 1$. Ist $\rho = 0$, so definieren wir die Vektoren $\mathbf{v}_l = (v_{l1}, \ldots, v_{lk}) \in \mathbb{Z}^k$ durch

$$v_{lj} := \begin{cases} 1 & j \in \{(l-1)(q-1)+1, \ldots, l(q-1)\} \\ 0 & \text{sonst} \end{cases}$$

für $1 \leq l \leq L''$. Ist andererseits $\rho > 0$, so setzen wir

$$\mathbf{v}_1 := (\underbrace{1, \ldots, 1}_{\rho \text{ Stück}}, 0, \ldots, 0) \in \mathbb{Z}^k$$

und für $2 \leq l \leq L''$ $\mathbf{v}_l = (v_{l1}, \ldots, v_{lk}) \in \mathbb{Z}^k$ mit

$$v_{lj} := \begin{cases} 1 & j \in \{(l-2)(q-1)+\rho+1, \ldots, (l-1)(q-1)+\rho\} \\ 0 & \text{sonst.} \end{cases}$$

Lemma 8.2

1. Es sei $f \in \mathcal{F}$ beliebig. Dann gilt

$$\Xi_{(\mathbf{0},0)_{1 \leq l \leq L'}}(f) = F_0,$$

falls $L' := \lfloor \frac{\log(k+1)}{\log q} \rfloor + 1$.

2. Es seien

$$i_1 := \begin{cases} 1 & \rho = 0 \\ q - \rho & \rho > 0 \end{cases} \quad \text{und}$$

$$i_l := 0 \quad (2 \leq l \leq L'').$$

Dann gilt $\Xi_{(\mathbf{v}_l, i_l)_{1 \leq l \leq L''}}(F_0) = F_1$.

Beweis: Für $f \in \mathcal{F}$ und beliebiges $S \subseteq \mathcal{M}$ gilt

$$\Xi_{\mathbf{0},0}(f)(S) = \left\lfloor \frac{f(S)}{q} \right\rfloor \leq \frac{f(S)}{q}.$$

Wendet man dies L'-mal an, so erhält man

$$\Xi_{(\mathbf{0},0)_{1 \leq l \leq L'}}(f)(S) \leq \left\lfloor \frac{f(S)}{q^{L'}} \right\rfloor < \frac{k+1}{k+1} = 1,$$

womit 1. bereits bewiesen ist.

Sei nun zunächst $\rho = 0$. Dann folgt aus den Definitionen

$$\begin{aligned}\Xi_{\mathbf{v}_1,i_1}(F_0)(S) &= \left\lfloor \frac{1 + \sum_{t \in S \cap \{1,\ldots,q-1\}} 1}{q} \right\rfloor \\ &= \begin{cases} 1 & \{1,2,\ldots,q-1\} \subseteq S \\ 0 & \text{sonst.} \end{cases}\end{aligned}$$

Ist andererseits $\rho > 0$, so folgt

$$\begin{aligned}\Xi_{\mathbf{v}_1,i_1}(F_0)(S) &= \left\lfloor \frac{q - \rho + \sum_{t \in S \cap \{1,\ldots,\rho\}} 1}{q} \right\rfloor \\ &= \begin{cases} 1 & \{1,2,\ldots,\rho\} \subseteq S \\ 0 & \text{sonst.} \end{cases}\end{aligned}$$

Der weitere Beweis wird induktiv geführt. Wir zeigen, dass

$$\Xi_{(\mathbf{v}_l,i_l)_{1 \leq l \leq j}}(F_0) = \begin{cases} 1 & \{1,\ldots,j(q-1)\} \subseteq S \\ 0 & \text{sonst} \end{cases}$$

gilt, falls $\rho = 0$, und

$$\Xi_{(\mathbf{v}_l,i_l)_{1 \leq l \leq j}}(F_0) = \begin{cases} 1 & \{1,\ldots,(j-1)(q-1)+\rho\} \subseteq S \\ 0 & \text{sonst,} \end{cases}$$

falls $\rho > 0$. Für $j = L''$ ergibt sich sich dann das Gewünschte, nämlich

$$\Xi_{(\mathbf{v}_l,i_l)_{1 \leq l \leq L''}}(F_0) = \begin{cases} 1 & \{1,\ldots,k\} \subseteq S \\ 0 & \text{sonst.} \end{cases}$$

Da offensichtlich $\mathcal{M} = \{1,\ldots,k\} \subseteq S$ bereits $S = \mathcal{M}$ impliziert, ist dies tatsächlich die Funktion F_1.

Es wird nur der erste Fall genauer behandelt, der zweite verläuft analog. Der Induktionsschritt von $j-1$ auf j lässt sich dann in folgender Weise bewerkstelligen:

$$\begin{aligned}\Xi_{(\mathbf{v}_l,i_l)_{1\le l\le j}}(F_0) &= \begin{cases}\left\lfloor \frac{1+\sum_{t\in S\cap\{(j-1)(q-1)+1,\ldots,j(q-1)\}} 1}{q}\right\rfloor & \{1,\ldots,(j-1)(q-1)\}\subseteq S\\ \left\lfloor \frac{\sum_{t\in S\cap\{(j-1)(q-1)+1,\ldots,j(q-1)\}} 1}{q}\right\rfloor & \text{sonst}\end{cases}\\ &= \begin{cases}1 & \{1,\ldots,j(q-1)\}\subseteq S\\ 0 & \text{sonst.}\end{cases}\end{aligned}$$

Damit ist das Lemma bewiesen. □

8.2 Rekursionen für einige Hilfsfunktionen

Definition 8.3 Es seien k,m,h,q,N positive ganze Zahlen mit $m\ge 2$, $q\ge 2$ und $m\nmid h(q-1)$. Weiters seien $I_1,\ldots,I_k,J$ Intervalle ganzer Zahlen mit $\sqrt{N}\le |I_j|,|J|\le N$ $(1\le j\le k)$. Wir definieren die folgenden Hilfsfunktionen:

$$Y(I_1,\ldots,I_k,J) := \sum_{h_1\in I_1}\cdots\sum_{h_k\in I_k}\left|\sum_{n\in J} e\Big(\frac{h}{m}\Delta_{h_k,\ldots,h_1}(s_q)(n)\Big)\right|^2,$$

$$\Phi(h_1,\ldots,h_k;J;f) := \sum_{n\in J} e\left(\frac{h}{m}\sum_{S\subseteq\mathcal{M}}(-1)^{k-|S|}s_q\Big(n+\sum_{t\in S}h_t+f(S)\Big)\right),$$

$$\Psi(h_1,\ldots,h_{k-1};I_k,J;f_1,f_2) := \sum_{h_k\in I_k}\Phi(h_1,\ldots,h_k;J;f_1)\overline{\Phi(h_1,\ldots,h_k;J;f_2)},$$

$$X(I_1,\ldots,I_k,J;f_1,f_2) := \sum_{h_1\in I_1}\cdots\sum_{h_{k-1}\in I_{k-1}}\Psi(h_1,\ldots,h_{k-1};I_k,J;f_1,f_2).$$

Dabei sind die h_i ganze Zahlen und $f,f_1,f_2\in\mathcal{F}$.

Nun gilt für den Differenzoperator aber (Beweis analog zu Lemma 6.1)

$$\Delta_{h_k,\ldots,h_1}(f)(n) = \sum_{S\subseteq\mathcal{M}}(-1)^{k-|S|}f\Big(n+\sum_{t\in S}h_t\Big)$$

und daher

$$\Phi(h_1,\ldots,h_k;J;F_0) = \sum_{n\in J} e\left(\frac{h}{m}\sum_{S\subseteq\mathcal{M}}(-1)^{k-|S|}s_q\Big(n+\sum_{t\in S}h_t\Big)\right) = \sum_{n\in J} e\Big(\frac{h}{m}\Delta_{h_k,\ldots,h_1}(s_q)(n)\Big).$$

Damit folgt

$$Y(I_1,\ldots,I_k,J) = X(I_1,\ldots,I_k,J;F_0,F_0). \tag{8.2}$$

Im Folgenden werden die Abkürzungen $\mathbf{r} = (r_1, \ldots, r_k)$ und $\mathbf{h} = (h_1, \ldots, h_k)$ verwendet. Darüber hinaus sei für ein Intervall $I = \{n \in \mathbb{Z} \mid a \leq n < b\}$ ganzer Zahlen das Intervall cI $(c \in \mathbb{N})$ durch

$$cI := \{n \in \mathbb{Z} \mid ca \leq n < cb\}$$

definiert.

Proposition 8.3 Es seien $f_1, f_2 \in \mathcal{F}$ und $I_1, \ldots, I_k, J$ Intervalle ganzer Zahlen. Dann gilt

$$\begin{aligned} X(qI_1, \ldots, qI_k, qJ; f_1, f_2) = \sum_{r_1=0}^{q-1} \cdots \sum_{r_k=0}^{q-1} \sum_{i_1=0}^{q-1} \sum_{i_2=0}^{q-1} \alpha(f_1, f_2, \mathbf{r}, i_1, i_2) \\ \cdot X(I_1, \ldots, I_k, J; \Xi_{\mathbf{r},i_1}(f_1), \Xi_{\mathbf{r},i_2}(f_2)). \end{aligned} \tag{8.3}$$

Dabei ist

$$\alpha(f_1, f_2, \mathbf{r}, i_1, i_2) := e\Big(\frac{h}{m} \sum_{S \subseteq \mathcal{M}} (-1)^{k-|S|} (b(f_1, S, \mathbf{r}, i_1) - b(f_2, S, \mathbf{r}, i_2))\Big).$$

$b(f, S, \mathbf{r}, i) \in \{0, \ldots, q-1\}$ ist als der Rest von $i + \sum_{t \in S} r_t + f(S)$ modulo q definiert.

Beweis: Zunächst wird die Funktion Φ behandelt. Für $1 \leq r_1, \ldots, r_k \leq q$ gilt

$$\Phi(q\mathbf{h} + \mathbf{r}; qJ; f) = \sum_{i=0}^{q-1} \sum_{n \in J} e\left(\frac{h}{m} \sum_{S \subseteq \mathcal{M}} (-1)^{k-|S|} s_q\Big(qn + \sum_{t \in S} qh_t + i + \sum_{t \in S} r_t + f(S)\Big)\right).$$

Nach der Definition von Ξ und $b(f, S, \mathbf{r}, i)$ gilt

$$i + \sum_{t \in S} r_t + f(S) = \Xi_{\mathbf{r},i}(f)(S)q + b(f, S, \mathbf{r}, i).$$

Unter Verwendung der q-Additivität von s_q folgt

$$\begin{aligned} s_q\Big(qn + \sum_{t \in S} qh_t + i + \sum_{t \in S} r_t + f(S)\Big) &= s_q\Big(qn + \sum_{t \in S} qh_t + q\Xi_{\mathbf{r},i}(f)(S) + b(f, S, \mathbf{r}, i)\Big) \\ &= s_q\Big(n + \sum_{t \in S} h_t + \Xi_{\mathbf{r},i}(f)(S)\Big) + b(f, S, \mathbf{r}, i). \end{aligned}$$

Setzt man dies in die Gleichung für Φ ein, so ergibt sich

$$\Phi(q\mathbf{h} + \mathbf{r}; qJ; f) = \sum_{i=0}^{q-1} e\Big(\frac{h}{m} \sum_{S \in \mathcal{M}} (-1)^{k-|S|} b(f, S, \mathbf{r}, i)\Big) \Phi(\mathbf{h}; J; \Xi_{\mathbf{r},i}(f)).$$

Damit erhält man aus der Definition von Ψ

$$\begin{aligned} \Psi(qh_1 + r_1, \ldots, qh_{k-1} + r_{k-1}; qI_k, qJ; f_1, f_2) = \sum_{r_k=0}^{q-1} \sum_{i_1=0}^{q-1} \sum_{i_2=0}^{q-1} \alpha(f_1, f_2, \mathbf{r}, i_1, i_2) \\ \cdot \Psi(h_1, \ldots, h_{k-1}; I_k, J; \Xi_{\mathbf{r},i_1}(f_1), \Xi_{\mathbf{r},i_2}(f_2)) \end{aligned}$$

und durch Summation über $h_1, \ldots, h_{k-1}$

$$X(qI_1, \ldots, qI_k, qJ; f_1, f_2) = \sum_{r_1=0}^{q-1} \cdots \sum_{r_k=0}^{q-1} \sum_{i_1=0}^{q-1} \sum_{i_2=0}^{q-1} \alpha(f_1, f_2, \mathbf{r}, i_1, i_2) \cdot X(I_1, \ldots, I_k, J; \Xi_{\mathbf{r},i_1}(f_1), \Xi_{\mathbf{r},i_2}(f_2)).$$

Damit ist die Rekursion für X bewiesen. □

Es werden einige Werte von α für spezielle Argumente benötigt. Dies ist Gegenstand des folgenden Lemmas:

Lemma 8.4 Es sei $\mathbf{0}$ der Nullvektor. Dann gilt

$$\alpha(F_0, F_0, \mathbf{0}, 0, 0) = e(0) = 1, \tag{8.4}$$

$$\alpha(F_1, F_0, \mathbf{0}, 0, 0) = e\Big(\frac{h}{m}\Big), \tag{8.5}$$

$$\alpha(F_1, F_0, \mathbf{0}, q-1, 0) = e\Big(\frac{h}{m}(1-q)\Big). \tag{8.6}$$

Beweis: α ist durch

$$\alpha(f_1, f_2, \mathbf{r}, i_1, i_2) = e\Big(\frac{h}{m} \sum_{S \subseteq \mathcal{M}} (-1)^{k-|S|} (b(f_1, S, \mathbf{r}, i_1) - b(f_2, S, \mathbf{r}, i_2))\Big)$$

definiert, wobei $b(f, S, \mathbf{r}, i)$ der Rest von $i + \sum_{t \in S} r_t + f(S)$ bei Division durch q ist.

- Ist $f = F_0$ und alle $r_t = 0$ sowie $i = 0$, so folgt $i + \sum_{t \in S} r_t + f(S) = 0$ und damit $b(f, S, \mathbf{r}, i) = 0$ für alle $S \subseteq \mathcal{M}$. Die erste Gleichung ergibt sich unmittelbar.

- Ist $f = F_1$ und alle $r_t = 0$ sowie $i = 0$, so folgt $i + \sum_{t \in S} r_t + f(S) = 0$, falls $S \neq \mathcal{M}$, andernfalls $i + \sum_{t \in S} r_t + f(S) = 1$. Gleiches gilt damit für $b(f, S, \mathbf{r}, i)$, womit die zweite Gleichung folgt.

- Ist schließlich $f = F_1$ und alle $r_t = 0$ sowie $i = q-1$, so folgt $i + \sum_{t \in S} r_t + f(S) = q-1$, falls $S \neq \mathcal{M}$, andernfalls $i + \sum_{t \in S} r_t + f(S) = q$. Daher ist $b(f, S, \mathbf{r}, i) = 0$ für $S = \mathcal{M}$, andernfalls $q-1$. Man erhält

$$\begin{aligned} \sum_{S \subseteq \mathcal{M}} (-1)^{k-|S|} b(F_1, S, \mathbf{0}, q-1) &= -(q-1) + \sum_{S \subseteq \mathcal{M}} (-1)^{k-|S|} (q-1) \\ &= -(q-1) + (q-1) \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} = 1 - q, \end{aligned}$$

womit sich der dritte Punkt ergibt.

□

8.3 Abschätzung von Exponentialsummen

Unter Verwendung der Abkürzungen $\mathcal{Q}_l := \{0, \dots, q-1\}^l$, $\mathbf{r}_l = (r_{l1}, \dots, r_{lk})$ und $\mathbf{i}_l = (i_{l1}, i_{l2})$ erhält man durch L-fache Iteration von Proposition 8.3

$$\begin{aligned} X(q^L I_1, \dots, q^L I_k, q^L J; f_1, f_2) = & \sum_{r_1,\dots,r_L \in \mathcal{Q}_k} \sum_{i_1,\dots,i_L \in \mathcal{Q}_2} \\ & \Big(\prod_{l=1}^{L} \alpha(\Xi_{(\mathbf{r}_j, i_{j1})_{1\le j\le l-1}}(f_1), \Xi_{(\mathbf{r}_j, i_{j2})_{1\le j\le l-1}}(f_2), \mathbf{r}_l, i_{l1}, i_{l2}) \Big) \\ & \cdot X(I_1, \dots, I_k, J; \Xi_{(\mathbf{r}_l, i_{l1})_{1\le l\le L}}(f_1), \Xi_{(\mathbf{r}_l, i_{l2})_{1\le l\le L}}(f_2)). \end{aligned} \tag{8.7}$$

Wir wählen nun $L := L' + L'' + 3$ mit L' und L'' wie in Definition 8.2 und Lemma 8.2. Weiters sei wieder $k = d(q-1) + \rho$ mit $0 \le \rho < q-1$. Nun setzen wir

$$\begin{array}{lrl} \mathbf{r}_l = (0, \dots, 0), & \mathbf{i}_l = (0,0) & (1 \le l \le L'), \\ \mathbf{r}_l = \mathbf{v}_1, & \mathbf{i}_l = \begin{cases} (1,1) & \rho = 0 \\ (q-\rho, q-\rho) & \rho > 0 \end{cases} & (l = L'+1), \\ \mathbf{r}_l = \mathbf{v}_{l-L'}, & \mathbf{i}_l = (0,0) & (L'+2 \le l \le L-3), \\ \mathbf{r}_l = (0, \dots, 0), & \mathbf{i}_l = (q-1, 0) & (l = L-2), \\ \mathbf{r}_l = (0, \dots, 0), & \mathbf{i}_l = (q-1, 0) & (l = L-1), \\ \mathbf{r}_l = (0, \dots, 0), & \mathbf{i}_l = (0,0) & (l = L). \end{array}$$

Der Summand, der für diese spezielle Wahl in der Summe (8.7) entsteht, wird V_1 genannt. Ebenso sei V_2 jener Summand, der für dieselbe Wahl mit der Modifikation $\mathbf{i}_{L-1} = (0,0)$ entsteht. Zudem wird die Abkürzung

$$A(f_1, f_2) := \prod_{l=1}^{L-2} \alpha(\Xi_{(\mathbf{r}_j, i_{j1})_{1\le j\le l-1}}(f_1), \Xi_{(\mathbf{r}_j, i_{j2})_{1\le j\le l-1}}(f_2), \mathbf{r}_l, i_{l1}, i_{l2})$$

verwendet. Zur Behandlung der Summe V_1 ergeben sich zunächst aus der Definition von Ξ die Identitäten

$$\Xi_{\mathbf{0},q-1}(F_0) = F_0, \quad \Xi_{\mathbf{0},0}(F_1) = F_0, \quad \Xi_{\mathbf{0},q-1}(F_1) = F_1,$$

wie leicht nachgerechnet werden kann. Durch Anwendung von Lemma 8.2 erhalten wir damit

$$\Xi_{(\mathbf{r}_j, i_{j1})_{1\le j\le L-2}}(f_1) = \Xi_{(\mathbf{r}_j, i_{j1})_{L'+1\le j\le L-2}}(F_0) = \Xi_{\mathbf{0},q-1}(F_1) = F_1$$

für die erste Wahl von $\mathbf{r}$ und $\mathbf{i}$. In analoger Weise ergibt sich

$$\Xi_{(\mathbf{r}_j, i_{j2})_{1\le j\le L-2}}(f_2) = F_0.$$

Dies alles setzen wir in Gleichung (8.7) ein und erhalten

$$\begin{aligned}V_1 &= A(f_1,f_2)\,\alpha(\Xi_{(\mathbf{r}_j,i_{j1})_{1\le j\le L-2}}(f_1),\Xi_{(\mathbf{r}_j,i_{j2})_{1\le j\le L-2}}(f_2),\mathbf{r}_{L-1},i_{L-1,1},i_{L-1,2})\\ &\quad \alpha(\Xi_{(\mathbf{r}_j,i_{j1})_{1\le j\le L-1}}(f_1),\Xi_{(\mathbf{r}_j,i_{j2})_{1\le j\le L-1}}(f_2),\mathbf{r}_L,i_{L1},i_{L2})\\ &\quad X(I_1,\dots,I_k,J;\Xi_{(\mathbf{r}_j,i_{j1})_{1\le j\le L}}(f_1),\Xi_{(\mathbf{r}_j,i_{j2})_{1\le j\le L}}(f_2))\\ &= A(f_1,f_2)\,\alpha(F_1,F_0,\mathbf{r}_{L-1},i_{L-1,1},i_{L-1,2})\\ &\quad \alpha(\Xi_{\mathbf{r}_{L-1},i_{L-1,1}}(F_1),\Xi_{\mathbf{r}_{L-1},i_{L-1,2}}(F_0),\mathbf{r}_L,i_{L1},i_{L2})\\ &\quad X(I_1,\dots,I_k,J;\Xi_{(\mathbf{r}_j,i_{j1})_{L-1\le j\le L}}(F_1),\Xi_{(\mathbf{r}_j,i_{j2})_{L-1\le j\le L}}(F_0))\\ &= A(f_1,f_2)\,\alpha(F_1,F_0,\mathbf{0},q-1,0)\,\alpha(F_1,F_0,\mathbf{0},0,0)\,X(I_1,\dots,I_k,J;F_0,F_0).\end{aligned}$$

bzw. wieder analog

$$V_2 = A(f_1,f_2)\,\alpha(F_1,F_0,\mathbf{0},0,0)\,\alpha(F_0,F_0,\mathbf{0},0,0)\,X(I_1,\dots,I_k,J;F_0,F_0).$$

Nun können wir die speziellen Werte aus Lemma 8.4 einsetzen, womit sich

$$V_1 = A(f_1,f_2)e\Big(\frac{h}{m}(2-q)\Big)X(I_1,\dots,I_k,J;F_0,F_0)$$

und

$$V_2 = A(f_1,f_2)e\Big(\frac{h}{m}\Big)X(I_1,\dots,I_k,J;F_0,F_0)$$

ergibt. Nun schreiben wir Gleichung (8.7) in der Form

$$\begin{aligned}X(q^LI_1,\dots,q^LI_k,q^LJ;f_1,f_2) = \sum_D\Big(\prod_{l=1}^{L}\alpha(\Xi_{(\mathbf{r}_j,i_{j1})_{1\le j\le l-1}}(f_1),\Xi_{(\mathbf{r}_j,i_{j2})_{1\le j\le l-1}}(f_2),\mathbf{r}_l,i_{l1},i_{l2})\Big)\\ \cdot X(I_1,\dots,I_k,J;\Xi_{(\mathbf{r}_l,i_{l1})_{1\le l\le L}}(f_1),\Xi_{(\mathbf{r}_l,i_{l2})_{1\le l\le L}}(f_2)) + V_1 + V_2,\end{aligned}$$

wobei D den gesamten Summationsbereich mit Ausnahme der für V_1 und V_2 speziell gewählten Werte bezeichnet. Verändert man hier die Summationsreihenfolge, so kommt man auf

$$\begin{aligned}X(q^LI_1,\dots,q^LI_k,q^LJ;f_1,f_2) = \left(\sum_{\substack{g_1,g_2\in\mathcal{F}\\(g_1,g_2)\ne(F_0,F_0)}} a'(f_1,f_2,g_1,g_2)\,X(I_1,\dots,I_k,J;g_1,g_2)\right)\\ + \Big(a'(f_1,f_2,F_0,F_0)+A(f_1,f_2)\Big(e\Big(\frac{h}{m}(2-q)\Big)+e\Big(\frac{h}{m}\Big)\Big)\Big)\\ \cdot X(I_1,\dots,I_k,J;F_0,F_0).\end{aligned}\tag{8.8}$$

Mit $a'(f_1,f_2,g_1,g_2)$ wird dabei die Summe aller Koeffizienten von $X(I_1,\dots,I_k,J;g_1,g_2)$ bezeichnet, die in der Summe über D auftauchen. Da D genau $q^{(k+2)L}-2$ Summanden hat und stets $|\alpha(\cdot)|=1$ gilt, haben wir für beliebige $f_1,f_2\in\mathcal{F}$

$$\sum_{g_1,g_2\in\mathcal{F}}|a'(f_1,f_2,g_1,g_2)| \le q^{(k+2)L}-2.$$

Nunmehr setzen wir

$$a(f_1, f_2, g_1, g_2) = a'(f_1, f_2, g_1, g_2),$$

falls $(g_1, g_2) \neq (F_0, F_0)$, und andernfalls

$$a(f_1, f_2, g_1, g_2) = a'(f_1, f_2, g_1, g_2) + A(f_1, f_2)\Big(e\Big(\frac{h}{m}(2-q)\Big) + e\Big(\frac{h}{m}\Big)\Big).$$

Wegen $m \nmid h(q-1)$ gilt

$$\Big|e\Big(\frac{h}{m}(2-q)\Big) + e\Big(\frac{h}{m}\Big)\Big| \leq \Big|1 + e\Big(\frac{1}{m}\Big)\Big| = 2\Big|\cos\Big(\frac{\pi}{m}\Big)\Big| \leq 2 - \Big(\frac{\pi}{2m}\Big)^2.$$

Damit folgt nun

$$\sum_{g_1, g_2 \in \mathcal{F}} |a(f_1, f_2, g_1, g_2)| \leq q^{(k+2)L} - \Big(\frac{\pi}{2m}\Big)^2. \tag{8.9}$$

Somit ist gezeigt, dass die Summe über alle $|a(\cdot)|$ zumindest um einen kleinen Term von der trivialen Schranke $q^{(k+2)L}$ wegbleibt. Es sei nun B die $|\mathcal{F}|^2 \times |\mathcal{F}|^2$-Matrix

$$B(|a(f_1, f_2, g_1, g_2)|)_{(f_1,f_2),(g_1,g_2) \in \mathcal{F}^2}.$$

Dann folgt zunächst die Ungleichung

$$(|X(q^L I_1, \dots, q^L I_k, q^L J; f_1, f_2)|)_{(f_1,f_2)\in\mathcal{F}^2} \leq B \cdot (|X(I_1, \dots, I_k, J; g_1, g_2)|)_{(g_1,g_2)\in\mathcal{F}^2}, \tag{8.10}$$

deren einzelne Komponenten sich aus (8.8) unter Verwendung der Dreiecksungleichung ergeben.

8.4 Ein Korrelationssatz für die Ziffernsumme

Satz 8.5 *Sei die Funktion $p(k,q)$ durch*

$$p(k,q) := (k+2)\Big\lceil \frac{k-1}{q-1} + \frac{\log(k+1)}{\log q} + 5 \Big\rceil + 1$$

definiert. Für die in Definition 8.3 eingeführte Hilfsfunktion Y gilt dann die Abschätzung

$$Y(I_1, \dots, I_k, J) \ll |I_1| \dots |I_k| |J|^2 N^{-\eta} \tag{8.11}$$

mit $\eta = \frac{1}{m^2 q^{p(k,q)}} > 0$.

Beweis: Wir setzen zunächst $q' := q^L$ und $\varepsilon := \frac{\pi^2}{4m^2 q'^{k+2}}$. Nach der Ungleichung (8.9) ist jede Zeilensumme von B kleiner oder gleich $q'^{k+2}(1-\varepsilon)$, d.h. (da die Einträge von B nichtnegativ sind) für die Zeilensummennorm gilt $\|B\|_\infty \leq q'^{k+2}(1-\varepsilon)$. Es folgt $\|B^l\|_\infty \leq \|B\|_\infty^l \leq q'^{(k+2)l}(1-\varepsilon)^l$, d.h. die Zeilensummen von B^l sind kleiner oder gleich $q'^{(k+2)l}(1-\varepsilon)^l$.

Darüber hinaus kann man $|X(\cdot)|$ in trivialer Weise abschätzen, indem man $|e(\cdot)| = 1$ zusammen mit der Dreiecksungleichung verwendet:

$$|X(I_1, \ldots, I_k, J; f_1, f_2)| \leq |I_1| \ldots |I_k||J|^2.$$

Zusammen mit der l-fachen Iteration der Ungleichung (8.10) erhält man nun

$$|X(q'^l I_1, \ldots, q'^l I_k, q'^l J; f_1, f_2)| \leq (1-\varepsilon)^l (q'^l |I_1|) \ldots (q'^l |I_k|)(q'^l |J|)^2. \tag{8.12}$$

Setzt man $t := \lfloor \frac{10 \log N}{21 \log q'} \rfloor$ (was für hinreichend großes N positiv ist), so ist zunächst $q'^t < \sqrt{N}$. Ist nun $I_j = [a_j, b_j]$ $(1 \leq j \leq k)$ und $J = [a_{k+1}, b_{k+1}]$, dann schreiben wir $a_j = q'^t u_j + r_j$ und $b_j = q'^t v_j + s_j$ mit $0 \leq r_j, s_j < q'^t$. Da die Intervalle nach Voraussetzung zumindest Länge $\sqrt{N}$ haben, gilt $|u_j - v_j| \geq 1$. Wir setzen nun $\tilde{I}_j := [u_j, v_j]$ $(1 \leq j \leq k)$ und $\tilde{J} := [u_{j+1}, v_{j+1}]$.

Aus der Definition von $X(I_1, \ldots, I_k, J; f_1, f_2)$ erhalten wir (wieder durch Verwendung der einfachen Tatsache, dass alle Summanden von der Form $e(\cdot)$ sind und damit Betrag 1 haben)

$$X(I_1, \ldots, I_k, J, f_1, f_2) = X(q'^t \tilde{I}_1, \ldots, q'^t \tilde{I}_k, q'^t \tilde{J}, f_1, f_2) + O\Big(|I_1| \ldots |I_k||J|^2 \frac{q'^t}{\sqrt{N}}\Big),$$

denn die Summationsbereiche von $X(I_1, \ldots, I_k, J, f_1, f_2)$ und $(q'^t \tilde{I}_1, \ldots, q'^t \tilde{I}_k, q'^t \tilde{J}, f_1, f_2)$ unterscheiden sich nur um eine Menge, deren Größe durch

$$(k+2)|I_1| \ldots |I_k||J|^2 \frac{\max_j(r_j) + q'^t - \max_j(s_j)}{\min_j(|I_j|, |J|)} \ll |I_1| \ldots |I_k||J|^2 \frac{q'^t}{\sqrt{N}}$$

beschränkt werden kann. Unter Verwendung von (8.12) (mit $\tilde{I}_j$ anstelle von I_j und $\tilde{J}$ anstelle von J) und $(1-\varepsilon)^t \leq e^{-t\varepsilon}$ ergibt sich

$$X(I_1, \ldots, I_k, J, f_1, f_2) \ll \Big(e^{-t\varepsilon} + \frac{q'^t}{\sqrt{N}}\Big)|I_1| \ldots |I_k||J|^2. \tag{8.13}$$

Aus der Definition von t erhalten wir für hinreichend großes N

$$-t\varepsilon \leq -\frac{10 \log N}{22 \log q'} \frac{\pi^2}{4m^2 q'^{k+2}} \leq -\frac{\log N}{m^2 q'^{k+2} \log q'} \leq -\frac{\log N}{m^2 q^{L(k+2)+1}}.$$

Wegen $L = L' + L'' + 3 \leq \left\lceil \frac{k-1}{q-1} + \frac{\log(k+1)}{\log q} + 5 \right\rceil$ folgt $-t\varepsilon \leq \frac{\log N}{m^2 q^{p(k,q)}}$ mit dem im Satz definierten $p(k,q)$. Zudem ist

$$\frac{q'^t}{\sqrt{N}} \leq \frac{\exp\left(\frac{10}{21} \log N\right)}{\sqrt{N}} = N^{-1/42} \leq N^{-1/(m^2 q^{p(k,q)})}.$$

Der Spezialfall $f_1 = f_2 = F_0$ von (8.13) ergibt wegen (8.2) das gewünschte Resultat. $\square$

8.5 Weyls Ungleichung und die Ziffernsumme

Die technischen Hilfsresultate der letzten Abschnitte gipfeln nunmehr im Beweis des folgenden Satzes, der es ermöglicht, Warings Problem mit Ziffernbedingungen auf die asymptotische Formel für das Originalproblem zurückzuführen. Es handelt sich um eine Variante der Ungleichung von Weyl, in die die Ziffernsumme miteinfließt.

Satz 8.6 *Es seien k, m, l, q und N positive ganze Zahlen mit $m \geq 2$, $q \geq 2$ und $m \nmid l(q-1)$. Dann gilt*

$$\Big|\sum_{n=1}^{N} e\Big(\theta n^k + \frac{l}{m} s_q(n)\Big)\Big| \ll N^{1-\gamma} \tag{8.14}$$

gleichmäßig in $\theta \in [0,1)$ mit $\gamma := \eta 2^{-(k+1)}$. Dabei ist η wie in Satz 8.5.

Beweis: Wir beginnen mit der Abschätzung aus Lemma 6.14: ist

$$S(f) = \sum_{n=N_1+1}^{N_2} e(f(n))$$

für eine arithmetische Funktion f (hier ist $f(n) = \theta n^k + \frac{l}{m} s_q(n)$), dann gilt

$$|S(f)|^{2^l} \leq (2N)^{2^l-l-1} \sum_{|d_1|<N} \cdots \sum_{|d_l|<N} S_{d_l,\ldots,d_1}(f)$$

mit

$$S_{d_l,\ldots,d_1}(f) = \sum_{n \in I(d_l,\ldots,d_1)} e(\Delta_{d_l,\ldots,d_1}(f)(n)),$$

wobei $I(d_l,\ldots,d_1)$ ein Intervall aufeinanderfolgender ganzer Zahlen ist, das in $[N_1+1, N_2]$ enthalten ist. Insbesondere ist hier $N_1 = 0, N_2 = N$. Die $I(d_l,\ldots,d_1)$ wurden im Beweis rekursiv durch

$$I(d_1) = [N_1+1-d_1, N_2-d_1] \cap [N_1+1, N_2]$$

und

$$I(d_l,\ldots,d_1) = I(d_{l-1},\ldots,d_1) \cap \{x \mid x + d_l \in I(d_{l-1},\ldots,d_1)\}$$

definiert.

Im Folgenden wird nun Information über die k-ten Differenzen

$$\Delta_{d_k,\ldots,d_1}(f)(n)$$

mit $f(n) = \theta n^k + \frac{l}{m} s_q(n)$ benötigt. Aufgrund der Linearität des Differenzoperators genügt es dafür, n^k und $s_q(n)$ einzeln zu studieren. Über ersteren Teil ist bereits aus Lemma 6.2 bekannt, dass

$$\Delta_{d_k,\ldots,d_1}(x^k) = d_1 \ldots d_k k!$$

ist. Damit folgt nun

$$\left|S\Big(\theta n^k + \frac{l}{m}s_q(n)\Big)\right|^{2^k} \leq (2N)^{2^k-k-1} \sum_{|d_1|<N} \cdots \sum_{|d_k|<N} \left|\sum_{n\in I(d_k,\dots,d_1)} e\Big(\theta d_1 \dots d_k k! + \frac{l}{m}\Delta_{d_k,\dots,d_1}(s_q)(n)\Big)\right|$$
$$= (2N)^{2^k-k-1} \sum_{|d_1|<N} \cdots \sum_{|d_k|<N} \left|\sum_{n\in I(d_k,\dots,d_1)} e\Big(\frac{l}{m}\Delta_{d_k,\dots,d_1}(s_q)(n)\Big)\right|. \tag{8.15}$$

Dieser Ausdruck ähnelt jenem für die Hilfsfunktion Y, mit dem wesentlichen Unterschied, dass der Summationsbereich der letzten Summe noch von den d_i abhängt. Zudem ist der Betrag der letzten Summe nicht quadriert. Diese Schwierigkeiten müssen nun noch behandelt werden. Wir wählen zunächst reelle Zahlen $\alpha, \beta, \varepsilon$ mit $\alpha > \frac{\eta}{2}$, $\beta \geq \frac{1}{2}$, $\alpha + \beta = 1$, $0 < \varepsilon \leq \alpha - \frac{\eta}{2}$. Nun zerlegen wir die Summe

$$S := \sum_{|d_1|<N} \cdots \sum_{|d_k|<N} \left|\sum_{n\in I(d_k,\dots,d_1)} e\Big(\frac{l}{m}\Delta_{d_k,\dots,d_1}(s_q)(n)\Big)\right|$$

in folgender Weise in Blöcke der Länge $\lfloor N^\beta \rfloor$:

$$S = \sum_{j_1=-\lfloor N^\alpha\rfloor-1}^{\lfloor N^\alpha\rfloor+1} \cdots \sum_{j_k=-\lfloor N^\alpha\rfloor-1}^{\lfloor N^\alpha\rfloor+1} R(j_1,\dots,j_k) + O(N^{k\beta+1})$$

mit

$$R(j_1,\dots,j_k) := \sum_{d_1=j_1\lfloor N^\beta\rfloor}^{(j_1+1)\lfloor N^\beta\rfloor-1} \cdots \sum_{d_k=j_k\lfloor N^\beta\rfloor}^{(j_k+1)\lfloor N^\beta\rfloor-1} \left|\sum_{n\in I(d_k,\dots,d_1)} e\Big(\frac{l}{m}\Delta_{d_k,\dots,d_1}(s_q)(n)\Big)\right|.$$

Nun schätzen wir die einzelnen Blöcke $R(j_1,\dots,j_k)$ ab. Dazu werden zwei verschiedene Fälle unterschieden:

1. Angenommen, es gelte $|I(d_k,\dots,d_1)| > N^{\beta+\varepsilon}$ für alle $d_1,\dots,d_k$ mit
$$j_r\lfloor N^\beta\rfloor \leq d_r < (j_r+1)\lfloor N^\beta\rfloor \quad (1 \leq r \leq k). \tag{8.16}$$
Die Intervallgrenzen von $I(d_k,\dots,d_1)$ hängen nach Definition linear von den d_i ab. Diese variieren aber nur in einem Intervall der Länge $\lfloor N^\beta\rfloor$. Es gibt somit Konstanten u und v mit
$$I(d_k,\dots,d_1) = [u + O(N^\beta), v + O(n^\beta)]$$

für jedes k-tupel, das (8.16) erfüllt. Ersetzt man nun $I(d_k,\dots,d_1)$ durch $I'(d_k,\dots,d_1) = [u,v]$, dann hängt I' im Gegensatz zu I innerhalb eines Blocks, der durch (8.16) gegeben ist, nicht mehr von den d_i ab, und es gilt für die symmetrische Differenz der beiden Mengen

$$|I(d_k,\dots,d_1)\,\Delta\, I'(d_k,\dots,d_1)| \ll N^\beta.$$

Es ergibt sich damit

$$R(j_1,\dots,j_k) = \sum_{d_1=j_1\lfloor N^\beta\rfloor}^{(j_1+1)\lfloor N^\beta\rfloor-1} \cdots \sum_{d_k=j_k\lfloor N^\beta\rfloor}^{(j_k+1)\lfloor N^\beta\rfloor-1} \left| \sum_{n\in I'(d_k,\dots,d_1)} e\Big(\frac{l}{m}\Delta_{d_k,\dots,d_1}(s_q)(n)\Big)\right| + O(N^{(k+1)\beta})$$

und durch Anwendung der Cauchy-Schwarz-Ungleichung

$$R(j_1,\dots,j_k) \le \left(N^{k\beta} \sum_{d_1=j_1\lfloor N^\beta\rfloor}^{(j_1+1)\lfloor N^\beta\rfloor-1} \cdots \sum_{d_k=j_k\lfloor N^\beta\rfloor}^{(j_k+1)\lfloor N^\beta\rfloor-1} \left| \sum_{n\in I'(d_k,\dots,d_1)} e\Big(\frac{l}{m}\Delta_{d_k,\dots,d_1}(s_q)(n)\Big)\right|^2 \right)^{\frac12} + O(N^{(k+1)\beta}).$$

Wegen $\beta \ge \frac12$ sind die Bedingungen von Satz 8.5 (über die Intervalllängen) erfüllt, und es folgt

$$R(j_1,\dots,j_k) \ll N^{k\beta+1-\frac{\eta}{2}} + N^{(k+1)\beta} \ll N^{k\beta+1-\frac{\eta}{2}}.$$

2. Sei andererseits $|I(d_k,\dots,d_1)| \le N^{\beta+\varepsilon}$ für zumindest ein k-tupel $(d_1,\dots,d_k)$, das (8.16) erfüllt. Da die Intervalllänge von $I(d_k,\dots,d_1)$ linear von den d_i abhängt, muss damit auch

$$|I(d_k,\dots,d_1)| \ll N^{\beta+\varepsilon}$$

für alle solchen k-tupel gelten. Die triviale Abschätzung der Exponentialsumme durch die Anzahl der Summanden führt dann auf

$$R(j_1,\dots,j_k) \ll N^{(k+1)\beta+\varepsilon} \ll N^{k\beta+1-\frac{\eta}{2}}.$$

$R(j_1,\dots,j_k)$ kann also für alle k-tupel $(j_1,\dots,j_k)$ in der Form $R(j_1,\dots,j_k) \ll N^{k\beta+1-\frac{\eta}{2}}$ abgeschätzt werden. Setzt man dies in den Ausdruck für S ein, so erhält man nun

$$S \ll N^{k\alpha+k\beta+1-\frac{\eta}{2}} = N^{k+1-\frac{\eta}{2}}$$

und schließlich unter Verwendung von (8.15)

$$\left| S\Big(\theta n^k + \frac{l}{m}s_q(n)\Big)\right|^{2^k} \ll N^{2^k-\frac{\eta}{2}}.$$

Durch Ziehen der 2^k-ten Wurzel erhält man das gewünschte Resultat. □

8.6 Die asymptotische Formel für Warings Problem mit Ziffernbedingungen

Nunmehr wird abermals die "circle method" angewandt. Wir setzen wieder $P := \lfloor N^{1/k} \rfloor$; bezeichnet

$$U_{h,m,q}(P) := \{n \leq P \mid s_q(n) \equiv h(m)\}$$

die Menge der natürlichen Zahlen $\leq P$ mit "erlaubter" Ziffernsume, so ist die erzeugende Funktion für das vorliegende Problem durch

$$f_{h,m,q}(z) = \sum_{n \in U_{h,m,q}(P)} z^{n^k} \tag{8.17}$$

gegeben. Bezeichnet $r'_{k,s}(N)$ hier die Anzahl der Repräsentationen von N in der Form

$$N = x_1^k + \dots x_s^k \text{ mit } s_{q_i} \equiv h_i \mod m_i \quad (1 \leq i \leq s),$$

so ist die erzeugende Funktion für $r'_{k,s}(N)$

$$\prod_{i=1}^{s} f_{h_i,m_i,q_i}(z) = \sum_{n=1}^{N} r'_{k,s}(N) z^N$$

Anwendung der Cauchy'schen Integralformel führt auf

$$\begin{aligned} r'_{k,s}(N) &= \frac{1}{2\pi i} \int_{\mathcal{C}} \prod_{i=1}^{s} f_{h_i,m_i,q_i}(z) z^{-(N+1)} \, dz \\ &= \int_0^1 \sum_{n_1 \in U_{h_1,m_1,q_1}(P)} \dots \sum_{n_s \in U_{h_s,m_s,q_s}(P)} e(\theta(n_1^k + \dots + n_s^k - N)) \, d\theta. \end{aligned} \tag{8.18}$$

Das Resultat, das nun bewiesen werden soll, lautet folgendermaßen:

Satz 8.7 (Thuswaldner/Tichy [32]) *Es seien s, k, h_i, m_i, q_i positive ganze Zahlen mit $m_i \geq 2$, $q_i \geq 2$ und $(q_i - 1, m_i) = 1$ $(1 \leq i \leq s)$. Dann gibt es für $s > 2^k$ ein $\delta > 0$, sodass*

$$r'_{k,s}(N) = \frac{1}{m_1 \dots m_s} \mathfrak{S}(N) \Gamma\Big(1 + \frac{1}{k}\Big)^s \Gamma\Big(\frac{s}{k}\Big)^{-1} N^{s/k-1} + O(N^{s/k-1-\delta}). \tag{8.19}$$

Die implizierte Konstante hängt nur von s, k und den m_i ab. $\mathfrak{S}$ ist die bereits bekannte singuläre Reihe, die durch Konstanten $0 < c_1 < c_2$ gleichmäßig in N von oben und unten beschränkt werden kann.

Bemerkung: Sind alle h_i, alle m_i und alle q_i gleich, so impliziert dies, dass die Menge $A_{k,h,m}$ eine asymptotische Basis der natürlichen Zahlen von endlicher Ordnung darstellt. Die asymptotische Ordnung ist dabei $\leq 2^k + 1$.

Beweis: Es wird eine bestimmte Hilfsfunktion eingeführt, um mit dem Summationsbereich $U_{h_i,m_i,q_i}(P)$ umgehen zu können. Diese Technik geht auf Gelfond [5] zurück. Wir setzen für eine arithmetische Funktion φ

$$H_l(\varphi, P) := \sum_{n=0}^{P} e\Big(\varphi(n) + \frac{l}{m} s_q(n)\Big).$$

Dann gilt

$$\begin{aligned}\sum_{l=0}^{m-1} e\Big(-\frac{lh}{m}\Big) H_l(\varphi, P) &= \sum_{n=0}^{P}\sum_{l=0}^{m-1} e\Big(l\,\frac{s_q(n)-h}{m}\Big) e(\varphi(n)) \\ &= m \sum_{n\in U_{h,m,q(P)}} e(\varphi(n)).\end{aligned}$$

Damit schreiben wir

$$\begin{aligned}\sum_{n\in U_{h,m,q(P)}} e(\theta n^k) &= \frac{1}{m}\sum_{l=0}^{m-1} e\Big(-\frac{lh}{m}\Big) H_l(\theta n^k, P) \\ &= \frac{1}{m}\sum_{l=0}^{m-1}\sum_{n=0}^{P} e\Big(l\,\frac{s_q(n)-h}{m}\Big) e(\theta n^k)\end{aligned}$$

und setzen dies in die Integraldarstellung von $r'_{k,s}(N)$ ein:

$$\begin{aligned} r'_{k,s}(N) = \frac{1}{m_1\dots m_s}\int_0^1 & \sum_{n_1\le P}\dots\sum_{n_s\le P}\sum_{l_1=0}^{m_1-1}\dots\sum_{l_s=0}^{m_s-1} \\ & e\Big(l_1\,\frac{s_{q_1}(n_1)-h_1}{m_1}\Big)\dots e\Big(l_s\,\frac{s_{q_s}(n_s)-h_s}{m_s}\Big) e(\theta(n_1^k+\dots+n_s^k-N))\,d\theta.\end{aligned}$$

Dieses Integral wird in zwei Teile gespalten: jenen mit $l_1 = \dots = l_s = 0$ und jenen mit $(l_1,\dots,l_s) \neq (0,\dots,0)$. Es ergibt sich

$$\begin{aligned} r'_{k,s}(N) = {} & \frac{1}{m_1\dots m_s}\int_0^1 \sum_{n_1\le P}\dots\sum_{n_s\le P} e(\theta(n_1^k+\dots+n_s^k-N))\,d\theta \\ & + \frac{1}{m_1\dots m_s}\int_0^1 \sum_{n_1\le P}\dots\sum_{n_s\le P}\underbrace{\sum_{l_1=0}^{m_1-1}\dots\sum_{l_s=0}^{m_s-1}}_{(l_1,\dots,l_s)\neq(0,\dots,0)} \\ & e\Big(l_1\,\frac{s_{q_1}(n_1)-h_1}{m_1}\Big)\dots e\Big(l_s\,\frac{s_{q_s}(n_s)-h_s}{m_s}\Big) e(\theta(n_1^k+\dots+n_s^k-N))\,d\theta.\end{aligned}$$

Das erste Integral wird mit $\mathcal{I}_1$ bezeichnet. Es enthält keinerlei Information über die Ziffernsumme mehr und ist gleich dem Integral für das gewöhnliche Problem von Waring.

Für diesen Teil kann daher die asymptotische Formel aus Satz 7.17 herangezogen werden. Kann gezeigt werden, dass sich das zweite Integral $\mathcal{I}_2$ in der Form $\mathcal{I}_2 \ll P^{s-k-\gamma}$ mit einem $\gamma > 0$ abschätzen lässt und somit nichts zum Hauptterm der Asymptotik beiträgt, dann ist die Behauptung bereits bewiesen. Dies muss nur für ein beliebiges, aber festes s-tupel $L = (l_1, \ldots, l_s) \neq (0, \ldots, 0)$ nachgeprüft werden. Wir schreiben dazu $\mathcal{J}_L$ für das Integral

$$\mathcal{J}_L := \int_0^1 \left(\sum_{n_1 \le P} e\Big(\theta n_1^k + l_1 \frac{s_{q_1}(n_1) - h_1}{m_1}\Big)\right) \cdots \left(\sum_{n_s \le P} e\Big(\theta n_s^k + l_s \frac{s_{q_s}(n_s) - h_s}{m_s}\Big)\right) e(-N\theta)\, d\theta.$$

Wir setzen weiters

$$S_j(\theta) := \sum_{n \le P} e\Big(\theta n^k + \frac{l_j}{m_j} s_{q_j}(n)\Big).$$

Der Integrand von $\mathcal{J}_L$ enthält $s > 2^k$ Faktoren $S_j(\theta) e\Big(-\frac{h_j}{m_j}\Big)$. Wir können daher $\mathcal{J}_L$ durch

$$\begin{aligned}
|\mathcal{J}_L| &\le \sup_{\theta,j} \Big(|S_j(\theta)|^{s-2^k}\Big) \int_0^1 \prod_{j=1}^{2^k} |S_j(\theta)|\, d\theta \\
&\le \sup_{\theta,j} \Big(|S_j(\theta)|^{s-2^k}\Big) \Big(\prod_{j=1}^{2^k} \int_0^1 |S_j(\theta)|^{2^k}\, d\theta\Big)^{2^{-k}} \\
&\le \sup_{\theta,j} \Big(|S_j(\theta)|^{s-2^k}\Big) \max_t \Big(\int_0^1 |S_t(\theta)|^{2^k}\, d\theta\Big)
\end{aligned}$$

abschätzen. Der zweite Schritt ergibt sich dabei durch wiederholte Anwendung der Cauchy-Schwarz-Ungleichung. Der erste Ausdruck lässt sich nach Satz 8.6 durch

$$\sup_{\theta,j} \Big(|S_j(\theta)|^{s-2^k}\Big) \ll P^{(1-\gamma)(s-2^k)} \le P^{s-2^k-\gamma}$$

mit γ wie in Satz 8.6 abschätzen. Das Integral im zweiten Ausdruck schreiben wir in folgender Weise um:

$$\begin{aligned}
\int_0^1 |S_t(\theta)|^{2^k}\, d\theta &= \int_0^1 S_t(\theta)^{2^{k-1}} \overline{S_t(\theta)}^{2^{k-1}}\, d\theta \\
&= \sum_{1 \le n_1, \ldots, n_{2^k} \le P} \int_0^1 e\Bigg(\theta\Big(n_1^k + \ldots + n_{2^{k-1}}^k - n_{2^{k-1}+1}^k - \ldots - n_{2^k}^k\Big) \\
&\quad + \frac{l_t}{m_t}\Big(s_{q_t}(n_1) + \ldots + s_{q_t}(n_{2^{k-1}}) - s_{q_t}(n_{2^{k-1}+1}) - \ldots - s_{q_t}(n_{2^k})\Big)\Bigg)\, d\theta \\
&= \sum_{n_1, \ldots, n_{2^k}} e\Bigg(\frac{l_t}{m_t} \sum_{r=1}^{2^{k-1}} s_{q_t}(n_r) - s_{q_t}(n_{2^{k-1}+r})\Bigg).
\end{aligned}$$

Die Summe geht dabei über alle $1 \leq n_1, \ldots, n_{2^k} \leq P$ mit

$$n_1^k + \ldots + n_{2^{k-1}}^k = n_{2^{k-1}+1}^k + \ldots + n_{2^k}^k$$

und lässt sich daher durch

$$\left| \left\{ 1 \leq n_1, \ldots, n_{2^k} \leq P \,\middle|\, n_1^k + \ldots + n_{2^{k-1}}^k = n_{2^{k-1}+1}^k + \ldots + n_{2^k}^k \right\} \right|$$

von oben abschätzen. Analog zur Umformung von $\int_0^1 |S_t(\theta)|^{2^k} \, d\theta$ kann man für das Integral $\int_0^1 |T(\theta)|^{2^k} \, d\theta$ mit $T(\theta) = \sum_{n=1}^P e(\theta n^k)$ Folgendes gewinnen:

$$\begin{aligned} \int_0^1 |T(\theta)|^{2^k} \, d\theta &= \sum_{n_1, \ldots, n_{2^k}} 1 \\ &= \left| \left\{ 1 \leq n_1, \ldots, n_{2^k} \leq P \,\middle|\, n_1^k + \ldots + n_{2^{k-1}}^k = n_{2^{k-1}+1}^k + \ldots + n_{2^k}^k \right\} \right|, \end{aligned}$$

wobei die Summe wieder über dieselbe Menge läuft. Aus dem Lemma von Hua (Satz 6.19) wissen wir nun, dass dies durch

$$\int_0^1 |T(\theta)|^{2^k} \, d\theta \ll P^{2^k - k + \varepsilon}$$

mit beliebigem ε abgeschätzt werden kann. Es folgt damit auch

$$\int_0^1 |S_t(\theta)|^{2^k} \, d\theta \ll P^{2^k - k + \varepsilon}$$

und insgesamt

$$\mathcal{J}_L \ll P^{s-k-\tilde{\gamma}},$$

wobei $\tilde{\gamma} = \gamma - \varepsilon$ für hinreichend kleines ε positiv ist. Es ergibt sich daher wie gewünscht

$$\mathcal{I}_2 \ll P^{s-k-\tilde{\gamma}},$$

womit der Beweis vollständig ist. □

Bemerkung: Da sich das Integral für diese Variante des Waring'schen Problems von dem des Originalproblems offenbar nur um einen Vorfaktor und einen vernachlässigbaren Term unterscheidet, erscheint es plausibel, dass die asymptotischen Ordnungen dieselben sind. Bezeichnet also $G_{h,m}(k)$ die asymptotische Ordnung von $A_{k,h,m}$, so kann vermutet werden, dass

$$G_{h,m}(k) = G(k)$$

für $k \geq 2$ ist (für $k = 1$ ist dies offensichtlich nicht der Fall, da $G_{h,m}(1) > 1$ und $G(1) = 1$ gilt). Noch schwieriger gestaltet sich aber die Frage nach der Ordnung $g_{h,m}(k)$, die selbst für $k = 1$ ein interessantes und nichttriviales Problem darstellt. Damit $A_{k,h,m}$ eine Basis von $\mathbb{N}$ ist, muss 1 (und 0) hinzugefügt werden, sonst hat 1 selbst bereits keine Darstellung. Im Gegensatz zur asymptotischen Ordnung hängt diese jedenfalls von den weiteren Parametern h, m und q ab. Einige Bemerkungen über den Spezialfall $k = 1$ sollen Gegenstand des letzten Abschnitts sein.

8.7 Bemerkungen über die Ordnungen $G_{h,m}(1)$ und $g_{h,m}(1)$

Während für $k \geq 2$ die Vermutung $G_{h,m}(k) = G(k)$ formuliert wurde, so ist dies für $k = 1$ sicherlich nicht mehr richtig. Besagter Fall lässt sich aber auf kombinatorischem Wege einfacher fassen. In der Tat lässt sich die asymptotische Ordnung noch leicht bestimmen:

Satz 8.8 *Sind h, m, q natürliche Zahlen mit $m, q \geq 2$ und $(m, q-1) = 1$, dann gilt für die asymptotische Ordnung der Menge $A_{1,h,m}$*

$$G_{h,m}(1) = 3.$$

Beweis: Zunächst wird gezeigt, dass $G_{h,m}(1) > 2$ ist. Dazu muss die Existenz beliebig großer Zahlen nachgewiesen werden, die sich nicht als Summe zweier Elemente von $A_{1,h,m}$ schreiben lassen. Betrachte dazu die Zahl $N_l := q^l - 1$ mit $(q-1)l \not\equiv 2h \bmod m$. Wegen $(m, q-1) = 1$ gibt es sicher ein (beliebig großes) l, das diese Bedingung erfüllt. Angenommen nun, es gäbe $a, b \in A_{1,h,m}$, sodass $a + b = N_l$. Es muss also $s_q(a) \equiv s_q(b) \equiv h \bmod m$ gelten.

Seien $a = \sum_{i=0}^{r} a_i q^i$ und $b = \sum_{i=0}^{r} b_i q^i$ mit $0 \leq a_i, b_i < q$ die Ziffernentwicklungen von a und b. Ohne Beschränkung der Allgemeinheit darf angenommen werden, dass beide dieselbe Länge haben, da man sonst gegebenenfalls mit führenden Nullen auffüllen kann. Wir zeigen durch Induktion nach j, dass aus $a + b = N_j = q^j - 1$ bereits $a_i + b_i = q - 1$ für $0 \leq i \leq j-1$ und $a_i = b_i = 0$ für $i \geq j$ folgt. Für $j = 0$ ist dies trivial. Für $j \geq 1$ gilt nun

$$a_0 + b_0 \equiv a + b = q^j - 1 \equiv -1 \mod q.$$

Aus $0 \leq a_0, b_0 \leq q - 1$ folgt $0 \leq a_0 + b_0 \leq 2q - 2$. Zusammen mit $a_0 + b_0 \equiv -1 \bmod q$ ist dies nur für $a_0 + b_0 = q - 1$ möglich. Setze nun $a' := \frac{a-a_0}{q} = \sum_{i=0}^{r-1} a_{i+1} q^i$ und $b' := \frac{b-b_0}{q} = \sum_{i=0}^{r-1} b_{i+1} q^i$. Dann ergibt sich

$$a' + b' = \frac{a + b - a_0 - b_0}{q} = \frac{q^j - 1 - (j-1)}{q} = q^{j-1} - 1.$$

Mit der Induktionsvoraussetzung folgt die Behauptung. Also muss für $a, b \in A_{1,h,m}$ mit $a + b = N_l$ die Beziehung

$$(q-1)l = \sum_{i=1}^{r} a_i + \sum_{i=1}^{r} b_i = s_q(a) + s_q(b) \equiv 2h \mod m$$

gelten, was einen Widerspruch darstellt.

Nun wird gezeigt, dass jede hinreichend große Zahl N Summe von drei Elementen aus $A_{1,h,m}$ ist. Dafür benötigen wir folgendes Lemma:

Lemma 8.9 Sei $n \geq q^{m^2-1}$ eine natürliche Zahl und $0 \leq l < m$. Dann kann man a, b, c derart wählen, dass $a + b + c = n$ und $s_q(a) + s_q(b) + s_q(c) \equiv l \bmod m$.

Beweis des Lemmas: n hat nach Voraussetzung zumindest m^2 Stellen. Es sei $n = \sum_{i=0}^{r} n_i q^i$ die Ziffernentwicklung von n. Wir unterscheiden zwei Fälle: hat n (zumindest) m von Null verschiedene Ziffern $n_{i_1}, \ldots, n_{i_m}$ mit $i_1 > \ldots > i_m \geq 0$, so setze $u_{i_1-1} = \ldots = u_{i_p-1} = 1$ mit einem noch festzulegenden $0 \leq p < m$ und $u_i = 0$ für alle übrigen $i \geq -1$, insbesondere also $u_{-1} = u_r = 0$. Es ist nun einfach, $0 \leq a_i, b_i, c_i < q$ derart zu wählen, dass

$$a_i + b_i + c_i = u_i q + n_i - u_{i-1}$$

gilt, da nach Voraussetzung stets $0 \leq u_i q + n_i - u_{i-1} \leq 2q-1$ ist. Es folgt mit $a := \sum_{i=0}^{r} a_i q^i$, $b := \sum_{i=0}^{r} b_i q^i$ und $c := \sum_{i=0}^{r} c_i q^i$

$$\begin{aligned} a + b + c &= \sum_{i=0}^{r} (a_i + b_i + c_i) q^i = \sum_{i=0}^{r} (u_i q + n_i - u_{i-1}) q^i \\ &= \sum_{i=0}^{r} u_i q^{i+1} + \sum_{i=1}^{r} n_i q^i - \sum_{i=-1}^{r-1} u_i q^{i+1} = \sum_{i=1}^{r} n_i q^i = n \end{aligned}$$

und

$$\begin{aligned} s_q(a) + s_q(b) + s_q(c) &= \sum_{i=0}^{r} (a_i + b_i + c_i) = \sum_{i=0}^{r} (u_i q + n_i - u_{i-1}) \\ &= \sum_{i=0}^{r} u_i (q-1) + \sum_{i=0}^{r} n_i = p(q-1) + s_q(n). \end{aligned}$$

Wegen $(m, q-1) = 1$ kann man p so wählen, dass dies $\equiv l \mod m$ ist.

Hat n nicht zumindest m Stellen ungleich Null, so gibt es auch höchstens $m-1$ Blöcke aufeinanderfolgender Nullen. Da die Anzahl aller Stellen zumindest m^2 ist und höchstens $m-1$ davon ungleich Null sind, gibt es nach dem Schubfachschlussprinzip einen Block mit zumindest $m+1$ aufeinanderfolgenden Nullen. Es gibt also ein j derart, dass $n_j = n_{j+1} = \ldots = n_{j+m} = 0$ und $n_{j+m+1} \neq 0$. Sei $n' := \sum_{i=0}^{j-1} n_i q^i$ und $n'' = \sum_{i=j+m+1}^{r} n_i q^i - q^{j+m+1}$. Nun setzen wir für ein noch festzulegendes $0 \leq p < m$

$$\begin{aligned} a &:= n' + n'' + q^{j+m+1} - q^{j+p+2} + q^{j+1} \\ &= \sum_{i=0}^{j-1} n_i q^i + q^{j+1} + \sum_{i=j+p+2}^{j+m} (q-1) q^i + (n_{j+m+1} - 1) q^{j+m+1} + \sum_{i=j+m+2}^{r} n_i q^i, \end{aligned}$$

$$b := q^{j+p+2} - q^{j+p+1} - q^j = \sum_{i=j}^{j+p} (q-1) q^i + (q-2) q^{j+p+1},$$

$$c := q^{j+p+1} - q^{j+1} + q^j = q^j + \sum_{i=j+1}^{j+p} (q-1) q^i,$$

womit sich $a+b+c=n'+n''+q^{j+m+1}=n$ und

$$\begin{aligned} s_q(a)+s_q(b)+s_q(c) &= s_q(n')+s_q(n'')+1+(q-1)(m-p-1) \\ &\quad +(q-1)(p+1)+(q-2)+1+(q-1)p \\ &= s_q(n)+(q-1)(m+p+1) \end{aligned}$$

ergibt. Wegen $(m,q-1)=1$ kann man p wieder so wählen, dass dies $\equiv l \mod m$ ist. Damit ist das Lemma bewiesen.

Nun setzen wir im Beweis des Satzes fort. Es wird behauptet, dass jede natürliche Zahl $N \geq q^{m^2+2m}$ sich als Summe dreier Zahlen $a,b,c \in A_{1,h,m}$ schreiben lässt. Dazu definieren wir zunächst rekursiv $0 \leq d_i < q$ $(0 \leq i < 2m)$ derart, dass

$$\sum_{i=0}^{j}(1+d_i)q^i \equiv N \mod q^{j+1}$$

für $0 \leq j < 2m$ gilt. Insbesondere ist damit $N' := \frac{N-\sum_{i=0}^{2m-1}(1+d_i)q^i}{q^{2m}} \in \mathbb{N}$. Auf N' wenden wir das Lemma an: wegen

$$N' \geq \frac{N-\sum_{i=0}^{2m-1}q^{i+1}}{q^{2m}} \geq \frac{q^{m^2+2m}-q^{2m+1}}{q^{2m}} = q^{m^2}-q \geq q^{m^2-1}$$

können wir a',b',c' derart finden, dass $a'+b'+c'=N'$ und

$$s_q(a')+s_q(b')+s_q(c') \equiv 3h - \sum_{i=0}^{2m-1}(1+d_i) \mod m.$$

Nun seien noch $0 \leq s_1, s_2 < m$ derart, dass

$$s_1 + s_q(a') + \sum_{i=0}^{m-1} d_i \equiv h \mod m$$

und

$$s_2 + s_q(b') + \sum_{i=m}^{2m-1} d_i \equiv h \mod m.$$

Setze

$$a = q^{2m}a' + \sum_{i=0}^{m-1} d_i q^i + \sum_{i=m}^{m+s_1-1} q^i,$$

$$b = q^{2m}b' + \sum_{i=m}^{2m-1} d_i q^i + \sum_{i=0}^{s_2-1} q^i$$

und

$$c = q^{2m}c' + \sum_{i=s_2}^{m-1} q^i + \sum_{i=m+s_1}^{2m-1} q^i.$$

Dann gilt zunächst

$$\begin{aligned} a+b+c &= q^{2m}(a'+b'+c') + \sum_{i=0}^{2m-1}(d_i+1)q^i \\ &= q^{2m}N' + \sum_{i=0}^{2m-1}(d_i+1)q^i \\ &= N - \sum_{i=0}^{2m-1}(1+d_i)q^i + \sum_{i=0}^{2m-1}(d_i+1)q^i = N, \end{aligned}$$

aber auch (unter Verwendung der q-Additivität)

$$s_q(a) = s_q(a') + \sum_{i=0}^{m-1} d_i + s_1 \equiv h \mod m,$$

$$s_q(b) = s_q(b') + \sum_{i=m}^{2m-1} d_i + s_2 \equiv h \mod m$$

und schließlich (aufgrund der Wahl von a', b', c' und der Definition von s_1 und s_2)

$$\begin{aligned} s_q(c) &= s_q(c') + (m-s_2) + (m-s_1) \\ &\equiv 3h - s_q(a') - s_q(b') - \sum_{i=0}^{2m-1}(1+d_i) - s_1 - s_2 \\ &\equiv 3h - h - h = h \mod m. \end{aligned}$$

D.h., $a, b, c \in A_{1,h,m}$. □

Wesentlich problematischer gestaltet sich die Frage nach der der Ordnung $g_{h,m}(1)$ von $A_{1,h,m} \cup \{0,1\}$. Wie die folgenden Propositionen zeigen, hängt diese Ordnung sehr stark von allen drei Parametern h, m, q ab. Die asymptotische Formel aus Satz 8.7, in der nur die m_i auftreten, legt nahe, dass m die wichtigste Rolle zukommt. In der Tat kann aber sogar für festes m die Ordnung sehr stark variieren.

Proposition 8.10 Sei $m \geq 2$ beliebig. Dann kann man q und h derart wählen, dass $g_{h,m}(1) \geq 2^m - 2$.

Beweis: Wähle $q = 2$ und $h = 0$. Offensichtlich ist $2^m - 1 = \sum_{i=0}^{m-1} 2^i$ die kleinste natürliche Zahl, deren 2-adische Ziffernsumme zumindest gleich m ist, also auch die kleinste Zahl $\neq 0$, deren 2-adische Ziffernsumme durch m teilbar ist. Damit ist die einzige Darstellung von $2^m - 2$ als Summe von Elementen aus $A_{1,h,m} \cup \{0,1\}$

$$2^m - 2 = 1 + \ldots + 1$$

und damit $g_{h,m}(1) \geq 2^m - 2$. □

Proposition 8.11 Sei $m \geq 2$ beliebig und $r \geq 2$. Dann kann man q und h derart wählen, dass $g_{h,m}(1) \leq \lfloor \frac{m}{r} \rfloor + r$.

Beweis: Wähle q groß genug, z.B. $q = 3m + r$, und $h = r$. Ich zeige zunächst, dass der maximale Abstand aufeinanderfolgender Zahlen aus $A_{1,h,m}$ höchstens gleich $2m - 1$ ist.

Sei dazu $a \in A_{1,h,m}$. Ist die letzte Ziffer a_0 von a kleiner oder gleich $q - 1 - m$, so ist $s_q(a + m) = s_q(a) + m$ und damit $a + m \in A_{1,h,m}$. Ist andererseits $a_0 \geq q - m$, dann endet $a - a_0 + q$ auf 0, und für $0 \leq j < m < q$ gilt $s_q(a - a_0 + q + j) = s_q(a - a_0 + q) + j$, d.h. $\{s_q(a - a_0 + q + j) : 0 \leq j < m\}$ durchläuft ein vollständiges Restsystem modulo m. Es gibt daher ein $j < m$, sodass $b = a - a_0 + q + j \in A_{1,h,m}$. Der Abstand zwischen a und b ist

$$b - a = -a_0 + q + j \leq -(q - m) + q + (m - 1) = 2m - 1,$$

womit diese Behauptung bewiesen ist.

Zu jedem $N \in \mathbb{N}$ gibt es damit ein j mit $0 \leq j \leq 2m - 2$, sodass $N - j \in A_{1,h,m}$. Es genügt damit zum Beweis der Proposition nachzuweisen, dass jede ganze Zahl $n \in [0, 2m - 2]$ sich als Summe von höchstens $\lfloor \frac{m}{r} \rfloor + r - 1$ Summanden aus $A_{1,h,m}$ schreiben lässt. Dazu benötigen wir nur die drei Elemente $1, r, m + r \in A_{1,h,m}$. Ist $n < m + r$, dann lässt sich n in der Form $n = x \cdot r + y \cdot 1$ schreiben, und es gilt $x = \lfloor \frac{n}{r} \rfloor$ und $y = n - rx$. Damit ergibt sich aber

$$\begin{aligned} x + y = n - (r-1)\left\lfloor \frac{n}{r} \right\rfloor &\leq n - (r-1)\frac{n - r + 1}{r} \\ &= \frac{rn - rn + n + (r-1)^2}{r} = \frac{n+1}{r} + r - 2 \\ &\leq \frac{m + r}{r} + r - 2 = \frac{m}{r} + r - 1, \end{aligned}$$

und wegen $x + y \in \mathbb{N}$ sogar $x + y \leq \lfloor \frac{m}{r} \rfloor + r - 1$. Damit ist für diesen Fall alles bewiesen. Ist andererseits $m + r \leq n \leq 2m - 2$ (was nur für $r \leq m - 2$ überhaupt möglich ist), so kann man n als $(m + r) + x \cdot r + y \cdot 1$ mit $x = \lfloor \frac{n-m-r}{r} \rfloor$ und $y = n - m - r - rx$ schreiben. Es gilt analog zur obigen Abschätzung $1 + x + y \leq \frac{m-1}{r} + r - 2 < \frac{m}{r} + r - 1$. Damit ist die Anzahl der Summanden auch in diesem Fall durch den gewünschten Ausdruck abschätzbar. □

Korollar 8.12 Wählt man insbesondere $r = \lfloor \sqrt{m} \rfloor$, so erhält man, dass zu beliebigem $m \geq 2$ stets h und q derart existieren, dass $g_{h,m}(1) \leq 2\sqrt{m} + O(1)$.

Abschließend möchte ich noch ein einzelnes Beispiel eines expliziten Resultats für $g_{h,m}(1)$ liefern:

Proposition 8.13 Sei $h \in \{0,1\}$ und $(q-1,2) = 1$, d.h. q gerade. Dann gilt $g_{h,2}(1) = 2$.

Beweis: Offenbar ist $q_{h,2}(1) \geq 2$. Sei $N = \sum_{i=0}^{r} n_i q^i$ beliebig. Es soll gezeigt werden, dass N sich als Summe von zwei Summanden aus $A_{1,h,2} \cup \{0,1\}$ schreiben lässt. Wegen $0 \in A_{1,h,2} \cup \{0,1\}$ ist dies trivial, falls $N \in A_{1,h,2}$. Wähle s minimal mit $n_s \neq 0$. Im Fall $h = 1$ muss $s_q(N)$ gerade sein, und damit sind $s_q(N - q^s) = s_q(N) - 1$ und $s_q(q^s) = 1$ ungerade. Somit ist aber bereits eine Zerlegung $N = (N - q^s) + q^s$ mit $N - q^s, q^s \in A_{1,1,2}$ gefunden.

Es sei daher nun $h = 0$ und damit $s_q(N)$ ungerade. Wäre $s = 0$, so hätte man wegen $s_q(N-1) = s_q(N) - 1$ bereits eine Zerlegung $N = (N-1) + 1$ mit $N-1, 1 \in A_{1,0,2} \cup \{0,1\}$ gefunden. Daher darf $s > 0$ angenommen werden. Ist nun $q > 2$, so lässt sich N durch $N = (N - 2q^{s-1}) + 2q^{s-1}$ mit

$$s_q(N - 2q^{s-1}) = s_q(N) - 1 + (q-2) = s_q(N) + q - 3 \equiv 0 \mod 2$$

und $s_q(2q^{s-1}) = 2$ schreiben, womit wieder eine geeignete Zerlegung gefunden wäre. Es verbleibt der Fall $q = 2$. Wegen $s_2(N-1) = s_2(N) - 1 + s$ wäre $N = (N-1) + 1$ für gerades s eine passende Zerlegung. Daher behandeln wir nur noch ungerades s. Es bleiben zwei Fälle:

- $s > 1$: in diesem Fall sei $a = 2^{s-1} - 1 = \sum_{i=0}^{s-2} 2^i$ und $b = N - a = 1 + 2^{s-1} + \sum_{i=s+1}^{r} n_i 2^i$. Die Ziffernsummen sind dann $s_2(a) = s - 1$ und $s_2(b) = s_2(N) + 1$, also beide gerade. Wieder sind zwei passende Summanden gefunden.

- $s = 1$: 2 kann als $1 + 1$ geschrieben werden. Ist $N \neq 2$, so gibt es ein $t > 1$ mit $n_t = 1$. Setze nun $a = 2^t + 1$ und $b = N - a$. Die Ziffernsummen sind hier $s_2(a) = 2$ und $s_2(b) = s_2(N) - 1$, also erneut gerade. Damit ist der letzte Fall behandelt.

$\square$

Literaturverzeichnis

[1] G. E. Andrews. *The theory of partitions.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1998. Reprint of the 1976 original.

[2] T. M. Apostol. *Modular functions and Dirichlet series in number theory*, volume 41 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1990.

[3] P. Erdös and J. Lehner. The distribution of the number of summands in the partitions of a positive integer. *Duke Math. J.*, 8:335–345, 1941.

[4] K. B. Ford. New estimates for mean values of Weyl sums. *Internat. Math. Res. Notices*, (3):155–171 (electronic), 1995.

[5] A. O. Gel′fond. Sur les nombres qui ont des propriétés additives et multiplicatives données. *Acta Arith.*, 13:259–265, 1967/1968.

[6] P. J. Grabner and A. Knopfmacher. Analysis of some new partition statistics. 2004.

[7] G. H. Hardy and J. E. Littlewood. A new solution of Waring's problem. *Q. J. Math.*, 48:272–293, 1919.

[8] G. H. Hardy and S. Ramanujan. Asymptotic formulae in combinatory analysis. *Proc. London Math. Soc.*, 17:75–115, 1918.

[9] H. Heuser. *Lehrbuch der Analysis. Teil 1.* Mathematische Leitfäden. B. G. Teubner, Stuttgart, ninth edition, 1991.

[10] D. Hilbert. Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl n-ter Potenzen. *Mat. Annalen*, 67:281–300, 1909.

[11] L. K. Hua. On Waring's problem. *Q. J. Math.*, 9:199–202, 1938.

[12] L. K. Hua. On the number of partitions of a number into unequal parts. *Trans. Amer. Math. Soc.*, 51:194–201, 1942.

[13] L. K. Hua. *Introduction to number theory.* Springer-Verlag, Berlin, 1982. Translated from the Chinese by Peter Shiu.

[14] G. J. Janusz. *Algebraic number fields*, volume 7 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 1996.

[15] A. Kempner. Bemerkungen zum Waringschen Problem. *Mat. Annalen*, 72:387–399, 1912.

[16] A. Knopfmacher, R. F. Tichy, S. Wagner, and V. Ziegler. Graphs, Partitions and Fibonacci numbers. 2004.

[17] L. Lewin. *Polylogarithms and associated functions*. North-Holland Publishing Co., New York, 1981.

[18] X. Li, Z. Li, and L. Wang. The inverse problems for some topological indices in combinatorial chemistry. *J. Computational Biology*, 10(1):47–55, 2003.

[19] C. Mauduit and A. Sárközy. On the arithmetic structure of sets characterized by sum of digits properties. *J. Number Theory*, 61(1):25–38, 1996.

[20] R. E. Merrifield and H. E. Simmons. *Topological Methods in Chemistry*. Wiley, New York, 1989.

[21] M. B. Nathanson. *Additive number theory. The classical bases*, volume 164 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996.

[22] H. Prodinger and R. F. Tichy. Fibonacci numbers of graphs. *Fibonacci Quart.*, 20(1):16–21, 1982.

[23] H. Rademacher. On the partition function $p(n)$. *Proc. London Math. Soc.*, 43:241–254, 1937.

[24] H. Rademacher. On the expansion of the partition function in a series. *Ann. of Math. (2)*, 44:416–422, 1943.

[25] B. Richmond and G. Szekeres. On the transformation equation for the partition generating function. *Aequationes Math.*, 21(1):44–48, 1980.

[26] C. L. Siegel. A simple proof of $\eta(-1/\tau) = \eta(\tau)\sqrt{\tau/i}$. *Mathematika*, 1:4, 1954.

[27] N. J. A. Sloane and S. Plouffe. *The encyclopedia of integer sequences*. Academic Press Inc., San Diego, 1995. Online edition available at `http://www.research.att.com/~njas/sequences`.

[28] G. Szekeres. An asymptotic formula in the theory of partitions. *Quart. J. Math., Oxford Ser. (2)*, 2:85–108, 1951.

[29] G. Szekeres. Some asymptotic formulae in the theory of partitions. II. *Quart. J. Math., Oxford Ser. (2)*, 4:96–111, 1953.

[30] G. Szekeres. Asymptotic distribution of the number and size of parts in unequal partitions. *Bull. Austral. Math. Soc.*, 36(1):89–97, 1987.

[31] G. Szekeres. Asymptotic distribution of partitions by number and size of parts. In *Number theory, Vol. I (Budapest, 1987)*, volume 51 of *Colloq. Math. Soc. János Bolyai*, pages 527–538. North-Holland, Amsterdam, 1990.

[32] J. M. Thuswaldner and R. F. Tichy. Waring's problem with digital restrictions. To appear in *Israel J. Math*, 2004.

[33] R. C. Vaughan. *The Hardy-Littlewood method*, volume 80 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1981.

[34] I. M. Vinogradov. On Waring's problem. *Ann. of Math. (2)*, 36(2):395–405, 1935.

[35] I. M. Vinogradov. *Metod trigonometricheskikh summ v teorii chisel* (The Method of Trigonometric Sums in Number Theory). "Nauka", Moscow, second edition, 1980.

[36] I. M. Vinogradov and A. A. Karatsuba. The method of trigonometric sums in number theory. *Trudy Mat. Inst. Steklov.*, 168:4–30, 1984.

[37] E. Waring. *Meditationes Algebraicae*. Cambridge University Press, Cambridge, 1770.

[38] G. N. Watson. *A treatise on the theory of Bessel functions*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1995. Reprint of the second (1944) edition.

[39] H. Weyl. Über die Gleichverteilung von Zahlen mod Eins. *Mat. Annalen*, 77:313–352, 1913.

[40] A. Wieferich. Beweis des Satzes, daß sich eine jede ganze Zahl als Summe von höchstens neun positiven Kuben schreiben läßt. *Mat. Annalen*, 66:95–101, 1909.

[41] T. D. Wooley. Large improvements in Waring's problem. *Ann. of Math. (2)*, 135(1):131–164, 1992.