Gitter

Jürgen Wolfart

Skriptum zum Kapitel III der Vorlesung *Höhere Algebra und Zahlentheorie*, WS 2000/01. Buchstaben in eckigen Klammern wie [Oe] beziehen sich auf das Literaturverzeichnis am Ende oder auf das Literaturverzeichnis von [Wo]. Wichtige Hinweise verdankt der Autor Dr. Sabine Ricker, Dr. Jörn Steuding und Dr. Manfred Streit. Weitere Anregungen, Korrekturen oder Kommentare sind sehr erwünscht. Keines der angeschnittenen Themen wird auch nur annähernd vollständig behandelt, und leider fehlen wichtige Dinge wie Reduktionstheorie und die Beziehung zu Kettenbrüchen oder algebraischen Zahlen völlig. Zu hoffen ist, daß es eine Anregung zu weiterem Studium darstellt.

Inhaltsverzeichnis

1 Grundbegriffe

L heißt **Gitter** im $\mathbb{R}^n$, wenn es n linear unabhängige $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ gibt mit

$$L \;=\; \mathbb{Z}\mathbf{a}_1 + \dots + \mathbb{Z}\mathbf{a}_n \;=\; \{\sum_{i=1}^{n} a_i \mathbf{a}_i \mid a_i \in \mathbb{R}\}\,.$$

Dabei heißt $\{\mathbf{a}_i \mid i = 1, \dots, n\}$ eine **Gitterbasis** für L. Bezeichnet man mit $\mathbf{e}_i$ die Standard–Einheitsvektoren des $\mathbb{R}^n$, so heißt

$$L_0 \;:=\; \mathbb{Z}\mathbf{e}_1 + , \dots + \mathbb{Z}\mathbf{e}_n$$

das *Standardgitter* des $\mathbb{R}^n$. Äquivalente Möglichkeiten, Gitter zu definieren, bieten sich durch den folgenden

Satz 1 *1. $L \subset \mathbb{R}^n$ ist genau dann ein Gitter, wenn ein linearer Automorphismus $A : \mathbb{R}^n \to \mathbb{R}^n$ existiert mit $L = AL_0$.*

2. Jedes Gitter ist additive, diskrete Untergruppe des $\mathbb{R}^n$.

3. Jede additive, diskrete Untergruppe des $\mathbb{R}^n$, welche n linear unabhängige Vektoren enthält, ist ein Gitter.

Der *Beweis* von 1) folgt daraus, daß A durch $A\mathbf{e}_i = \mathbf{a}_i$ definiert werden kann. 2) ist klar für L_0, und der allgemeine Fall folgt daraus, daß die Transformation A in 1) gleichzeitig Gruppenisomorphismus und Homöomorphismus ist. Zum Beweis von 3) sei die Untergruppe $L \subset \mathbb{R}^n$ topologisch diskret, d.h. es existiere eine Umgebung $U = U(\mathbf{0})$ des Nullpunkts mit $U \cap L = \{\mathbf{0}\}$. Da es n linear unabhängige Vektoren in L gibt, existiert insbesondere ein $\mathbf{a}_1 \in L$ mit $\mathbf{a}_1 \neq \mathbf{0}$, $r\mathbf{a}_1 \notin L$ für alle $0 < r < 1$. Dieses ist das erste Element einer induktiv zu konstruierenden Gitterbasis. Seien linear unabhängige $\mathbf{a}_1, \dots, \mathbf{a}_{i-1} \in L$, $i \leq n$, bereits gefunden mit der Eigenschaft, daß das $(i-1)$–dimensionale Parallelepiped

$$P_{i-1} \;=\; \{\sum_{j=1}^{i-1} r_j \mathbf{a}_j \mid 0 \leq r_j < 1 \quad \text{für alle} \quad j\}$$

nur den Punkt $\mathbf{0}$ aus L enthält. Nach Voraussetzung gibt es ein von $\mathbf{a}_1, \dots, \mathbf{a}_{i-1}$ linear unabhängiges $\mathbf{a}_i \in L$. Dieses läßt sich so wählen, daß es von dem $(i-1)$–dimensionalen linearen Unterraum

$$V_{i-1} \;=\; \mathbb{R}\mathbf{a}_1 + \dots + \mathbb{R}\mathbf{a}_{i-1} \;\supset\; P_{i-1}$$

minimalen euklidischen Abstand > 0 besitzt: Andernfalls gäbe es eine Folge solcher Punkte in L, deren Abstände zu V_{i-1} gegen 0 konvergiert oder einen Wert $d > 0$, der nicht selbst als Abstand eines Gitterpunkts von V_{i-1} vorkommt. Man könnte durch Addition geeigneter

$\mathbb{Z}$–Linearkombinationen der $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$ — hier verwenden wir die Gruppeneigenschaft von L — sogar dafür sorgen, daß diese Folge gegen einen Punkt der abgeschlossenen Hülle von P_{i-1} konvergiert (Bolzano–Weierstraß) bzw. gegen einen Punkt, der von P_{i-1} einen Abstand $\leq d$ besitzt, in beiden Fällen im Widerspruch zur Diskretheit von L. Man überzeuge sich davon, daß das neu entstehende i–dimensionale Parallelepiped

$$P_i \;=\; \{\sum_{j=1}^{i} r_j \mathbf{a}_j \,|\, 0 \leq r_j < 1 \quad \text{für alle} \quad j\} \;\subset\; P$$

wieder nur den Punkt $\mathbf{0}$ aus L enthält, da andere einen kleineren Abstand von dem Unterraum V_{i-1} hätten als $\mathbf{a}_i$.
Behauptung: Die so induktiv konstruierten $\mathbf{a}_1, \ldots, \mathbf{a}_n$ bilden eine Gitterbasis für L. Die lineare Unabhängigkeit ist klar, und jedes $\mathbf{b} \in L$ besitzt eine Darstellung

$$\mathbf{b} \;=\; \sum_{i=1}^{n} s_i \mathbf{a}_i \;, \quad \text{alle} \quad s_i \in \mathbb{R} \,.$$

Durch Addition einer geeigneten $\mathbb{Z}$–Linearkombination der $\mathbf{a}_i$ dürfen wir sogar annehmen, daß o.B.d.A. $0 \leq s_i < 1$ für alle i ist — hier verwenden wir wieder die Gruppeneigenschaft von L —, daß also $\mathbf{b} \in P_n \cap L$, was aber nach Konstruktion nur aus $\mathbf{0}$ besteht. $\square$

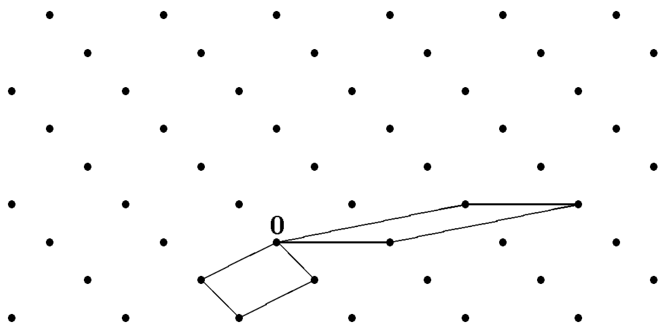

Abbildung 1: Zwei mögliche P_2 in einem Gitter $L \subset \mathbb{R}^2$

Das in diesem Beweis konstruierte P_n nennt man ein **Fundamentalparallelotop** des Gitters L. Wie die Abbildung 1 zeigt, ist es ebensowenig eindeutig durch L bestimmt wie die Gitterbasis, wohl aber sein euklidisches Volumen, wie wir gleich sehen werden. Der Name kommt daher, daß es die Rolle eines *Fundamentalbereichs* für die Operation von L auf $\mathbb{R}^n$ besitzt, d.h. die L–Translate von P_n pflastern $\mathbb{R}^n$ lückenlos und ohne Überlappungen:

$$\mathbb{R}^n \;=\; L + P_n \;=\; \bigcup_{x \in L} x + P_n \quad \text{und} \quad x + P_n \cap y + P_n \quad \text{für alle} \quad x \neq y \in L$$

Für den nun folgenden Satz, der den Wechsel zwischen verschiedenen Gitterbasen beschreibt, stelle man sich die Basis als n–tupel von Spaltenvektoren des $\mathbb{R}^n$ vor.

Satz 2 *Sei $L \subset \mathbb{R}^n$ Gitter mit Gitterbasis $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$. Das n–tupel $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ ist genau dann eine Gitterbasis von L, wenn eine ganzzahlige $n \times n$–Matrix U mit Determinante ± 1 existiert (also $\in \mathrm{GL}_n(\mathbb{Z})$, d.h.* unimodular*) mit $B = AU$.*

Beweis. Sind die $\mathbf{b}_j$ Gittervektoren, so müssen sie ganzzahlige Linearkombinationen der $\mathbf{a}_i$ sein, was bereits die Existenz einer ganzzahligen $n \times n$–Matrix U mit $B = AU$ sichert. Ist darüber hinaus B eine Gitterbasis, muß U eine ganzzahlige Inverse besitzen. Daraus folgt $\det U = \pm 1$ nach dem Determinantenmultiplikationssatz, da ± 1 die einzigen Einheiten in $\mathbb{Z}$ sind. Wenn umgekehrt $\det U = \pm 1$, so besitzt U nach der Kramerschen Regel eine ganzzahlige Inverse; damit sind alle $\mathbf{a}_i$ ganzzahlige Linearkombinationen der $\mathbf{b}_j$, folglich sogar alle Elemente von L.
□

Folgerung 3 *und* **Definition.** *Sei $L \subset \mathbb{R}^n$ Gitter mit Basis A und zugehörigem Fundamentalparallelotop P. Dessen Volumen $V(P) = |\det A| =: d(L)$ ist unabhängig von der Basiswahl und wird daher als* **Gitterdeterminante** *bezeichnet.*

2 Untergitter und Elementarteiler

2.1 Untergitter und Basiswechsel

Hilfssatz 4 *Seien $\Lambda \subseteq L$ Gitter im $\mathbb{R}^n$. Dann ist*

$$\frac{d(\Lambda)}{d(L)} =: D \in \mathbb{N}$$

und das Gitter $DL = \{D\mathbf{a} \mid \mathbf{a} \in L\}$ erfüllt $DL \subseteq \Lambda \subseteq L$.

Beweis. Sei $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ eine Basis von Λ und $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ eine Basis von L. Dann existiert eine ganzzahlige $n \times n$–Matrix V mit $B = AV$, welche offenbar $D = |\det V|$ erfüllt. Die Restklassen von $L \bmod \Lambda$ werden z.B. durch jene Gitterpunkte von L repräsentiert, welche in einem Fundamentalparallelotop P_Λ von Λ liegen, und man könnte durch einen Vergleich mit dem Volumen von P_L schon jetzt einsehen, daß die Anzahl $[L : \Lambda]$ dieser Repräsentanten gerade D ist. Jedenfalls ist DA Basis von DL, die Matrix DV^{-1} hat ebenfalls ganzzahlige Koeffizienten (Kramersche Regel), und aus $DA = B \cdot DV^{-1}$ folgt, daß DL Untergitter von Λ ist. □

Satz 5 *Seien $\Lambda \subseteq L$ Gitter im $\mathbb{R}^n$. Zu jeder Basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ von L gibt es eine Basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ von Λ mit*

$$\begin{aligned} \mathbf{b}_1 &= v_{11}\mathbf{a}_1 \\ \mathbf{b}_2 &= v_{12}\mathbf{a}_1 + v_{22}\mathbf{a}_2 \\ &\vdots \\ \mathbf{b}_n &= v_{1n}\mathbf{a}_1 + \quad \ldots \quad + v_{nn}\mathbf{a}_n \,. \end{aligned}$$

Dabei sind (natürlich) alle $v_{ij} \in \mathbb{Z}$ und alle $v_{ii} \neq 0$, o.B.d.A. sogar > 0.

Beweis. Für alle i ist $D\mathbf{a}_i \in \Lambda$, also gibt es jedenfalls Elemente $\mathbf{b}_i = v_{1i}\mathbf{a}_1 + \ldots + v_{ii}\mathbf{a}_i \in \Lambda$ mit $v_{ii} > 0$. Die $\mathbf{b}_i$ sind von dieser Form zu wählen mit kleinstmöglichem $v_{ii} > 0$. Behauptung: Dann bilden sie eine Gitterbasis von Λ. Gäbe es nämlich ein $\mathbf{c} \in \Lambda$, welches nicht in $\mathbb{Z}\mathbf{b}_1 + \ldots + \mathbb{Z}\mathbf{b}_n$ läge, so könnte man so ein $\mathbf{c}$ auch von der Form $c_1\mathbf{a}_1 + \ldots + c_k\mathbf{a}_k$, $c_k \neq 0$, wählen mit minimalem k. Division mit Rest von c_k durch v_{kk} ergibt ein

$$\mathbf{c} - s\mathbf{b}_k = (c_1 - sv_{1k})\mathbf{a}_1 + \ldots + (c_k - sv_{kk})\mathbf{a}_k \in \Lambda \,,$$

dessen k–ter Koeffizient entweder verschwindet — dann wäre k nicht minimal — oder zwischen 0 und v_{kk} liegt im Widerspruch zur Minimalität von v_{kk} bei der Konstruktion von $\mathbf{b}_k$. □

Satz 6 *Unter den gleichen Voraussetzungen kann man umgekehrt zu jeder Basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ von Λ eine Basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ von L finden, deren Übergang durch das Gleichungssystem aus Satz 5 gegeben wird.*

Zum *Beweis* schreibe man die Basis von Λ wie in Satz 2 in Matrixform als B und wende Satz 5 auf die Inklusion $DL \subseteq \Lambda$ an: Es existiert also eine Basis DA von DL von der Form $DA = BW$ mit einer oberen Dreiecksmatrix W. Daraus folgt $B = A \cdot DW^{-1}$, und dabei muß auch DW^{-1} obere Dreiecksmatrix sein. Daß sie außerdem ganzzahlig ist, folgt aus der Gitterinklusion $\Lambda \subseteq L$. □

Zusatzbemerkung. In den Sätzen 5 und 6 lassen sich die Basen so wählen, daß in den Übergangsmatrizen V für alle i und j

$$0 \leq v_{ij} < v_{ii} \quad \text{in Satz 5},$$

$$0 \leq v_{ij} < v_{jj} \quad \text{in Satz 6}.$$

In Satz 5 ist diese Eigenschaft für $i = 1$ trivial erfüllt. Sei k der kleinste Index, für den diese Ungleichungen in

$$\mathbf{b}_k = v_{1k}\mathbf{a}_1 + \ldots + v_{kk}\mathbf{a}_k$$

nicht alle gelten, und zwar sei darin i der größte Index, für den v_{ik} nicht im angegebenen Intervall liegt. Division mit Rest von v_{ik} durch v_{ii} zeigt die Existenz eines $s \in \mathbb{Z}$, welches $0 \leq v_{ik} - sv_{ii} =: v'_{ik} < v_{ii}$ erfüllt. Ersetzt man also $\mathbf{b}_k$ durch $\mathbf{b}'_k := \mathbf{b}_k - s\mathbf{b}_i$, so hat man weiterhin eine Basis von Λ, deren neuer Übergangskoeffizient v'_{ik} nun aber auch die fragliche Ungleichung erfüllt. Durch eine doppelte Induktion (aufsteigend in j, absteigend in i) läßt sich so ein Basiswechsel durchführen, daß die erste Ungleichung für alle i und j erfüllt ist. Im Fall von Satz 6 nimmt man ganz entsprechend einen Basiswechsel von A zu A' vor.

2.2 Basisergänzung

Folgerung 7 *Seien* $\mathbf{b}_1, \ldots, \mathbf{b}_m$ *linear unabhängige Elemente des Gitters* $L \subset \mathbb{R}^n$. *Dieses besitzt eine Basis* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *so daß*

$$\begin{aligned} \mathbf{b}_1 &= v_{11}\mathbf{a}_1 \\ \mathbf{b}_2 &= v_{12}\mathbf{a}_1 + v_{22}\mathbf{a}_2 \\ &\vdots \\ \mathbf{b}_m &= v_{1m}\mathbf{a}_1 + \quad \ldots \quad + v_{mm}\mathbf{a}_m \end{aligned}$$

mit ganzzahligen Übergangskoeffizienten, die alle die Ungleichungen

$$0 \leq v_{ij} < v_{jj}$$

erfüllen.

Man ergänze dazu die $\mathbf{b}_j$ durch weitere Gittervektoren $\mathbf{b}_{m+1}, \ldots, \mathbf{b}_n \in L$ zu einer Basis des $\mathbb{R}^n$. Sie bilden dann eine Basis eines Untergitters Λ, auf das man Satz 6 und die Zusatzbemerkung anwenden kann. $\square$

Folgerung 8 *Seien* $\mathbf{b}_1, \ldots, \mathbf{b}_m$ *linear unabhängige Elemente des Gitters* $L \subset \mathbb{R}^n$. *Diese lassen sich genau dann zu einer Gitterbasis von* L *ergänzen, wenn für alle* $\mathbf{c} \in L$ *mit*

$$\mathbf{c} = c_1\mathbf{b}_1 + \ldots + c_m\mathbf{b}_m \quad \textit{gilt: alle} \quad c_i \in \mathbb{Z}\,.$$

Die Notwendigkeit der angegebenen Bedingung ist klar. Daß die Bedingung auch hinreichend ist, sieht man so ein: Man nehme die Basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ des Gitters L aus Folgerung 7 und wende die Bedingung nacheinander an auf

$$\mathbf{c} = \mathbf{a}_1 = v_{11}^{-1}\mathbf{b}_1 \quad \Rightarrow \quad v_{11} = 1 \quad \Rightarrow \quad \mathbf{a}_1 = \mathbf{b}_1\,,$$

$$\mathbf{c} = \mathbf{a}_2 = v_{22}^{-1}\mathbf{b}_2 - v_{22}^{-1}v_{12}\mathbf{b}_1 \quad \Rightarrow \quad v_{22} = 1\,,\, v_{12} = 0 \quad \Rightarrow \quad \mathbf{a}_2 = \mathbf{b}_2$$

u.s.w. und zeige so $\mathbf{b}_i = \mathbf{a}_i$ für alle $i \leq m$. Die Ergänzung zu einer Gitterbasis ist dann durch die übrigen $\mathbf{a}_i$ möglich.

Folgerung 9 *Sei* $\mathbf{a}_1, \dots, \mathbf{a}_n$ *eine Basis des Gitters* $L \subset \mathbb{R}^n$ *und*

$$\mathbf{b} \;=\; b_1\mathbf{a}_1 \,+\, \dots \,+\, b_n\mathbf{a}_n \;.$$

$\mathbf{a}_1, \dots, \mathbf{a}_{m-1}, \mathbf{b}$ *lassen sich genau dann zu einer Basis von* L *ergänzen, wenn der ggT* $(b_m, \dots, b_n) = 1$ *ist.*

Klar, daß eine solche Basisergänzung nur möglich ist, wenn $\mathbf{a}_1, \dots, \mathbf{a}_{m-1}, \mathbf{b}$ linear unabhängig sind, wenn also $b_m, \dots, b_n$ nicht alle $= 0$ sind. Nun setze man in Folgerung 8

$$\mathbf{b}_1 \;=\; \mathbf{a}_1 \;,\; \dots \;,\; \mathbf{b}_{m-1} \;=\; \mathbf{a}_{m-1} \;,\; \mathbf{b}_m \;=\; \mathbf{b} \;.$$

Die Bedingung

$$\mathbf{c} \;=\; c_1\mathbf{b}_1 \,+\, \dots \,+\, c_m\mathbf{b}_m \;=\; c_1\mathbf{a}_1 \,+\, \dots \,+\, c_{m-1}\mathbf{a}_{m-1} \,+\, c_m\mathbf{b}_m \;=$$

$$=\; (c_1 + b_1c_m)\mathbf{a}_1 \,+\, \dots \,+\, (c_{m-1} + b_{m-1}c_m)\mathbf{a}_{m-1} \,+\, c_m b_m\mathbf{a}_m \,+\, \dots \,+\, c_m b_n\mathbf{a}_n \in L$$

bedeutet, daß alle Koeffizienten der letzten Zeile ganz sind. Wenn nun $(b_m, \dots, b_n) = 1$ ist, folgt daraus notwendig $c_m \in \mathbb{Z}$ und $c_1, \dots, c_{m-1} \in \mathbb{Z}$. Nach Folgerung 8 ist also $\mathbf{a}_1, \dots, \mathbf{a}_{m-1}, \mathbf{b}$ Teil einer Basis von L. Wenn andererseits $(b_m, \dots, b_n) = b > 1$ ist, läßt sich $\mathbf{c} \in L$ so wählen, daß $c_m = 1/b \notin \mathbb{Z}$, $c_i = -b_i/b$ für alle $i < m$ sind, und dann ist $\mathbf{a}_1, \dots, \mathbf{a}_{m-1}, \mathbf{b}$ nicht Teil einer Basis von L.

2.3 Elementarteilertheorie

Der Sinn der Sätze 5 und 6 läßt sich kurz so beschreiben: Man möchte sich durch geschickte Basiswahl von Gitter bzw. Untergitter eine möglichst bequeme Übergangsmatrix verschaffen. Es ist plausibel, daß man dieses Ziel am besten erreichen kann, wenn man beide Basen frei wählen kann.

Satz 10 *Seien* $\Lambda \subseteq L \subset \mathbb{R}^n$ *Gitter. Es gibt Basen* $\mathbf{b}_1, \dots, \mathbf{b}_n$ *von* Λ *und* $\mathbf{a}_1, \dots, \mathbf{a}_n$ *von* L *so, daß*

$$\mathbf{b}_j \;=\; v_j\mathbf{a}_j \;,\; v_j \in \mathbb{N} \quad \textit{für alle} \quad j \;=\; 1, \dots, n$$

ist und außerdem $v_1|v_2|\dots|v_n$.

Beweis durch Induktion über n. 1.) Für $n = 1$ ist die Aussage des Satzes trivial. Wir nehmen an, der Satz sei für Gitter $\Lambda' \subseteq L'$ in $(n-1)$–dimensionalen $\mathbb{R}$–Vektorräumen V bereits bewiesen.
2.) Man wähle die Basen von Λ und L so, daß in der durch $\mathbf{b}_j = \sum v_{ij}\mathbf{a}_i$ bestimmten Übergangsmatrix der $v_{ij} \in \mathbb{Z}$ ein $v_{ij} \neq 0$ mit möglichst kleinem Betrag vorkommt. Durch Umnumerierung und eventuellen Vorzeichenwechsel kann man erreichen, daß $i = j = 1$ und

daß nun $v_1 := v_{11} > 0$ ist. Der Übergang von der einen zur anderen Basis wird dann durch das Gleichungssystem

$$\begin{aligned}
\mathbf{b}_1 &= v_{11}\mathbf{a}_1 + v_{21}\mathbf{a}_2 + \ldots + v_{n1}\mathbf{a}_n \\
\mathbf{b}_2 &= v_{12}\mathbf{a}_1 + v_{22}\mathbf{a}_2 + \ldots + v_{n2}\mathbf{a}_n \\
&\vdots \\
\mathbf{b}_n &= v_{1n}\mathbf{a}_1 + v_{2n}\mathbf{a}_2 + \ldots + v_{nn}\mathbf{a}_n
\end{aligned}$$

beschrieben, in dem alle $|v_{ij}| = 0$ oder $\geq v_{11} > 0$ sind.

3.) Behauptung: $v_1 | v_{j1}$ für alle j. Wäre dies z.B. für $j = 2$ falsch, so gäbe es ein $s \in \mathbb{Z}$ mit $0 < v_{21} - sv_1 < v_1$. Ersetzt man nun $\mathbf{a}_1$ durch $\mathbf{a}'_1 = \mathbf{a}_1 + s\mathbf{a}_2$ (man überzeuge sich, daß dabei die Basiseigenschaft erhalten bleibt), so hat man die erste Zeile des Gleichungssystems zu ersetzen durch

$$\mathbf{b}_1 = v_1\mathbf{a}'_1 + (v_{21} - sv_1)\mathbf{a}_2 + \ldots$$

im Widerspruch zur Minimalität von v_1. Wendet man die gleiche Idee im Fall $v_1 | v_{21}$ an mit $s = v_{21}/v_1$, so verschwindet nun der Koeffizient von $\mathbf{a}_2$, und entsprechend ergibt sich für die ganze erste Zeile: Die Basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ von L läßt sich sogar so wählen, daß $v_{21} = \ldots = v_{n1} = 0$ sind.

4.) Behauptung: $v_1 | v_{1j}$ für alle j. Wie eben führe man eine Division mit Rest der v_{1j} durch v_1 durch. Wenn z.B. $0 \leq v_{12} - sv_1 < v_1$, ersetze $\mathbf{b}_2$ durch $\mathbf{b}'_2 = \mathbf{b}_2 - s\mathbf{b}_1$. Nun beginnt die zweite Zeile des neuen Systems von Übergangsgleichungen mit

$$\mathbf{b}'_2 = (v_{12} - sv_1)\mathbf{a}_1 + \ldots ,$$

und der erste Koeffizient muß wieder wegen der Minimalität von v_1 verschwinden. Entsprechendes gilt für die übrigen v_{1j}, man darf also o.B.d.A. annehmen, daß $v_{1j} = 0$ für alle $j > 1$ ist.

5.) Behauptung: Es ist $v_1 = v_{11} | v_{ij}$ für alle $i, j > 1$. Wäre dies nämlich für ein Indexpaar (i, j) falsch, so gäbe es ein $s \in \mathbb{Z}$ mit $0 < v_{ij} - sv_1 < v_1$, und man könnte $\mathbf{a}_1$ durch $\mathbf{a}'_1 = \mathbf{a}_1 + s\mathbf{a}_i$ und gleichzeitig $\mathbf{b}_j$ durch $\mathbf{b}'_j = \mathbf{b}_j + \mathbf{b}_1$ ersetzen (man überzeuge sich wieder, daß dabei die Basiseigenschaft erhalten bleibt) und hätte in der neuen Transformationsgleichung für $\mathbf{b}'_j$ an der i–ten Stelle den Koeffizienten $v_{ij} - sv_1$ im Widerspruch zur Minimalität von v_1.

6.) Man darf also annehmen, daß der Basiswechsel von L zu Λ durch das folgende Gleichungssystem beschrieben wird:

$$\begin{aligned}
\mathbf{b}_1 &= v_1\mathbf{a}_1 \\
\mathbf{b}_2 &= \phantom{v_1\mathbf{a}_1} v_{22}\mathbf{a}_2 + v_{32}\mathbf{a}_3 + \ldots + v_{n2}\mathbf{a}_n \\
\mathbf{b}_3 &= \phantom{v_1\mathbf{a}_1} v_{23}\mathbf{a}_2 + v_{33}\mathbf{a}_3 + \ldots + v_{n3}\mathbf{a}_n \\
&\vdots \\
\mathbf{b}_n &= \phantom{v_1\mathbf{a}_1} v_{2n}\mathbf{a}_2 + v_{3n}\mathbf{a}_3 + \ldots + v_{nn}\mathbf{a}_n ,
\end{aligned}$$

dabei teilt v_1 alle v_{ij}. Die Gittervektoren $\mathbf{a}_2, \ldots, \mathbf{a}_n$ erzeugen ein Gitter L' in einem $(n-1)$–dimensionalen Unterraum V des $\mathbb{R}^n$, und dieses enthält ein von $\mathbf{b}_2, \ldots, \mathbf{b}_n$ erzeugtes Untergitter Λ', auf das man die Induktionsannahme anwenden kann. □

Die Rolle des in Hilfssatz 4 eingeführten Quotienten der Gitterdeterminanten wird durch Satz 10 nun deutlicher:

Folgerung 11 *Seien $\Lambda \subseteq L$ Gitter mit den Basen aus Satz 10. Dann ist*

$$D = \frac{d(\Lambda)}{d(L)} = v_1 v_2 \cdot \ldots \cdot v_n = [L : \Lambda]$$

der Index von Λ in L. Ein Repräsentantensystem der Restklassen von L/Λ wird durch

$$\{\, \mathbf{c} = c_1\mathbf{a}_1 + \ldots + c_n\mathbf{a}_n \mid 0 \le c_i < v_i \quad \textit{für alle} \quad i \,\}$$

gegeben und es ist

$$L/\Lambda \;\cong\; \mathbb{Z}/v_1\mathbb{Z} \times \ldots \times \mathbb{Z}/v_n\mathbb{Z}\,.$$

2.4 Endlich erzeugte abelsche Gruppen

Die in Satz 10 und Folgerung 11 aufgetretenen *Elementarteiler* sind bereits in [Wo], Satz 2.38, aufgetreten. Diesen Satz über die Klassifikation endlicher abelscher Gruppen kann man nun in allgemeinerem Rahmen beweisen.

Satz 12 *und* **Definition.** *Jede endlich erzeugte abelsche Gruppe A ist isomorph zu einem direkten Produkt*

$$\mathbb{Z}/v_1\mathbb{Z} \times \ldots \times \mathbb{Z}/v_n\mathbb{Z} \times \mathbb{Z}^{m-n}\,,$$

dabei ist $v_1|v_2|\ldots|v_n$. Diese v_i lassen sich o.B.d.A. > 1 wählen und sind dann ebenso wie der (torsionsfreie) **Rang** *$m-n$ von A eindeutig bestimmt.*

Zum *Beweis* nehmen wir an, A besitze m Erzeugende. Dann gibt es einen surjektiven Homomorphismus $h : \mathbb{Z}^m \to A$ mit einem Kern K, der einen n–dimensionalen Untervektorraum V des $\mathbb{R}^n$ erzeugt. $\mathbb{Z}^m \cap V$ ist ein Gitter L mit Untergitter $\Lambda = V \cap K$, auf das sich Folgerung 11 anwenden läßt. Nach Konstruktion von L läßt sich Folgerung 8 anwenden, um eine beliebige Basis von L zu einer Basis von $\mathbb{Z}^m$ zu ergänzen, natürlich durch $m-n$ Elemente, deren h–Bilder in A eine torsionsfreie Untergruppe erzeugen. Bei dieser Konstruktion kann es vorkommen, daß $v_1 = \ldots = v_k = 1$ ist und erst $v_{k+1} > 1$ (wenn man nämlich mit überflüssigen Erzeugenden von A gearbeitet hat). Es ist klar, daß die entsprechenden Faktoren aus A trivial sind und weggelassen werden können. Daß die verbleibenden dann eindeutig bestimmt sind, läßt sich ebenso wie im Beweis von [Wo], Satz 2.38, zeigen, indem man Schritt für Schritt zyklische Untergruppen maximaler Ordnung aus A herausfaktorisiert. Letztlich bleibt eine *torsionsfreie* abelsche Gruppe übrig, d.h. ohne Elemente endlicher Ordnung außer der 0, welche mit der gleichen Begründung wie eben zu einem Gitter $\mathbb{Z}^r$ isomorph ist. Daß der Rang r dieses Gitters unter Isomorphismen invariant ist, folgt z.B. aus der Invarianz der Dimension: Jeder solche Isomorphismus induziert einen Isomorphismus des umgebenden $\mathbb{R}^r$. □

3 Der Minkowskische Gitterpunktsatz

3.1 Konvexe Körper

Eine Untermenge H des $\mathbb{R}^n$ heißt **konvex**, wenn zu je zwei Punkten von H auch ihre Verbindungsstrecke zu H gehört, also wenn mit $\mathbf{x}, \mathbf{y} \in H$ auch $\mathbf{x} + r(\mathbf{y} - \mathbf{x}) \in H$ ist für alle $r \in [0, 1]$. Wenn H nicht in einer Hyperebene enthalten ist, heißt H **konvexer Körper**. Die folgende Eigenschaft ist leicht zu beweisen, indem man aus den Richtungsvektoren $\mathbf{x} - \mathbf{x}_0$ von einem festen Punkt $\mathbf{x}_0 \in H$ zu allen anderen Punkten $\mathbf{x} \in H$ eine Basis des $\mathbb{R}^n$ auswählt:

Hilfssatz 13 *Eine konvexe Menge $H \subseteq \mathbb{R}^n$ ist genau dann ein konvexer Körper, wenn sie ein nichtleeres Inneres besitzt.* □

Beispiele konvexer Körper sind Würfel, Parallelotope, Kugeln, Ellipsoide, Halbräume, alle Durchschnitte (mit nichtleerem Inneren) von konvexen Körpern, insbesondere *Polytope*, d.h. nichtleere Durchschnitte endlich vieler Halbräume. Für alle konvexen Mengen H und alle affinen Transformationen A ist auch $A(H)$ konvexe Menge. Es läßt sich zeigen, daß beschränkte konvexe Mengen (Jordan–) meßbar sind und unbeschränkte konvexe Körper unendliches Volumen haben; wir werden diese Sachverhalte hier nicht beweisen, zumal sie in wichtigen Spezialfällen, beispielsweise für kompakte und offene konvexen Mengen, offensichtlich sind. Das Volumen wird wie schon bei Parallelotopen mit V bezeichnet werden. **Symmetrische konvexe Körper** sind dadurch ausgezeichnet, daß sie mit jedem Punkt $\mathbf{x}$ auch den Punkt $-\mathbf{x}$ enthalten, also bezüglich des Nullpunkts $\mathbf{0}$ punktsymmetrisch sind.

3.2 Spezialfall: das ebene Standardgitter

Gibt es ganzzahlige Lösungen der Gleichung $2x^2 + 10xy + 13y^2 = 1$? Da diese quadratische Form positiv definit ist, kann man sie in ein Problem über Gitterpunkte umformulieren: Hat das Standardgitter $L_0 \subset \mathbb{R}^n$ mit der Ellipse

$$2x^2 + 10xy + 13y^2 < 2$$

außer $\mathbf{0}$ noch einen weiteren Gitterpunkt gemeinsam? Die Antwort könnte man in diesem Fall etwas mühsam vermöge quadratischer Ergänzung geben. Die Frage erlaubt aber die Anwendung eines sehr viel allgemeineren Kriteriums, das sich auf alle symmetrischen konvexen Körper anwenden läßt und nur noch deren Volumen benutzt (hier: 2π). Es ist der erstmals 1892 bewiesene Gitterpunktsatz von MINKOWSKI (1864–1909), den wir zunächst für das Standardgitter $L_0 = \mathbb{Z}^2$ des $\mathbb{R}^2$ vorstellen.

Satz 14 *Sei $K \subset \mathbb{R}^2$ ein symmetrischer konvexer Körper mit Volumen $V(K) > 4$. Dann liegt in K ein Punkt $\neq \mathbf{0}$ des Gitters L_0.*

Unter den vielen bis heute bekannten *Beweisen* des Satzes ist der nun folgende von MORDELL besonders elegant. Für alle $t \in \mathbb{N}$ sei $N(t)$ die Anzahl der Punkte in $\frac{2}{t}L_0 \cap K$. Da K Jordan–meßbar ist, erhält man

$$\lim_{t\to\infty} \frac{4}{t^2} N(t) \ > \ 4 \,,$$

also $N(t) > t^2$ für alle hinreichen großen $t > t_0 \in \mathbb{N}$. Die Restklassengruppe L_0/tL_0 enthält nur t^2 Elemente, also gibt es für $t > t_0$ unter den $N(t)$ Punkten von $\frac{2}{t}L_0 \cap K$ sicher Punkte $\mathbf{x} \neq \mathbf{y}$ mit

$$\mathbf{x} - \mathbf{y} \in \frac{2}{t} tL_0 \,,$$

oder — anders gesagt — mit

$$\mathbf{0} \neq \frac{1}{2}(\mathbf{x} - \mathbf{y}) \in L_0 \,.$$

Wegen der Symmetrie von K ist auch $-\mathbf{y} \in K$, und aus der Konvexität von K folgt daher auch

$$\frac{1}{2}(\mathbf{x} - \mathbf{y}) \in K \,. \quad \square$$

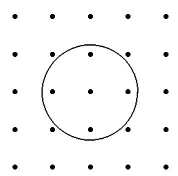

Abbildung 2: Standardgitter mit Kreis vom Radius 1.2, $V(K) = 1.44\pi > 4$

Daß der MINKOWSKIsche Gitterpunktsatz scharf ist, kann man an dem (offenen) Parallelogramm der nächsten Abbildung erkennen: Hier hat K das Volumen 4 und enthält keinen Gitterpunkt außer $\mathbf{0}$.

3.3 Der Satz von MINKOWSKI – VAN DER CORPUT

Man könnte den obenstehenden Beweis von MORDELL ohne weiteres auf höhere Dimensionen übertragen. Zur Abwechslung sei eine andere Methode vorgestellt, die zu einem allgemeineren Satz führt. Sie beruht auf dem folgenden Satz von BIRKHOFF und BLICHFELDT.

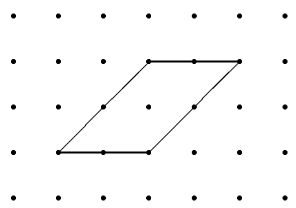

Abbildung 3: Standardgitter mit Parallelogramm K der Fläche 4

Satz 15 *Sei $L \subset \mathbb{R}^n$ ein Gitter, $S \subseteq \mathbb{R}^n$ eine meßbare Menge, $m \in \mathbb{N}$, und*

$$V(S) > m\, d(L) \qquad \text{oder} \qquad V(S) = m\, d(L)\,,\ S \quad \text{kompakt.}$$

Dann existieren paarweise verschiedene $\mathbf{x}_1, \ldots, \mathbf{x}_{m+1} \in S$, so daß alle Differenzen $\mathbf{x}_i - \mathbf{x}_j$ in L liegen.

Beweis. Wegen Satz 1.1 und Folgerung 3 genügt es, den Satz für das Standardgitter L_0 mit $d(L_0) = 1$ zu beweisen. P sei ein Fundamentalparallelotop für L_0, dann gibt es zu jedem $\mathbf{x} \in \mathbb{R}^n$ eindeutig bestimmte $\mathbf{y} \in L_0$ und $\mathbf{z} \in P$ mit $\mathbf{x} = \mathbf{y} + \mathbf{z}$. Für alle $\mathbf{y} \in L_0$ sei

$$R(\mathbf{y}) := \{\mathbf{z} \in P \,|\, \mathbf{y} + \mathbf{z} \in S\}\,, \quad \text{also} \quad \sum_{\mathbf{y} \in L_0} V(R(\mathbf{y})) = V(S)\,.$$

Wenn nun $V(S) > m$ ist, muß wegen $R(\mathbf{y}) \subseteq P$ ein $\mathbf{z}_0$ existieren, welches in mindestens $m+1$ Mengen $R(\mathbf{y})$ enthalten ist. Es gibt also paarweise verschiedene $\mathbf{y}_1, \ldots, \mathbf{y}_{m+1} \in L_0$ mit $\mathbf{x}_i = \mathbf{z}_0 + \mathbf{y}_i \in S$ für alle i. Diese $\mathbf{x}_i$ erfüllen diese Behauptung.
Nun sei S kompakt mit $V(S) = m$, dazu B_k die kompakte Kugel $||\mathbf{x}|| \leq \frac{1}{k}$ (euklidische Norm). Dann existiert für alle $k \in \mathbb{N}$ ein $(m+1)$–Tupel paarweise verschiedener Punkte $\mathbf{x}_{1k}, \ldots, \mathbf{x}_{m+1,k}$ in $S_k := S + B_k$, deren Differenzen alle im Gitter liegen. Da $S = \bigcap_k S_k$ und alle S_k kompakt sind, gibt es eine Teilfolge dieser $(m+1)$–Tupel, die gegen ein $(m+1)$–Tupel $(\mathbf{x}_1, \ldots, \mathbf{x}_{m+1}) \in S^{m+1}$ konvergieren (der Einfachheit halber führen wir für diese Teilfolge keine Doppelindizes ein). Weil die Differenzen $\mathbf{x}_{ik} - \mathbf{x}_{jk}$ stets in L_0 liegen und für $k \to \infty$ ebenfalls konvergieren müssen, sind diese Differenzenfolgen für hinreichend große k konstant und insbesondere $\neq \mathbf{0}$, die Limites $\mathbf{x}_i$ also paarweise verschieden. □

Satz 16 *Sei $m \in \mathbb{N}$, $L \subset \mathbb{R}^n$ ein Gitter und $K \subseteq \mathbb{R}^n$ ein symmetrischer konvexer Körper mit*

$$V(K) > m\, 2^n\, d(L) \qquad \text{oder} \qquad V(K) = m\, 2^n\, d(L)\,, \quad K \ \text{kompakt.}$$

Dann enthält $L \cap K$ außer $\mathbf{0}$ noch mindestens m paarweise verschiedene Punktepaare $\pm\mathbf{y}_1, \ldots, \pm\mathbf{y}_m$.

Beweis. $S = \frac{1}{2}K$ erfüllt nämlich die Voraussetzungen von Satz 15. Es gibt also $m+1$ paarweise verschiedene Punkte

$$\frac{1}{2}\mathbf{x}_i \in S \quad \text{mit} \quad \frac{1}{2}\mathbf{x}_i - \frac{1}{2}\mathbf{x}_j \in L \quad \text{für alle} \quad i, j\,.$$

Definiert man $\mathbf{y}_j := \frac{1}{2}\mathbf{x}_{m+1} - \frac{1}{2}\mathbf{x}_j$, so sind $\mathbf{0}, \pm\mathbf{y}_1, \dots, \pm\mathbf{y}_m$ paarweise verschiedene Gitterpunkte, und ebenso wie im Beweis von Satz 14 folgt aus Symmetrie und Konvexität, daß sie alle zu K gehören. $\square$

4 Anwendungen des Gitterpunktsatzes

4.1 Der Linearformensatz von Minkowski

Satz 17 *Sei $L \subset \mathbb{R}^n$ ein Gitter, $A = (a_{ij})$ eine reelle $n \times n$–Matrix und $c_1 > 0, \dots, c_n > 0$. Wenn*

$$c_1 \cdot \ldots \cdot c_n \;\geq\; |\det A|\, d(L)$$

ist, so existiert ein Gittervektor $\mathbf{u} = (u_1, \dots, u_n) \in L$, $\mathbf{u} \neq \mathbf{0}$, mit

$$|\sum a_{1j}u_j| \;\leq\; c_1\,, \qquad |\sum a_{ij}u_j| \;<\; c_j \quad \text{für } j \;=\; 2, \dots, n\,.$$

Zum *Beweis* wähle man den symmetrischen konvexen Körper

$$K_0 \;:=\; \{\,\mathbf{y} = (y_1, \dots, y_n)\,|\,|y_1| \leq c_1,\, |y_2| < c_2,\, \dots,\, |y_n| < c_n\,\}$$

vom Volumen $2^n c_1 \cdot \ldots \cdot c_n$. Faßt man A als lineare Transformation auf, so ist auch das Urbild

$$K \;:=\; \{\,\mathbf{u}\,|\,A\mathbf{u} \in K_0\,\}$$

ein symmetrischer konvexer Körper vom Volumen $2^n c_1 \cdot \ldots \cdot c_n |\det A|^{-1}$ (der Wert ∞ ist zugelassen). Wenn nun $c_1 \cdot \ldots \cdot c_n > |\det A| d(L)$, so ist die Aussage des Satzes richtig nach dem Minkowskischen Gitterpunktsatz (Satz 16 mit $m = 1$). Steht anstelle von „$>$" das Gleichheitszeichen, so ist die Aussage jedenfalls für beliebige $c_1 + \frac{1}{k}$, $k \in \mathbb{N}$, anstelle von c_1 richtig. Man erhält also die Existenz von nichttrivialen Gitterpunkten $\mathbf{u}_k$ in diesen — in y_1–Richtung etwas verbreiterten — konvexen Körpern. Da in diesem Fall $\det A \neq 0$ ist, sind diese Körper beschränkt, und zumindest eine Teilfolge (die dann wie im Beweis von Satz 15 konstant wird) konvergiert gegen ein $\mathbf{u} \in L \cap K_0$, $\mathbf{u} \neq \mathbf{0}$. $\square$

Daß für alle Linearformen mit einer Ausnahme „$<$" anstelle von „$\leq$" stehen darf, hat Hurwitz bemerkt. Weiter läßt sich der Satz aber nicht verschärfen: Wählt man für A die Einheitsmatrix, so hat das Ungleichungssystem $|u_j| < 1$ für alle j im Standardgitter L_0 nur die triviale Lösung.

4.2 Werte quadratischer Formen

Die zu Beginn des Abschnitts 3.2 gegebene Motivation läßt sich verallgemeinern:

Satz 18

$$Q(\mathbf{x}) \;=\; Q(x_1,\dots,x_n) \;=\; \sum_{i,j=1}^{n} s_{ij}x_ix_j \qquad (\,s_{ij}=s_{ji}\in\mathbb{R}\,)$$

sei eine positiv definite quadratische Form der Diskriminante $D := \det(s_{ij})$ *. Dann gibt es* $u_1,\dots,u_n \in \mathbb{Z}$ *, nicht alle* $=0$ *, mit der Eigenschaft*

$$Q(u_1,\dots,u_n) \;\le 4\,(\kappa_n^{-2}D)^{1/n} \;=\; \frac{4}{\pi}\sqrt[n]{\Gamma(1+\frac{n}{2})^2 D}\,.$$

Dabei bezeichnet Γ die Gammafunktion und $\kappa_n = \pi^{n/2}\cdot\Gamma(1+\frac{n}{2})^{-1}$ das Volumen der n–dimensionalen Einheitskugel. Zum *Beweis* beachte man, daß für jedes $\lambda > 0$ der Körper

$$K \;:=\; \{\,\mathbf{x}\in\mathbb{R}^n \,|\, Q(\mathbf{x}) \le \lambda\,\}$$

ein symmetrisches konvexes kompaktes Ellipsoid vom Volumen $V(K) = \kappa_n\lambda^{n/2}D^{-1/2}$. Nach dem MINKOWSKIschen Gitterpunktsatz (16 mit $m=1$) enthält K einen nichttrivialen Punkt des Standardgitters L_0 jedenfalls dann, wenn $V(K)\ge 2^n$ ist, also $\lambda \ge 4\sqrt[n]{\kappa_n^{-2}D}$. □

4.3 Primzahlen als Summe zweier Quadrate

Der MINKOWSKIsche Gitterpunktsatz verschafft uns einen neuen Zugang zu der schon in [Wo], Folgerung 4.29, mit völlig anderen Methoden diskutierten Frage nach der arithmetischen Natur von Quadratsummen.

Hilfssatz 19 *Seien* $n,m,k_1,\dots,k_m \in \mathbb{N}$ *und* $a_{ij}\in\mathbb{Z}$ *für alle* $i=1,\dots,m$ *,* $j=1,\dots,n$ *. Dann ist*

$$L \;:=\; \{\mathbf{u}\in L_0 \,|\, \sum_{j=1}^{n} a_{ij}u_j \equiv 0 \bmod k_i \quad \textit{für alle} \quad i=1,\dots,m\}$$

ein Gitter mit $d(L) \le k_1\cdot\ldots\cdot k_m$ *.*

Beweis. L ist Untergruppe von L_0 und enthält das Untergitter $k_1\cdot\ldots\cdot k_mL_0$. Aus Satz 1.3 folgt daher, daß L selbst Gitter ist. Folgerung 11 sagt $d(L) = [L_0 : L]$, und für alle $\mathbf{u},\mathbf{v}\in L_0$ gilt die Äquivalenz

$$\mathbf{u}\equiv\mathbf{v} \bmod L \quad\Leftrightarrow\quad \sum_{j=1}^{n} a_{ij}u_j \equiv \sum_{j=1}^{n} a_{ij}v_j \bmod k_i \quad \text{für alle} \quad i=1,\dots,m\,.$$

Die Restklassen $\mathbf{u}+L\in L_0/L$ sind also charakterisiert durch die m–Tupel von Restklassen $\sum_{j=1}^{n} a_{ij}u_j \bmod k_i$, die Anzahl aller dieser Restklassen ist also $\le k_1\cdot\ldots\cdot k_m$. □

Satz 20 *Jede Primzahl* $p \equiv 1 \bmod 4$ *ist Summe von zwei Quadraten in* $\mathbb{Z}$ *.*

(Daß Primzahlen $p \equiv 3 \bmod 4$ keine solchen Quadratsummen sind, und was für die Darstellbarkeit beliebiger natürlicher Zahlen als Summen zweier Quadrate daraus folgt, wurde bereits in Abschnitt 4.6.3 erörtert.) Zum *Beweis* sei zunächst daran erinnert, daß $(\frac{-1}{p}) = 1$ ist, also $i^2 + 1 \equiv 0 \bmod p$ eine Lösung $i \in \mathbb{Z}$ besitzt. Nach Hilfssatz 19 ist

$$L \; := \; \{\mathbf{u} = (u_1, u_2) \in L_0 \,|\, u_2 \equiv i u_1 \bmod p\}$$

ein Untergitter von L_0 mit Determinante $d(L) \le p$.

$$K \; := \; \{\, \mathbf{x} \in \mathbb{R}^2 \,|\, x_1^2 + x_2^2 < 2p \,\}$$

ist ein konvexer symmetrischer Körper mit Flächeninhalt $V(K) = 2\pi p > 4p \ge 4d(L)$, nach dem Minkowskischen Gitterpunktsatz gibt es also einen nichttrivialen Gitterpunkt in K, mit anderen Worten eine ganzzahlige Lösung von $0 < u_1^2 + u_2^2 < 2p$. Wegen $u_1^2 + u_2^2 \equiv u_1^2(1 + i^2) \equiv 0 \bmod p$ kann diese nur von der Form $u_1^2 + u_2^2 = p$ sein. $\square$

4.4 Darstellbarkeit durch Summen von vier Quadraten

Satz 21 (Lagrange) *Jede natürliche Zahl ist Summe von vier Quadraten ganzer Zahlen.*

Den *Beweis* kann man natürlich auf die Darstellbarkeit *quadratfreier* $m \in \mathbb{N}$ beschränken, d.h. man darf $m = p_1 \cdot \ldots \cdot p_g$ annehmen mit paarweise verschiedenen Primfaktoren p_i. Man überlege sich zunächst, daß für jede Primzahl p die Kongruenz

$$a_p^2 \; + \; b_p^2 \; + \; 1 \; \equiv \; 0 \bmod p$$

durch ganze Zahlen a_p, b_p erfüllt werden kann. Für $p = 2$ ist diese Behauptung evident, und für $p > 2$ durchlaufen a_p^2 und $-b_p^2 - 1 \bmod p$ jeweils genau $(p+1)/2$ verschiedene Restklassen, sie müssen also mindestens eine gemeinsame Restklasse enthalten. Diese ergibt die Lösung.
Für diese a_p, b_p betrachtet man nun nach einer Idee von Davenport das System linearer Kongruenzen

$$LK_p \; : \quad u_1 \; \equiv \; a_p u_3 + b_p u_4 \; , \quad u_2 \; \equiv \; b_p u_3 - a_p u_4 \bmod p$$

und definiert damit das Gitter

$$L \; := \; \{\mathbf{u} = (u_1, u_2, u_3, u_4) \in L_0 \subset \mathbb{R}^4 \,|\, LK_p \quad \text{für alle} \quad p = p_1, \ldots, p_g \} \; ,$$

das nach Hilfssatz 19 eine Gitterdeterminante $d(L) \le p_1^2 p_2^2 \cdot \ldots \cdot p_g^2 = m^2$ hat. Nach dem Gitterpunktsatz liegt in dem symmetrischen konvexen Körper $K = \{\mathbf{x} \in \mathbb{R}^4 | x_1^2 + \ldots + x_4^2 < 2m\}$ vom Volumen

$$V(K) \; = \; \frac{1}{2}\,\pi^2\,(2m)^2 \; > \; 2^4 m^2 \; \ge \; 2^4\, d(L)$$

ein nichttrivialer Gitterpunkt $\mathbf{u}$, der also $u_1^2+u_2^2+u_3^2+u_4^2 < 2m$ erfüllt und die Kongruenzen

$$u_1^2+u_2^2+u_3^2+u_4^2 \equiv (a_p^2+b_p^2+1)u_3^2 + (a_p^2+b_p^2+1)u_4^2 \equiv 0 \bmod p$$

für alle $p = p_1, \ldots, p_g$, somit auch

$$u_1^2+u_2^2+u_3^2+u_4^2 \equiv 0 \bmod m ,$$

was zusammen mit der Größenabschätzung und $\mathbf{u} \neq \mathbf{0}$ nur für $u_1^2+u_2^2+u_3^2+u_4^2 = m$ möglich ist. □

Ähnlich wie bei der Darstellbarkeit durch zwei Quadrate ist auch der hier vorgestellte Beweis nicht der einzige. Wie so oft erhält man mit mehr methodischem Aufwand reicheren Ertrag, beispielsweise mit *Thetafunktionen*, einem mächtigen analytischen Hilfsmittel der Gittertheorie, oder mit Hilfe von *Quaternionen*: Die Anzahl der verschiedenen $\mathbf{u} \in L_0 = \mathbb{Z}^4$, welche zu Darstellungen von m als Summe von vier Quadraten führen, läßt sich berechnen als

$$8 \cdot \sum_{d|m, 4 \nmid d} d ,$$

also die achtfache Summe der Teiler $d \not\equiv 0 \bmod 4$ von m.

5 Das Kreis– und Kugelproblem

5.1 Problemstellung

Im Gitterpunktsatz geht es darum, daß bereits in relativ kleinen konvexen Körpern mindestens ein Gitterpunkt $\mathbf{u} \neq \mathbf{0}$ liegt, aber schon in der Beweisidee von MORDELL spielt die Anzahl der Gitterpunkte in sehr großen Körpern (genauer: im gleichen Körper, aber für Gitter mit immer kleinerer Masche) eine wichtige Rolle. Was wir dort eigentlich benutzt haben, ist folgende Tatsache, die aus der Meßbarkeit der n–dimensionalen Einheitskugel folgt: Das Volumen $V_n(t)$ der n–dimensionalen Kugel

$$K(t) := \{ \mathbf{x} \in \mathbb{R}^n \,|\, x_1^2 + \ldots + x_n^2 \leq t \}$$

vom Radius $\sqrt{t}$ ist asymptotisch (für $t \to \infty$) gleich der Anzahl $A_n(t)$ der in ihr enthaltenen Punkte des Standardgitters $L_0 = \mathbb{Z}^n$, d.h. $\lim_{t\to\infty} A_n(t)/V_n(t) = 1$. Dabei ist

$$V_n(t) = \kappa_n \cdot t^{n/2} = \pi^{n/2} \cdot \Gamma(1+\frac{n}{2})^{-1} \cdot t^{n/2}$$

mit dem schon in Abschnitt 4.2 erwähnten Volumen κ_n der n–dimensionalen Einheitskugel. Wie groß ist in dieser Asymptotik das *Restglied*, d.h. der Fehler $A_n(t) - V_n(t)$?

5.2 Eine einfache obere Abschätzung

Satz 22

$$A_n(t) \; = \; V_n(t) \; + \; O(t^{(n-1)/2}) \; ,$$

d.h. es existiert eine Konstante C *, so daß* $|A_n(t) - V_n(t)| \le C t^{(n-1)/2}$ *für alle genügend großen* t *ist.*

(Die O–Schreibweise hatten wir schon in Abschnitt 1.4.2 eingeführt.) Zum *Beweis* seien $P + \mathbf{x}$, $\mathbf{x} \in L_0$ die Translate des Standard–Fundamentalquaders $P = P_n$ (vgl. Abschnitt 1), das der Standardbasis $\mathbf{e}_i$, $i = 1, \ldots n$, zugeordnet ist. $A_n(t)$ ist gleichzeitig das Volumen von

$$P(t) \; := \; \bigcup_{\mathbf{x} \in K(t)} P + \mathbf{x} \; ,$$

und weil P den euklidische Durchmesser $\sqrt{n}$ besitzt, zeigt die Dreiecksungleichung für $t > n$

$$K((\sqrt{t} - \sqrt{n})^2) \; \subset \; P(t) \; \subset \; K((\sqrt{t} + \sqrt{n})^2) \; .$$

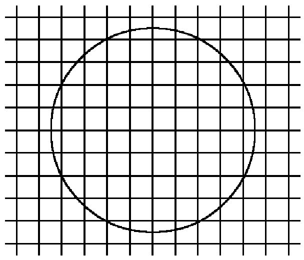

Abbildung 4: Beispiel $n = 2$, $t = 81/4$, $V_2(t) = 63.6\ldots$, $A_2(t) = 69$

Daraus folgt

$$V_n((\sqrt{t} - \sqrt{n})^2) \; = \; \kappa_n (\sqrt{t} - \sqrt{n})^n \; < \; A_n(t) \; < V_n((\sqrt{t} + \sqrt{n})^2) \; < \; \kappa_n (\sqrt{t} + \sqrt{n})^n \; ,$$

durch Anwenden der binomischen Formel also

$$A_n(t) \; = \; \kappa_n \, t^{n/2} \; + \; O(t^{(n-1)/2}) \; . \qquad \Box$$

5.3 Eine einfache untere Abschätzung

Wie klein ist das Restglied wirklich? Um diese Frage präzise stellen zu können, verwenden wir die in der analytischen Zahlentheorie übliche Ω–*Notation*: Für rellwertige Funktionen f, g, h , welche für genügend große t definiert sind, schreibt man

$$f(t) \; = \; g(t) \; + \; \Omega(h(t)) \qquad (\text{für} \quad t \to \infty) \; ,$$

wenn $(f(t) - g(t))/h(t)$ für $t \to \infty$ nicht gegen 0 konvergiert. Ω–Resultate geben also untere Schranken für obere Abschätzungen, also für O–Resultate. Für das Kreis– und Kugelproblem zeigen wir hier

Satz 23 $\quad A_n(t) \;=\; V_n(t) \;+\; \Omega(t^{(n-2)/2})$

Der *Beweis* beruht auf der Beobachtung, daß für alle $m \in \mathbb{N}$

$$A_n(m) \;=\; A_n(m+\frac{1}{2})$$

ist, weil $A_n(t)$ die Anzahl der ganzzahligen Lösungen von $x_1^2 + \ldots + x_n^2 \leq t$ ist. Wäre die Behauptung des Satzes falsch, so hätte man insbesondere

$$\lim_{m\to\infty} \frac{A_n(m) - V_n(m)}{m^{(n-2)/2}} \;=\; 0\,, \qquad \lim_{m\to\infty} \frac{A_n(m+\frac{1}{2}) - V_n(m+\frac{1}{2})}{(m+\frac{1}{2})^{(n-2)/2}} \;=\; 0\,,$$

und daraus durch Differenzbildung

$$0 \;=\; \lim_{m\to\infty} \left(\frac{V_n(m+\frac{1}{2})}{(m+\frac{1}{2})^{(n-2)/2}} - \frac{V_n(m)}{m^{(n-2)/2}} + A_n(m) \left(\frac{1}{m^{(n-2)/2}} - \frac{1}{(m+\frac{1}{2})^{((n-2)/2}} \right) \right) \;\geq$$

$$\geq\; \kappa_n \lim_{m\to\infty} \left(\frac{(m+\frac{1}{2})^{n/2}}{(m+\frac{1}{2})^{(n-2)/2}} - \frac{m^{n/2}}{m^{(n-2)/2}} \right) \;=\; \kappa_n \lim_{m\to\infty} (m+\frac{1}{2}-m) \;=\; \frac{\kappa_n}{2}\,,$$

Widerspruch. □

5.4 Verbesserungen und offene Fragen

Die Beziehungen zu den Darstellungsanzahlen natürlicher Zahlen durch Quadratsummen, die wir eben genutzt haben, machen deutlich, warum man an einem möglichst genauen Restglied interessiert ist. Aus den letzten Sätzen folgt, daß der wahre Exponent α_n des Restglieds zwischen $\frac{n}{2}$ und $\frac{n}{2}-1$ liegen muß. Seit Landau und Hardy (1915) weiß man, daß für $n=2$ die untere Abschätzung $\alpha_2 \geq \frac{1}{4}$ erfüllt, und Verbesserungen hat es seitdem nur in log–Faktoren gegeben. Der Rekord liegt bei Hafner 1981: Es gibt ein $c > 0$ mit

$$A_2(t) \;=\; \pi\, t \;+\; \Omega(t^{1/4}(\log t)^{1/4}(\log\log t)^{(\log 2)/4}\exp(-c(\log\log\log t)^{1/2})\,.$$

Die beste bekannte obere Abschätzung stammt von Huxley 1993 nach Vorarbeiten von Bombieri und Ivaniec, nämlich

$$A_2(t) \;=\; \pi\, t \;+\; O(t^{23/73}(\log t)^{315/146})\,,$$

jedenfalls ist also α_2 schon sehr nahe an $1/4$. Für das Kugelproblem im Raum lauten die besten heute bekannten Resultate

$$A_3(t) \;=\; V_3(t) \;+\; O(t^{\frac{2}{3}}\log^6 t) \;=\; V_3(t) \;+\; \Omega(t^{\frac{1}{2}}\log^{\frac{1}{2}} t)$$

Alle diese Verbesserungen (I.M. VINOGRADOV 1963, SZEGÖ 1926) erfordern einen erheblichen Aufwand! Merkwürdigerweise wird das Problem in höheren Dimensionen wieder einfacher (WALFISZ 1924): Für $n = 4$ ist man mit

$$A_4(t) \;=\; V_4(t) \;+\; O(t\log t) \;=\; V_4(t) \;+\; \Omega(t\log\log t)$$

schon sehr nahe an der unteren Abschätzung des Restglieds durch Satz 23, und für $n > 4$ schließt sich die Lücke ganz vermöge

$$A_n(t) \;=\; V_n(t) \;+\; O(t^{(n-2)/2})\;.$$

6 Der Satz von MINKOWSKI – HLAWKA

Auch dieser Abschnitt beschäftigt sich mit einer Fragestellung, die sich aus dem MINKOWSKIschen Gitterpunktsatz ergibt. Gegeben eine meßbare Menge $S \subset \mathbb{R}^n$, was läßt sich dann über die kleinstmögliche Gitterdeterminante $d(L)$ aussagen, für die ein Gitter $L \subset \mathbb{R}^n$ existiert, für das $L \cap S$ nur aus $\mathbf{0}$ und Randpunkten von S besteht? Wir werden die Fragestellung einschränken auf **Sternkörper** $S \subset \mathbb{R}^n$, die wir durch folgende Eigenschaft definieren: S ist nichtleer, und für jeden Punkt $\mathbf{x}$ der abgeschlossenen Hülle $\overline{S}$ von S und jedes r aus dem halboffenen Intervall $[0,1)$ gehört der Punkt $r\mathbf{x}$ zum offenen Kern S^0 von S . Insbesondere ist auch S^0 nichtleer, und jede von $\mathbf{0}$ ausgehende Halbgerade trifft den Rand von S höchstens in einem Punkt. Beispiele von Sternkörpern sind alle konvexen Körper, die $\mathbf{0}$ im Innern enthalten, aber auch der durch $|x_1x_2| < 1$ definierte Bereich des $\mathbb{R}^2$. Wie schon konvexe Körper heißen auch Sternkörper *symmetrisch*, wenn für jedes $\mathbf{x} \in S$ auch $-\mathbf{x}$ zu S gehört.

6.1 Zulässige Gitter und kritische Determinante

Sei $S \subset \mathbb{R}^n$ meßbar. Ein Gitter $L \subset \mathbb{R}^n$ heißt (S–)**zulässig**, wenn $L \cap S^0 = \mathbf{0}$ ist, und **streng** (S–)**zulässig**, wenn sogar $L \cap S = \mathbf{0}$ ist. Die **Gitterkonstante** oder **kritische Determinante** von S ist

$$\Delta(S) \;:=\; \inf\{\, d(L) \,|\, L \quad \text{streng zulässig}\,\}\;.$$

Wenn kein streng zulässiges Gitter für S existiert, setzen wir $\Delta(S) := \infty$. Anstelle streng zulässiger Gitter kann man bei Sternkörpern ebensogut mit zulässigen Gittern arbeiten wegen der ersten der im folgenden Hilfssatz zusammengestellten elementaren Eigenschaften.

Hilfssatz 24 *$S \subseteq \mathbb{R}^n$ sei ein Sternkörper. Dann gilt*

1. *$\Delta(S) \; = \; \inf\{\, d(L) \mid L$ zulässig $\}$,*
2. *$\Delta(S) \neq 0$,*
3. *jedes Gitter $L \subset \mathbb{R}^n$ mit $d(L) < \Delta(S)$ enthält einen Punkt $\mathbf{0} \neq \mathbf{x} \in L \cap S$,*
4. *für jeden kleineren Sternkörper $K \subset S$ ist $\Delta(K) \leq \Delta(S)$,*
5. *für alle $r \in \mathbb{R}^*$ ist $\Delta(rS) = |r|^n \Delta(S)$,*
6. *für alle $A \in \mathrm{GL}_n(\mathbb{R})$ ist $\Delta(AS) = |\det A| \cdot \Delta(S)$,*
7. *für konvexe symmetrische Körper K ist $\Delta(K) \geq 2^{-n} V(K)$.*

(3. bis 6. sind für beliebige $M \subseteq \mathbb{R}^n$ gültig.) *Beweis.* 1. Wenn L zulässig ist, so ist rL streng zulässig für jedes reelle $r > 1$, denn wir wissen bereits, daß auf jeder von $\mathbf{0}$ ausgehenden Halbgeraden höchstens ein Randpunkt von S liegen kann. Für 6. ersetze man alle Gitter L durch AL , und 5. ist ein Spezialfall davon. 7. ergibt sich aus dem Gitterpunktsatz von Minkowski. 3. und 4. ergeben sich direkt aus der Definition von Δ . Zum Beweis von 2. beachte man, daß $\mathbf{0}$ innerer Punkt von S ist, und daß darum ein konvexer symmetrischer Körper $K \subset S$ existiert, auf den man 7. und 4. anwenden kann. □

Beispiel. $S := \{\, \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid |x_1 x_2 \cdot \ldots \cdot x_n| < 1 \,\}$ ist ein Sternkörper und enthält nach der Ungleichung vom geometrischen und arithmetischen Mittel den konvexen symmetrischen Körper $K = \{\, \mathbf{x} \mid |x_1| + |x_2| + \ldots + |x_n| < n \,\}$ vom Volumen $2^n n^n / n!$, also ist $\Delta(S) \geq n^n/n!$.

Der letzte Punkt des Hilfssatzes gibt eine untere Abschätzung für den Quotienten $\Delta(K)/V(K)$. Läßt sich auch eine *obere* Abschätzung finden, d.h. eine Schranke $\delta/V(K)$, so daß für jedes $d \geq \delta$ ein zulässiges Gitter L mit $d(L) = d$ existiert?

6.2 Hilfsbetrachtungen

Hilfssatz 25 *$f : \mathbb{R}^n \to \mathbb{R}$ sei eine stetige Funktion mit kompaktem Träger $\overline{\{\, \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \neq 0 \,\}}$, und für alle $x \in \mathbb{R}$ sei*

$$V(x) \; := \; \int_{\mathbb{R}^{n-1}} f(x_1, \ldots, x_{n-1}, x)\, d(x_1, \ldots, x_{n-1}) \; .$$

Λ sei ein Gitter in der Hyperebene $x_n = 0$ und $a > 0$ fest gewählt. Dann gibt es ein $\mathbf{z} = (z_1, \ldots, z_{n-1}, a)$, so daß für das Gitter $L_{\mathbf{z}} := \Lambda + \mathbb{Z}\mathbf{z}$ gilt

$$\sum_{\mathbf{x} \in L_{\mathbf{z}},\, x_n \neq 0} f(\mathbf{x}) \; = \; \frac{1}{d(\Lambda)} \sum_{r \in \mathbb{Z},\, r \neq 0} V(ra) \; .$$

Zum *Beweis* dürfen wir vermöge einer Koordinatentransformation in der Hyperebene $x_n = 0$ annehmen, daß Λ das Standardgitter des $\mathbb{R}^{n-1}$ ist; die Integrale ändern sich dabei um den gleichen Faktor wie $d(\Lambda)$. Dann wird

$$\sum_{\mathbf{x}\in L_{\mathbf{z}},\, x_n\neq 0} f(\mathbf{x}) \;=\; \sum_{r\in\mathbb{Z},\, r\neq 0}\;\sum_{u_1,\dots,u_{n-1}\in\mathbb{Z}} f(u_1+rz_1,\dots,u_{n-1}+rz_{n-1},ra)\,.$$

Integriert man die innere Summe über die z_i und setzt $rz_i = y_i$, so wird

$$\int_0^1 \dots \int_0^1 \sum_{u_i\in\mathbb{Z}} f(u_1+rz_1,\dots,u_{n-1}+rz_{n-1},ra)dz_1\cdot\ldots\cdot dz_{n-1} \;=$$

$$= r^{-n+1}\int_0^r \dots \int_0^r \sum_{u_i\in\mathbb{Z}} f(u_1+y_1,\dots,u_{n-1}+y_{n-1},ra)\,dy_1\cdot\ldots\cdot dy_{n-1} \;=$$

$$= \int_0^1 \dots \int_0^1 \sum_{u_i\in\mathbb{Z}} f(u_1+y_1,\dots,u_{n-1}+y_{n-1},ra)\,dy_1\cdot\ldots\cdot dy_{n-1} \;=$$

$$= \int_{\mathbb{R}^{n-1}} f(y_1,\dots,y_{n-1},ra)\,d(y_1,\dots,y_{n-1}) \;=\; V(ra)\,.$$

Summiert man nun über r , so ergibt sich

$$\int_0^1 \dots \int_0^1 \left(\sum_{\mathbf{x}\in L_{\mathbf{z}},\, x_n\neq 0} f(\mathbf{x})\right) dz_1\cdot\ldots\cdot dz_{n-1} \;=\; \sum_{r\in\mathbb{Z},\, r\neq 0} V(ra)\,,$$

und mit dem Mittelwertsatz der Integralrechnung erhält man daraus die Behauptung des Hilfssatzes (für $d(\Lambda) = 1$). $\Box$

Hilfssatz 26 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine beschränkte Riemann–integrierbare Funktion mit kompaktem Träger. Für alle $\epsilon > 0$ gibt es ein Gitter $L \subset \mathbb{R}^n$ mit $d(L) = 1$ und*

$$\sum_{\mathbf{x}\in L,\, \mathbf{x}\neq \mathbf{0}} f(\mathbf{x}) \;<\; \int_{\mathbb{R}^n} f(\mathbf{x})\,d\mathbf{x} \;+\; \epsilon\,.$$

Zum *Beweis* sei zunächst f stetig, $a > 0$ und Λ das Gitter in der Hyperebene $x_n = 0$ mit der Basis $b\mathbf{e}_1,\dots,b\mathbf{e}_{n-1}$, $b = a^{-1/(n-1)}$, so daß $d(\Lambda) = 1/a$ wird und die im vorigen Hilfssatz betrachteten Gitter $L_{\mathbf{z}}$ die Determinante 1 haben. Für $a \to 0$, d.h. $b \to \infty$, ist $f(\mathbf{x}) = 0$ für alle $\mathbf{x} \in \Lambda$, $\mathbf{x} \neq \mathbf{0}$, weil f kompakten Träger hat. Andererseits ist

$$\lim_{a\to 0}\, a \sum_{r\in\mathbb{Z},\, r\neq 0} V(ra) \;=\; \int_{\mathbb{R}^n} f(\mathbf{x})\,d\mathbf{x}\,.$$

Es gibt also zu jedem $\epsilon > 0$ ein hinreichend kleines $a > 0$ und dazu nach Hilfssatz 25 ein Gitter $L = L_{\mathbf{z}}$ mit

$$\sum_{\mathbf{x}\in L,\, \mathbf{x}\neq \mathbf{0}} f(\mathbf{x}) \;=\; \sum_{\mathbf{x}\in L,\, x_n\neq 0} f(\mathbf{x}) \;=\; a \sum_{r\in\mathbb{Z},\, r\neq 0} V(ra) \;<\; \int_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x} \;+\; \epsilon\;,$$

wie behauptet. Verzichtet man nun auf die Voraussetzung, daß f stetig ist, so kann man f als Riemann–integrierbare Funktion von oben durch eine Treppenfunktion und darum (durch Glätten) sogar durch eine stetige Funktion f_1 so approximieren, daß $f_1(\mathbf{x}) \geq f(\mathbf{x})$ ist für alle $\mathbf{x} \in \mathbb{R}^n$ und trotzdem $\int_{\mathbb{R}^n} f_1(\mathbf{x}) d\mathbf{x} < \int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} + \epsilon$, also

$$\sum_{\mathbf{x}\in L,\, \mathbf{x}\neq \mathbf{0}} f(\mathbf{x}) \;\leq\; \sum_{\mathbf{x}\in L,\, \mathbf{x}\neq \mathbf{0}} f_1(\mathbf{x}) \;<\; \int_{\mathbb{R}^n} f_1(\mathbf{x})\, d\mathbf{x} \;+\; \epsilon \;<\; \int_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x} \;+\; 2\epsilon\;. \Box$$

Satz 27 *Sei $M \subset \mathbb{R}^n$ beschränkt und Jordan–meßbar. Dann ist*

$$\Delta(M) \;\leq\; V(M)\;.$$

Wenn darüber hinaus M symmetrisch ist, gilt sogar

$$\Delta(M) \;\leq\; \frac{1}{2} V(M)\;.$$

Sei zunächst $V(M) < 1$ und f die charakteristische Funktion von M. Aus Hilfssatz 26 folgt, daß es ein Gitter $L \subset \mathbb{R}^n$ mit $d(L) = 1$ gibt und

$$\sum_{\mathbf{x}\in L,\, \mathbf{x}\neq \mathbf{0}} f(\mathbf{x}) \;<\; 1 \quad\Rightarrow\quad \sum_{\mathbf{x}\in L,\, \mathbf{x}\neq \mathbf{0}} f(\mathbf{x}) \;=\; 0 \quad\Rightarrow\quad L\cap M \;\subseteq\; \{\mathbf{0}\}\;,$$

L ist also streng M–zulässig, somit $\Delta(M) \leq 1$. Da diese Schlußweise auf alle rM, $r \geq 0$, anwendbar ist, solange nur $V(rM) < 1$ ist, folgt daraus die erste Behauptung. Für die zweite Behauptung sei M symmetrisch und

$$M_{\pm} \;:=\; \{\, \mathbf{x} \in M \,|\, x_n \geq 0 \quad \text{bzw.} \quad x_n \leq 0\,\}\;.$$

Dann ist offenbar $M_- = -M_+$ und $V(M_\pm) = \frac{1}{2}V(M)$, andererseits — wegen der Symmetrie des Gitters — $\Delta(M_\pm) = \Delta(M)$. $\Box$

6.3 Primitive Gitterpunkte

Mit dem letzten Satz hat man eine bereits eine obere Abschätzung von $\Delta(M)/V(M)$ gewonnen. Wie von Minkowski vermutet und von Hlawka bewiesen, kann man diese Abschätzung im Fall von Sternkörpern um einiges verbessern, indem man die Summation über die Gitterpunkte auf eine Summation über eine sehr spezielle Auswahl von Gitterpunkten zurückspielt. Sei dazu $L \subset$

$\mathbb{R}^n$ ein Gitter mit Basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Einen Gitterpunkt $\mathbf{x} \in L$ nennen wir **primitiv**, wenn er $\neq \mathbf{0}$ und kein echtes Vielfaches eines anderen Gitterpunkts ist. An der Koordinatendarstellung

$$\mathbf{x} = a_1\mathbf{a}_1 + \ldots + a_n\mathbf{a}_n$$

ist das daran erkennbar, daß die Koeffizienten $a_1, \ldots, a_n$ teilerfremd sind. Geometrisch läßt sich diese Eigenschaft dadurch ausdrücken, daß die primitiven Gitterpunkte genau jene sind, die vom Nullpunkt aus „sichtbar" sind (schwarze Punkte in Abb. 5, Punkt $(3, 2)$ ist sichtbar vom Nullpunkt aus, Punkt $(6, 4)$ unsichtbar).

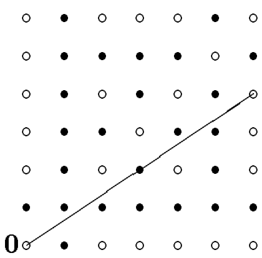

Abbildung 5: Primitive Gitterpunkte im ersten Quadranten eines ebenen Standardgitters

Satz 28 *Sei* $n > 1$ *und* $S \subset \mathbb{R}^n$ *ein Sternkörper. Dann ist*

$$\Delta(S) \leq \frac{1}{\zeta(n)} V(S) .$$

Ist S *ein symmetrischer Sternkörper, so gilt sogar*

$$\Delta(S) \leq \frac{1}{2\zeta(n)} V(S) .$$

Dabei ist $\zeta(n) = \sum_{\mathbb{N}} k^{-n}$ die in [Wo], Abschnitt 1.4.2 eingeführte RIEMANNsche Zetafunktion. Zum *Beweis* sei daran erinnert, daß $1/\zeta(n) = \sum_{\mathbb{N}} \mu(k)k^{-n}$ ist für die in [Wo], Abschnitt 4.1.2 behandelte MÖBIUSsche μ–Funktion (leicht zu beweisen mit Hilfe der EULERschen Produktdarstellung der Zetafunktion, vgl. [Wo], Aufgaben 4.7.1 und 2).
Den Beweis werden wir für *beschränkte* Sternkörper führen. Wie im Beweis von Satz 27 genügt es, zu zeigen, daß $\Delta(S) \leq 1$ ist, wenn $V(S) < \zeta(n)$ gilt bzw. $V(S) < 2\zeta(n)$ im symmetrischen Fall. Sei nun f_0 die charakteristische Funktion von S und

$$f(\mathbf{x}) := \sum_{k=1}^{\infty} \mu(k)\, f_0(k\mathbf{x}) .$$

Für ein zunächst beliebiges Gitter $L \subset \mathbb{R}^n$ bezeichnet man mit $\sum_{L^*}$ die Summation über die primitiven Gitterpunkte von L; die MÖBIUSsche Umkehrformel ([Wo], Abschnitt 4.1.6) ergibt

$$\sum_{\mathbf{x}\in L,\, \mathbf{x}\neq\mathbf{0}} f(\mathbf{x}) = \sum_{m=1}^{\infty} \sum_{L^*} f(m\mathbf{x}) = \sum_{L^*} \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} \mu(k)\, f_0(km\mathbf{x}) =$$

$$= \sum_{L^*} \sum_{s=km=1}^{\infty} f_0(s\mathbf{x}) \sum_{d|s} \mu(d) \; = \; \sum_{L^*} f_0(\mathbf{x}) \; .$$

Aus Hilfssatz 26 folgt nun für jedes $\epsilon > 0$ die Existenz eines Gitters L mit $d(L) = 1$ und

$$\sum_{L^*} f_0(\mathbf{x}) \; = \sum_{\mathbf{x} \in L, \, \mathbf{x} \neq \mathbf{0}} f(\mathbf{x}) \; < \; \int_{\mathbb{R}^n} f(\mathbf{x}) \, d\mathbf{x} \; + \; \epsilon \; =$$

$$= \sum_{k=1}^{\infty} \mu(k) \int_{\mathbb{R}^n} f_0(k\mathbf{x}) \, d\mathbf{x} \; + \; \epsilon \; = \; \sum_{k=1}^{\infty} \frac{\mu(k)}{k^n} V(S) \; + \; \epsilon \; < \; 1$$

(bzw. < 2 für symmetrische S). Daraus folgt, daß S keinen primitiven Gitterpunkt enthält; weil S Sternkörper ist, enthält S überhaupt keinen Gitterpunkt $\neq \mathbf{0}$, somit ist L streng zulässig, damit ist die Behauptung für beschränkte S bewiesen.
Für unbeschränkte S und $V(S) = \infty$ ist nichts zu zeigen. Für unbeschränkte S mit endlichem Volumen $V(S)$ darf man o.B.d.A. annehmen, daß S offen ist und ausgeschöpft wird durch eine Folge beschränkter Sternkörper $S_r := \{ \mathbf{x} \in S | \; ||\mathbf{x}|| < r \}$, für die jeweils ein zulässiges Gitter L_r mit $d(L_r)\zeta(n) \leq V(S_r)$ (bzw. $\leq \frac{1}{2}V(S_r)$) existiert. Da die Länge der beteiligten Gittervektoren nach unten beschränkt ist, kann man mit Hilfe eines geeigneten Konvergenzbegriffs für die Gitter L_r (Chabauty–Topologie, Kompaktheitssatz von Mahler) zeigen, daß auch für S ein zulässiges Gitter L mit der gewünschten oberen Abschätzung für die Gitterdeterminante existiert. Da die Details einigen Aufwand erfordern und dieser Teil des Satzes für uns in der Folge keine Rolle spielen wird, sei hierzu auf speziellere Literatur über die Geometrie der Zahlen verwiesen ([Ca], [GL]). □

Für $n = 2$ und konvexe symmetrische Körper hat man damit die Abschätzungen

$$3.2 \; < \; \frac{\pi^2}{3} \; \leq \; \frac{V(K)}{\Delta(K)} \leq 4$$

und entsprechend für $n = 3$

$$2.4 \; < \; 2\zeta(3) \; \leq \; \frac{V(K)}{\Delta(K)} \leq 8 \; .$$

Hier deutet sich schon an, daß mit wachsender Dimension die vorhandenen Abschätzungen immer weniger aussagekräftig werden. Beweistechnisch liegt das daran, daß in Hilfssatz 25 ein Mittelungsprozeß über alle Gitter eines bestimmten Typs vorgenommen wird und somit ein durchschnittliches und keineswegs ein besonders günstiges zulässiges Gitter ausgewählt wird. Wie man dieses für Kreise und Kugeln gewinnen kann, werden wir im folgenden Abschnitt sehen — leider aber mit Methoden, die auf kleine Dimensionen beschränkt sind. Zu erwähnen ist hier, daß Rogers und W.M. Schmidt Verbesserungen des Satzes von Minkowski–Hlawka erzielt haben, welche einen erheblichen Aufwand erfordern und für große n leider nur eine geringe Verbesserung erbringen ([Ca], [Ro]).

7 Packungsdichte

7.1 Gitterpackungen und zulässige Gitter

Sei $K \subset \mathbb{R}^n$ ein offener beschränkter konvexer Körper. Ein Gitter $L \subset \mathbb{R}^n$ heißt **Packungsgitter** für K, wenn die Menge der Translate $K+L$ disjunkt ist, d.h. wenn für alle $\mathbf{x} \neq \mathbf{y} \in L$ gilt

$$K + \mathbf{x} \cap K + \mathbf{y} = \emptyset .$$

In diesem Fall nennt man $K+L$ eine **Gitterpackung**. Die Voraussetzungen, daß L ein Gitter ist und daß K konvex und beschränkt ist, spielt für die Begriffsbildung eigentlich keine Rolle, wird aber Sätze und Beweise dieses Abschnitts ganz erheblich vereinfachen. Auf den allgemeineren Packungsbegriff werden wir nur gelegentlich verweisen. Gitterpackungen stehen in engem Zusammenhang mit den im vorigen Abschnitt behandelten zulässigen Gittern:

Satz 29 *$K \subset \mathbb{R}^n$ sei offener beschränkter konvexer Körper. Ein Gitter $L \subset \mathbb{R}^n$ ist genau dann ein Packungsgitter für K, wenn L zulässig ist für die* Differenzmenge

$$DK := \{\mathbf{x} - \mathbf{y} \,|\, \mathbf{x}, \mathbf{y} \in K\} .$$

Wenn K symmetrischer konvexer Körper ist, so ist $DK = 2K$.

(Man erinnere sich, daß Differenzmengen schon in den Beweisen des Minkowskischen Gitterpunktsatzes eine wesentliche Rolle gespielt haben.) Zum *Beweis* sei zunächst L kein Packungsgitter für K. Dann gibt es $\mathbf{x} \neq \mathbf{y} \in L$ mit

$$K + \mathbf{x} \cap K + \mathbf{y} \neq \emptyset ,$$

also $\mathbf{a}, \mathbf{b} \in K$ mit

$$\mathbf{0} \neq \mathbf{x} - \mathbf{y} = \mathbf{a} - \mathbf{b} \in DK \cap L ,$$

L ist also nicht DK–zulässig. Der Umkehrschluß funktioniert genauso, und der Zusatz über symmetrische konvexe Körper ist evident. $\square$

7.2 Dichte

Für eine Packung von K kann man nun in naheliegender Weise einen Dichtebegriff einführen: Sei etwa Q ein achsenparalleler und $\mathbf{0}$–symmetrischer Quader der Seitenlänge s und

$$\delta(L, K) := \limsup_{s \to \infty} V((K+L) \cap Q)/s^n .$$

Daß für Gitterpackungen beschränkter offener Mengen K rechts sogar der Limes existiert, ist eine Übungsaufgabe der Analysis II oder III: Man mache sich klar, daß man Q durch Parallelotope ersetzen darf und arbeite mit Vielfachen eines Fundamentalparallelotops von L. Dann steht rechts gerade der Anteil von $K+L$ am Fundamentalparallelotop von L, und die Translationsinvarianz des Volumens sowie die Pflasterungseigenschaft des Fundamentalparallelotops (vgl. Abschnitt 1) ergeben

Hilfssatz 30 *Für offene beschränkte $K \subset \mathbb{R}^n$ mit Packungsgitter L ist die Gitterpackungsdichte*

$$\delta(L, K) \;=\; V(K)/d(L)\,.$$

Es ist plausibel, daß man sich bei festem K für besonders dichte Packungen interessiert und darum versucht, die **obere (Gitter)packungsdichte**

$$\rho(K) \;:=\; \sup_L \delta(L, K)$$

zu bestimmen oder zumindest einzugrenzen. Das Supremum werden wir hier über alle Packungsgitter L für K nehmen, man beachte allerdings die Bemerkungen am Ende dieses Abschnitts 7 über Nicht–Gitter–Packungen. Bei festem konvexem symmetrischem K das Supremum über die $V(K)/d(L)$ zu nehmen, ist äquivalent zur Bestimmung des Infimums über die $d(L)$ für $2K$–zulässige Gitter L (Satz 29), also zur Bestimmung von $\Delta(2K) = 2^n\Delta(K)$ (Hilfssatz 24.5), und nach dem Satz von Minkowski–Hlawka heißt das

Satz 31 *Sei $K \subset \mathbb{R}^n$, $n > 1$, ein konvexer symmetrischer Körper. Für seine obere Gitterpackungsdichte gilt*

$$2^{1-n}\,\zeta(n) \;\le\; \rho(K)\,.$$

7.3 Dichteste Kreisgitterpackungen

Am interessantesten — auch für praktische Anwendungen, wie sich im kommenden Abschnitt zeigen wird — sind dichte Kugelpackungen. Hierzu sind keine allgemeinen Resultate bekannt, die wesentlich besser wären als die eben gewonnene Abschätzung für $\rho(K)$, allerdings sind für viele kleine n dichte und z.T. sogar dichteste Kugelgitterpackungen bekannt [CS], die eine weit bessere Packungsdichte ergeben als Satz 31. Wir wollen uns hier mit zwei klassischen Resultaten begnügen und beginnen mit einem Satz von Lagrange (1773).

Satz 32 *Das hexagonale Gitter $L \subset \mathbb{R}^2$ (d.h. mit zwei Gitterbasisvektoren gleicher Länge, die den Winkel $\pi/3$ einschließen) liefert die — bis auf Streckungen und Drehungen sogar eindeutig bestimmte — dichteste Kreisgitterpackung. Die Dichte ist*

$$\rho(K) \;=\; \delta(L, K) \;=\; \frac{\pi}{2\sqrt{3}} \;\approx\; 0.907\,.$$

(Die Schranke aus Satz 31 würde hier $\zeta(2)/2 = \pi^2/12 \approx 0.822$ ergeben.)
Beweis. 1. O.B.d.A. dürfen wir annehmen, daß K der offene Einheitskreis ist. Da Packungsgitter $L \subset \mathbb{R}^2$ gesucht sind, haben alle Gitterpunkte einen euklidischen Abstand ≥ 2 voneinander, insbesondere erfüllt der Gitterpunkt mit kleinster euklidischer Norm die Ungleichung $||\mathbf{a}|| \geq 2$.
2. Wäre nun $||\mathbf{a}|| > 2$, so hätten keine zwei Kreise der Gitterpackung $L+K$ gemeinsame Randpunkte, und es gäbe ein $\epsilon > 0$, so daß auch noch $(1-\epsilon)L$ ein Packungsgitter für K wäre, und zwar mit größerer Dichte als L. Bei der Suche nach dichtesten Gitterpackungen dürfen wir also annehmen, daß der kürzeste Gittervektor $\mathbf{a} \neq \mathbf{0}$ die euklidische Länge 2 hat. Durch Drehung des Koordinatensystems darf man o.B.d.A. annehmen, daß $\mathbf{a} = (2,0)$ ist, also in Richtung der positiven x–Achse zeigt.
3. Würde der Kreis K nun nur die Nachbarkreise $K \pm \mathbf{a}$ berühren (d.h. gemeinsame Randpunkte mit ihnen besitzen), so würde die Gitterpackung in isolierte Kreisketten $K + \mathbf{b} + \mathbb{Z}\mathbf{a}$ zerfallen und man könnte L in y–Richtung mit einem Faktor $(1-\epsilon)$ multiplizieren, ohne die Gitterpackungseigenschaft zu zerstören, aber mit Verdichtung der Packung. Man darf also annehmen, daß K nicht nur seine beiden Nachbarn in x–Richtung berührt, daß also ein Gitterpunkt $\mathbf{b} \in L - \mathbb{Z}\mathbf{a}$ existiert mit euklidischer Länge $||\mathbf{b}|| = 2$.
4. Durch eventuellen Übergang zu $-\mathbf{b}$ darf man annehmen, daß $\mathbf{a}$ und $\mathbf{b}$ einen Winkel $\leq \pi/2$ einschließen. Dieser Winkel ist außerdem $\geq \pi/3$, da andernfalls $||\mathbf{b}-\mathbf{a}|| < 2$ wäre. Die Punkte $\mathbf{a}$ und $\mathbf{b}$ bilden eine Gitterbasis von L: Andernfalls gäbe es in dem von $\mathbf{a}$ und $\mathbf{b}$ aufgespannten Parallelogramm einen weiteren Gitterpunkt, der von einem der Eckpunkte notwendig einen euklidischen Abstand < 2 hätte im Widerspruch zu der Packungseigenschaft.

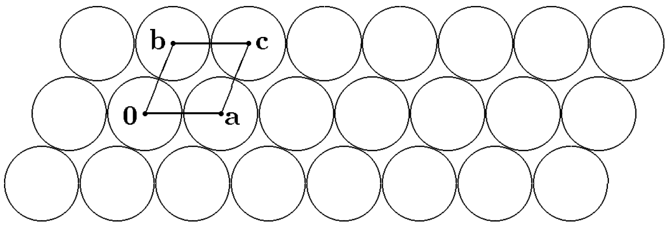

Abbildung 6: Eine Kreisgitterpackung im $\mathbb{R}^2$

5. Das von $\mathbf{a}$ und $\mathbf{b}$ aufgespannte Parallelogramm ist also ein Fundamentalparallelogramm P von L, und es ist zu untersuchen, wann unter den oben gefundenen Bedingungen

$$||\mathbf{a}|| = ||\mathbf{b}|| = 2\,, \qquad \frac{\pi}{3} \leq \angle(\mathbf{a},\mathbf{b}) \leq \frac{\pi}{2}$$

$d(L) = V(P)$ sein Minimum annimmt. Es ist leicht zu sehen, daß dies gerade für den Winkel $\pi/3$ der Fall ist. $\square$

7.4 Die dichteste Kugelgitterpackung

Satz 33 *(*Gauss *1831) Die (bis auf Streckungen und Drehungen) eindeutige dichteste Kugelgitterpackung im* $\mathbb{R}^3$ *wird durch das* flächenzentriert–kubische Gitter $L \subset \mathbb{R}^3$ *realisiert, dessen Basispunkte mit* $\mathbf{0}$ *ein reguläres Tetraeder bilden.*

$$\rho(K) \;=\; \frac{\pi}{3\sqrt{2}} \;\approx\; 0.74$$

(Die Schranke aus Satz 31 würde hier $\zeta(3)/4 \approx 0.3$ ergeben!) Die *Beweis*schritte 1. bis 3. aus dem Beweis von Satz 32 können wir direkt übernehmen. Sie zeigen, daß ein dichtestes Packungsgitter L für die Einheitskugel in der xy–Ebene eines geeigneten cartesischen Koordinatensystems ein Teilgitter Λ bilden mit Erzeugenden $\mathbf{a}, \mathbf{b}$, welche die Bedingungen

$$||\mathbf{a}|| \;=\; ||\mathbf{b}|| \;=\; 2\,, \qquad \frac{\pi}{3} \;\le\; \angle(\mathbf{a}, \mathbf{b}) \;\le\; \frac{\pi}{2}$$

erfüllen. Der Schnitt der Kugelpackung mit der xy–Ebene sieht also aus wie in Abb. 6 angegeben.

4. Wie schon in der Basiskonstruktion aus Abschnitt 1 zerfällt L in Schichten parallel zu Λ. Wenn $\Lambda + K$ keine andere Kugel der Kugelpackung $L + K$ berührt, kann man die z–Koordinaten des Gitters L wieder mit $(1-\epsilon)$ multiplizieren, um die Dichte zu erhöhen und ohne die Packungseigenschaft zu zerstören. Bei der Suche nach dichtesten Kugelgitterpackungen darf man also zusätzlich annehmen, daß ein dritter Gittervektor $\mathbf{d} \in L - \Lambda$ existiert mit euklidischer Länge $||\mathbf{d}|| = 2$. Die Projektion $\mathbf{d}'$ von $\mathbf{d}$ auf die xy–Ebene liegt o.B.d.A. in dem Fundamentalparallelogramm von Λ, welches von $\mathbf{a}, \mathbf{b}$ erzeugt wird, o.B.d.A. sogar in dem abgeschlossenen Dreieck $\mathbf{0}, \mathbf{a}, \mathbf{b}$. Analog zum Beweis von Satz 32 bilden $\mathbf{a}, \mathbf{b}, \mathbf{d}$ eine Gitterbasis von L, weil das von ihnen erzeugte Parallelotop P nur Punkte enthält, die von einem der Eckpunkte einen Abstand < 2 haben. Die Abbildung 7 zeigt das Teilgitter Λ (schwarze Punkte) und die Translate $\Lambda + \mathbf{d}'$ (weiße Punkte), also die Projektion der nächsten Schicht von L.

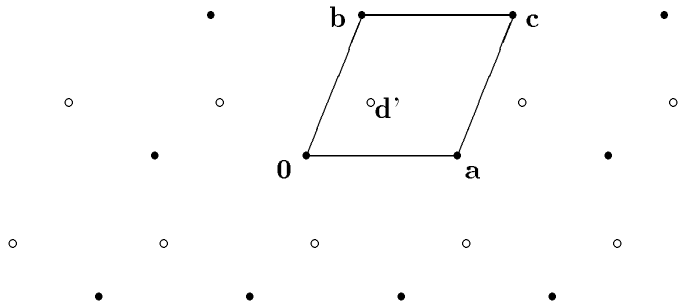

Abbildung 7: Zwei Schichten einer Kugelgitterpackung im $\mathbb{R}^3$, Grundriß

5. Die Gitterdeterminante $d(L) = V(P)$ ist nun $2Fh$, wobei F die Fläche des Dreiecks $\mathbf{0}, \mathbf{a}, \mathbf{b}$ bezeichnet und h den Absolutbetrag der z–Koordinate von $\mathbf{d}$. Wegen

$$4 \;=\; ||\mathbf{d}||^2 \;=\; ||\mathbf{d}'||^2 + h^2$$

läuft die Suche nach einer dichtesten Kugelgitterpackung bei gegebenem Λ darauf hinaus, $||\mathbf{d}'||$ zu maximieren. Da die Kugel $K+\mathbf{d}$ sich nicht mit den Kugeln $K+\mathbf{a}$, $K+\mathbf{b}$ schneiden darf, muß dabei

$$||\mathbf{a}-\mathbf{d}'||^2\ ,\ ||\mathbf{b}-\mathbf{d}'||^2\ \geq\ 4\ -\ h^2\ =\ ||\mathbf{d}'||^2$$

sein. Etwas Elementargeometrie zeigt: Unter diesen Bedingungen wird $||\mathbf{d}'||$ genau dann maximal, wenn hier das Gleichheitszeichen steht, wenn also $\mathbf{d}'$ der Umkreismittelpunkt des Dreiecks $\mathbf{0},\mathbf{a},\mathbf{b}$ ist und $R=||\mathbf{d}'||$ der Umkreisradius, denn jeder andere Punkt $\mathbf{d}''$ im Dreieck hat von einem der Eckpunkte einen kleineren Abstand als R.

6. Um ein Kugelpackungsgitter maximaler Dichte zu finden, ist also $2Fh$ zu minimieren, wo F die Fläche eines gleichschenkligen Dreiecks mit Schenkellänge 2 und Basislänge u ist, $2\leq u\leq 2\sqrt{2}$, und h zum Umkreisradius R des Dreiecks in der Beziehung $h^2=4-R^2$ steht. Etwas mehr Elementargeometrie (Übungsaufgabe zur Trigonometrie) zeigt

$$\frac{u^2}{16}+\frac{1}{R^2}\ =\ 1\quad\text{und}\quad F\ =\ \frac{u}{R}\,.$$

Es ist zweckmäßig, $v:=1/R^2$ als Hilfsgröße einzuführen ($\frac{1}{2}\leq v\leq\frac{3}{4}$) und zum Quadrat der Gitterdeterminante überzugehen. Dieses wird dann

$$4F^2h^2\ =\ 4u^2v(4-\frac{1}{v})\ =\ 64(1-v)(4v-1)$$

und nimmt sein Minimum (32) genau an den beiden Randpunkten des Intervalls an, also für das gleichseitige und das rechtwinklige Dreieck. Mit $V(K)=\frac{4}{3}\pi$ ergibt sich in beiden Fällen die im Satz angegebene obere Gitterpackungsdichte.

7. Um die Eindeutigkeitsaussage zu zeigen, ist noch nachzuweisen, daß beide Fälle das gleiche Gitter in verschiedenen Koordinatensystemen beschreiben. Im Fall des hexagonalen Untergitters Λ wähle man die Gitterbasis

$$(2,0,0)\ ,\ (1,\sqrt{3},0)\ ,\ (1,\frac{1}{\sqrt{3}},\sqrt{\frac{8}{3}})$$

und im Fall des quadratischen Gitters Λ die Gitterbasis

$$(2,0,0)\ ,\ (1,1,\sqrt{2})\ ,\ (1,-1,\sqrt{2})$$

und überzeuge sich davon, daß jedesmal mit $\mathbf{0}$ die Ecken eines regulären Tetraeders entstehen. □

7.5 Schlußbemerkungen

Im Fall ebener Kreispackungen kann man zeigen, daß keine Nicht–Gitter–Packung existiert, welche die Dichte der hexagonalen Gitterpackung aus Satz 32 erreicht. Dieser sehr anschauliche und lange vermutete Sachverhalt ist durchaus nicht leicht zu beweisen. Ein erster Beweis wurde von THUE 1910 gegeben, allerdings nicht in allen Details überzeugend. Den ersten allgemein akzeptierten Beweis hat L. FEJES TÓTH 1940 geliefert, vgl. [Ro].

Im Fall von Kugelpackungen, also für $n = 3$, liegen die Dinge schon viel komplizierter, was man sich bereits anhand der dichtesten Kugelgitterpackung aus Satz 33 klarmachen kann. Wir dürfen davon ausgehen, daß der Schnitt von L mit der xy–Ebene das hexagonale Gitter Λ ist (Schicht a, schwarze Punkte in Abbildung 8). Die in der z–Richtung nächstgelegene Schicht b von Gitterpunkten ist in Abb. 8 ebenfalls auf die xy–Ebene projiziert und als weiße Punkte dargestellt, ebenso die darauffolgende Schicht c (als Kreuze auf die xy–Ebene projiziert). Für die Kugelgitterpackung $L+K$ wiederholen sich beim Durchlaufen der z–Achse die Projektionen der Schichten periodisch als abcabcabc u.s.w.

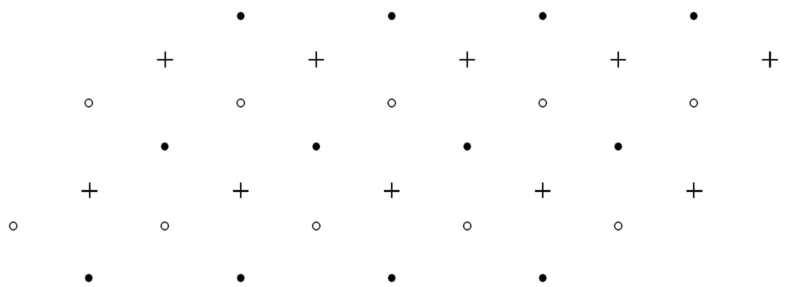

Abbildung 8: Drei Schichten des dichtesten Kugelpackungsgitters im $\mathbb{R}^3$, Grundriß

Man könnte aber ebensogut die Schichten der Kugelpackung anders aufeinandersetzen, etwa als ababacbcaba u.s.w., solange nur keine zwei Schichten gleichen Typs aufeinander folgen. Dabei entsteht eine Kugelpackung, die jedenfalls in z–Richtung nicht einmal periodisch sein muß, wohl aber die gleiche maximale Packungsdichte $\pi/3\sqrt{2}$ hat. Sowohl die Kugel–Gitterpackung wie die Nicht–Gitter–Kugelpackungen dieser Dichte kommen in der Natur vor, z.B. in der Anordnung der Atome in metallischem Gold und diversen Kristallgittern. In der Natur hat man es natürlich häufig mit Packungen aus endlich vielen Kugeln zu tun (Tennisbälle, Orangen). Verallgemeinert man den Dichtebegriff auf solche endlichen Packungen, so stellt man fest, daß für wenige Kugeln zunächst lineare Anordnungen dichter sind als Teilbereiche der hier diskutierten Gitter– oder Nicht–Gitter–Kugelpackungen. Mit wieviel Kugeln die Dichte dieser linearen Packungen überboten wird durch sogenannte *Clusterpackungen* oder schließlich Teilbereiche von Gitterpackungen, ist durchaus nicht klar und führt auf interessante geometrisch–kombinatorische Probleme, vgl. Arbeiten von WILLS und anderen ([Le]).

Könnte es eine unendliche Nicht–Gitter–Kugelpackung geben, welche die Dichte $\pi/3\sqrt{2}$ sogar überbietet? Schon KEPLER hat vermutet, daß dies unmöglich ist, und im Lauf der Mathematikgeschichte sind einige erfolglose Beweisversuche für diese Vermutung unternommen worden. Erst in jüngster Zeit ist ein offenbar überzeugender Beweis von TH. HALES gegeben worden, vgl. [Oe]. Ähnlich wie der Beweis des Vierfarbensatzes durch APPEL und HAKEN hat er leider den Nachteil, daß er die Behauptung auf so viele Fallunterscheidungen reduziert, daß diese nur mit Computerhilfe abzuarbeiten sind.

Die entsprechende Fragestellung in höheren Dimensionen ist völlig offen. Für einige Dimensionen kennt man immerhin noch die dichtesten Kugelgitterpackungen [CS], z.B. ist für $n = 4$

$$L \ := \ \{ (m_1, m_2, m_3, m_4) \in \mathbb{Z}^4 \,|\, m_1 + m_2 + m_3 + m_4 \equiv 0 \bmod 2 \}$$

das dichteste Packungsgitter (für Kugeln vom Radius $\sqrt{2}/2$, Dichte $\pi^2/16$, Korkine und Zolotareff 1872). Die Beweise der Sätze 32 und 33 auf höhere Dimensionen zu verallgemeinern, ist nicht nur technisch sehr schwierig, wie z.B. der Punkt 4 dieser Beweise zeigt: Man erweitere das Standardgitter $L_0 = \mathbb{Z}^5 \subset \mathbb{R}^5$ um den Mittelpunkt des Fundamentalquaders zu einem Obergitter L und benutze dieses als Packungsgitter für Kugeln K vom Radius $1/2$. Der Mittelpunkt des Fundamentalquaders hat dann von allen Ecken einen Abstand > 1; folglich ist es nicht möglich, aus den Mittelpunkten der K berührenden Kugeln eine Gitterbasis von L zu gewinnen, man erhält nur eine Basis des Untergitters L_0.

Hinsichtlich der Packungen endlich vieler Kugeln gibt es ebenfalls eine Überraschung [Le]: Mit wachsendem $n > 4$ werden selbst die dichtesten Kugelgitterpackungen so dünn, daß für beliebig große Anzahlen von Kugeln lineare Anordnungen dichter sind als Teilbereiche von Gitterpackungen!

8 Packungsdichte und Codierungstheorie

8.1 Lineare Codes in Kurzfassung

Ein **Code** der **Länge** n ist eine Untermenge C eines Vektorraums $V = \mathbb{F}_q^n$ über einem endlichen Körper $\mathbb{F}_q$ mit q Elementen, die Elemente von C heißen dann *Nachrichten* oder **Codewörter**. Wenn C sogar Untervektorraum von V ist, nennt man C einen **linearen Code**, und im Fall $q = 2$ heißt C **binär**. Wir werden hier zumeist binäre lineare Codes betrachten. Die **Hammingdistanz** von $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{w} = (w_1, \ldots, w_n) \in V$ ist als die Anzahl der Koordinaten definiert, in denen sich $\mathbf{v}$ und $\mathbf{w}$ unterscheiden, also

$$d(\mathbf{v}, \mathbf{w}) \ := \ \#\{ i = 1, \ldots, n \,|\, v_i \neq w_i \} \ .$$

Hilfssatz 34 *Die Hammingdistanz definiert eine Metrik auf* V. $\square$

Dann ist klar, was man unter einer *Hammingkugel*

$$S_r(\mathbf{v}) \ := \ \{ \mathbf{w} \in V \,|\, d(\mathbf{v}, \mathbf{w}) \leq r \}$$

zu verstehen hat. Das **Minimalgewicht** $w(C)$ eines Codes C ist die kleinste Hammingdistanz zwischen zwei verschiedenen Punkten von C. Wenn $w(C) > 1$ ist, kann bei der Übermittlung von Codewörtern zumindest das Vorliegen eines einzelnen Übertragungsfehlers, d.h. ein Fehler in *einer* Koordinate erkannt werden, etwa durch Vergleich mit der Liste aller Codewörter (was ein sehr aufwendiges Verfahren ist; bei linearen Codes gibt es dazu viel schnellere Möglichkeiten, s.u.). Ein Code C heißt t–**fehlerkorrigierend**, wenn in jeder Hammingkugel $S_t(\mathbf{v})$ höchstens ein Punkt von C liegt. C heißt **perfekter** t–fehlerkorrigierender Code, wenn in jeder Hammingkugel $S_t(\mathbf{v})$ *genau* ein Codewort liegt. Die Bezeichnung kommt natürlich daher, daß dann (für vorgegebenes t) eine möglichst große Anzahl von Codewörtern mit möglichst kleiner Länge n existieren. Es ist klar, daß man einerseits ein großes t haben möchte — abhängig davon, wie viele Übertragungsfehler erfahrungsgemäß auftreten und wie gravierend diese sind — und andererseits aber eine große *Informationsrate*

$$R \;:=\; \frac{\log_q \#C}{\log_q \#\mathbb{F}_q^n}\;.$$

Für lineare Codes C der Dimension k über $\mathbb{F}_q$ ist die Informationsrate einfach $R = k/n$. Wir stellen einige einfache Fakten zusammen:

Hilfssatz 35 *Sei C ein Code in $V = \mathbb{F}_q^n$ und $t \in \mathbb{N}$.*

1. *Wenn die Hammingkugeln $S_t(\mathbf{v})$ um die Punkte $\mathbf{v} \in C$ disjunkt sind, ist C ein t–fehlerkorrigierender Code.*
2. *C ist t–fehlerkorrigierend genau dann, wenn sein Minimalgewicht $w(C) \geq 2t+1$ erfüllt.*
3. *Wenn C linearer Code ist, ist sein Minimalgewicht*
$$w(C) \;=\; \min\{\, d(\mathbf{0},\mathbf{v}) \mid \mathbf{v} \in C\,,\ \mathbf{v} \neq \mathbf{0}\,\}\;.$$
4. *C ist genau dann ein perfekter t–fehlerkorrigierender Code, wenn V die disjunkte Vereinigung der Hammingkugeln $S_t(\mathbf{v}),\ \mathbf{v} \in C$, ist.*
5. *Ein 1–fehlerkorrigierender binärer linearer Code erfüllt*
$$\#C \;\leq\; \frac{2^n}{n+1}\;.$$
Die Gleichheit gilt genau dann, wenn er perfekt ist. □

Um die entscheidenden Daten eines Codes $C \subset \mathbb{F}_q^n$ mit $M = \#C$ Codewörtern und Minimalgewicht $w = w(C)$ zusammenzufassen, spricht man von einem (n, M, w)–Code, bei binären linearen Codes, also im Fall $q = 2, \#C = 2^k, k = \dim C$, von einem $[n, k, w]$–Code. Es gibt zwei Möglichkeiten, lineare Codes einfacher zu beschreiben als durch Angabe einer Liste von $M = q^k$ Codewörtern: Erstens lassen sich k Basisvektoren von C als Zeilen einer $k \times n$–**Generatormatrix** G auffassen, zweitens kann man C als Kern einer surjektiven linearen Abbildung $\mathbb{F}_q^n \to \mathbb{F}_q^{n-k}$ beschreiben, die durch eine $(n-k) \times n$–**Kontrollmatrix** H gegeben ist. Bezeichnet man die transponierte Matrix mit H^t, so ist eine Kontrollmatrix dadurch charakterisiert, daß sie den Rang $n-k$ hat und daß GH^t die $(n-k)\times(n-k)$–Nullmatrix ist. Bei binären linearen Codes hat die Kontrollmatrix noch eine andere Interpretation. Zu einem Code C definiert man einen **dualen Code**

$$C^{\perp} := \{ (w_1,\ldots,w_n) \mid w_1z_1+\ldots+w_nz_n = 0 \quad \text{für alle} \quad (z_1,\ldots,z_n)\in C \} ,$$

also eigentlich die Menge jener Linearformen des Dualraums von V, die auf C verschwinden.

Hilfssatz 36 *Sei C ein linearer binärer $[n,k,w]$–Code. Dann gilt*

1. $\dim C^{\perp} = n-k$
2. *Jede Kontrollmatrix H von C ist Generatormatrix von $C^{\perp}$ und umgekehrt.*
3. $C^{\perp\perp} = C$
4. *Wenn $2k = n$ ist und $C\subseteq C^{\perp}$ oder $C^{\perp}\subseteq C$, so gilt sogar $C = C^{\perp}$, der Code heißt dann* selbstdual.
5. *Ein Code C ist selbstdual genau dann, wenn jede Generatormatrix auch Kontrollmatrix und umgekehrt.*
6. *Ein selbstdualer Code ist insbesondere* von geradem Gewicht, *d.h. für jedes Codewort $\mathbf{v}=(v_1,\ldots,v_n)$ ist $d(\mathbf{0},\mathbf{v})$ gerade, und es erfüllt die* Paritätsbedingung

$$v_1+\ldots+v_n = 0 . \quad \square$$

8.2 Beispiele

1. Der triviale $[n,n,1]$–Code $C=\mathbb{F}_2^n=V$, bei dem weder Fehlererkennung noch Fehlerkorrektur möglich ist. Der duale Code besteht hier nur aus dem Wort $\mathbf{0}$.

2. Der $[3,1,3]$–Wiederholungscode mit Generatormatrix $(1,1,1)$ ist zwar 1–fehlerkorrigierend und perfekt, hat aber die äußerst schlechte Informationsrate $R=1/3$.

3. Die **Hammingcodes** H_r, benannt nach R.W. Hamming, einem der Gründungsväter der Codierungstheorie, kann man wie folgt konstruieren. Sei $r\in\mathbb{N}, r>1$ und $n:=2^r-1$ die Anzahl der Vektoren $\neq\mathbf{0}$ in $\mathbb{F}_2^r$; diese Vektoren denkt man sich als Spalten einer $r\times n$–Matrix H, z.B. im Fall $r=3,\ n=7$

$$\begin{pmatrix} 1&0&0&1&1&0&1\\ 0&1&0&1&0&1&1\\ 0&0&1&0&1&1&1 \end{pmatrix} .$$

Schon anhand der Einheits–Spaltenvektoren sieht man, daß H den Rang r besitzt und somit als Kontrollmatrix eines Codes H_r der Dimension $k = n - r = 2^r - 1 - r$ dienen kann. Hat nun $\mathbf{v} \in C \subseteq V = \mathbb{F}_2^n$ unter seinen Koordinaten $d = w(C)$ Einsen, so müssen wegen $\mathbf{v}H^t = \mathbf{0}$ jedenfalls d Spalten von H linear abhängig sein, und das kann weder für $d = 1$ vorkommen (die Nullspalte kommt nicht vor) noch für 2, weil im $\mathbb{F}_2^n$ zwei linear abhängige Vektoren $= \mathbf{0}$ sind oder übereinstimmen; es tritt erst für $d = 3$, ein. Mit Hilfe von Hilfssatz 35, Punkte 2,3, und 5, sieht man darum:

Satz 37 *Für $r > 1$ sind die Hammingcodes H_r perfekte 1–fehlerkorrigierende $[2^r - 1, 2^r - 1 - r, 3]$–Codes.*

(Für $r = 2$ entsteht nur wieder der Wiederholungscode aus Beispiel 2, aber für $r > 2$ erhält man brauchbare Informationsraten.)

4. Aus jedem binären linearen Code C läßt sich ein **erweiterter Code** C^* als Bild von C unter der Einbettung

$$\mathbb{F}_2^n \hookrightarrow \mathbb{F}_2^{n+1} \ : \ (x_1, \ldots, x_n) \ \mapsto \ (x_1, \ldots, x_n, x_1 + \ldots + x_n)$$

definieren. Offensichtlich ist C^* ein $[n+1, k, w(C^*)]$–Code von geradem Gewicht, da man durch die letzte Koordinate die Gültigkeit der Paritätsbedingung erzwungen hat. Aus einer Kontrollmatrix von C eine Kontrollmatrix von C^* zu erzeugen, ist leicht: Der letzte Zeilenvektor besteht nur aus Einsen (Paritätsbedingung), die übrigen Zeilen werden durch 0 in der letzten Koordinate ergänzt. Für den $[3,3,1]$–Code aus Beispiel 1 ist

$$(1\,1\,1\,1)$$

die Kontrollmatrix von C^*, und die Kontrollmatrix des erweiterten Hammingcodes H_3^* ist

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} .$$

H_3^* ist aus verschiedenen Gründen besonders interessant. Hilfssatz 36.5 zeigt, daß er selbstdual ist, dazu vom Typ $[8,4,4]$ und *doppelt–gerade*, d.h. alle Hammingabstände zum Nullpunkt sind sogar durch 4 teilbar.

8.3 Codes und Gitter

Sei $C \subseteq \mathbb{F}_2^n$ ein binärer linearer Code der Dimension k. Der komponentenweise verwendete kanonische Homomorphismus

$$\rho \ : \ \mathbb{Z}^n \ \to \ (\mathbb{Z}/2\mathbb{Z})^n \ = \ \mathbb{F}_2^n$$

zeigt, daß man dem Code ein Gitter $\rho^{-1}(C)$ zwischen $\mathbb{Z}^n$ und $2\mathbb{Z}^n$ zuordnen kann. Die Homomorphiesätze der Gruppentheorie und Folgerung 11 zeigen

$$[\mathbb{Z}^n : \rho^{-1}(C)] = 2^{n-k}, \quad [\rho^{-1}(C) : 2\mathbb{Z}^n] = 2^k, \quad d(\rho^{-1}(C)) = 2^{n-k}.$$

Aus Gründen, die gleich einsichtig werden, versieht man das Gitter üblicherweise mit einem Normierungsfaktor $1/\sqrt{2}$:

$$\Gamma_C := \frac{1}{\sqrt{2}}\rho^{-1}(C), \quad d(\Gamma_C) = 2^{\frac{n}{2}-k}$$

Schreibt man das Standardskalarprodukt $\mathbf{x}\mathbf{y}^t$ der (Zeilen)vektoren des $\mathbb{R}^n$ einfach als $\mathbf{x}\cdot\mathbf{y}$ (identifiziert also $\mathbb{R}^n$ mit seinem Dualraum), so kann man jedem Gitter $L \subset \mathbb{R}^n$ ein **duales Gitter** durch

$$L^* := \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x}\cdot\mathbf{y} \in \mathbb{Z} \quad \text{für alle} \quad \mathbf{y} \in L\}$$

zuordnen (nicht verwechseln mit den primitiven Gitterpunkten aus Abschnitt 6 !). Als Basis von L^* kann man stets eine zur Basis A von L duale Basis nehmen, also die Transponierte der Matrix–Inversen A^{-1}. Wenn $L \subseteq L^*$, also wenn

$$\mathbf{x}\cdot\mathbf{y} \in \mathbb{Z} \quad \text{für alle} \quad \mathbf{x},\mathbf{y} \in L,$$

so heißt L **ganzes Gitter**, im Falle $L = L^*$ heißt L **unimodular**.

Hilfssatz 38 *$d(L^*) = d(L)^{-1}$, d.h. insbesondere: Unimodulare Gitter L besitzen die Determinante $d(L) = 1$.* □

Ein ganzes Gitter L heißt **gerade**, wenn für alle $\mathbf{x} \in L$ gilt

$$||\mathbf{x}||^2 = \mathbf{x}\cdot\mathbf{x} \equiv 0 \bmod 2.$$

Der folgende Satz schlägt eine Brücke von diesen Gittereigenschaften zu den Begriffen der Codierungstheorie.

Satz 39 *Sei $C \subseteq \mathbb{F}_2^n$ ein linearer binärer Code und Γ_C sein zugeordnetes Gitter.*

1. *$C \subseteq C^\perp$ genau dann, wenn Γ_C ein ganzes Gitter ist.*
2. *C ist selbstdual genau dann, wenn Γ_C unimodulares Gitter ist.*
3. *C ist doppelt–gerade genau dann, wenn Γ_C gerades Gitter ist.*

Der *Beweis* ist naheliegend; für Punkt 1 seien $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n) \in C$ Codeworte mit Repräsentanten $\mathbf{x},\mathbf{y} \in \mathbb{Z}^n$ im Standardgitter. Dann ist

$$\frac{1}{\sqrt{2}}\mathbf{x}\cdot\frac{1}{\sqrt{2}}\mathbf{y} = \frac{1}{2}\mathbf{x}\cdot\mathbf{y} \equiv \frac{1}{2}(u_1v_1 + \ldots + u_nv_n) \bmod \mathbb{Z}$$

(die rechte Seite ist wohldefiniert in $\frac{1}{2}\mathbb{Z}/\mathbb{Z}$), also genau dann ganz, wenn $u_1v_1 + \ldots + u_nv_n = 0 \in \mathbb{F}_2$. Aus Punkt 1 ergibt sich Punkt 2, und Punkt 3 folgt aus einer entsprechenden Kongruenzbetrachtung mod 4. □

8.4 Packungsdichte und Gewicht

Wir haben schon in Hilfssatz 35.4 gesehen, daß Codes etwas mit dichten Packungen zu tun haben: Perfekte Codes liefern eine dichtestmögliche Packung von Hammingkugeln im $\mathbb{F}_2^n$. Das ist zwar nicht die Metrik, die man bei Packungsgittern zu betrachten hat, trotzdem zeigt es sich, daß gute Codes C etwas mit dichten Kugelpackungen bei den zugehörigen Gittern Γ_C zu tun haben. Was *gut* heißen soll, werden wir gleich präzisieren.

Satz 40 *Sei C ein binärer linearer $[n,k,w]$–Code mit Minimalgewicht $w \le 4$. Dann ist Γ_C ein Packungsgitter für Kugeln K vom Radius $\frac{1}{2}\sqrt{w/2}$ und mit Packungsdichte*

$$\delta(\Gamma_C, K) \;=\; \kappa_n\, w^{n/2}\, 2^{k-2n}\,.$$

(κ_n ist wieder das Volumen der n–dimensionalen Einheitskugel.)
Beweis. Wegen $w \le 4$ haben paarweise verschiedene Punkte von $\rho^{-1}(C)$ voneinander einen euklidischen Abstand $\ge \sqrt{w}$. Daraus folgt, daß Γ_C Packungsgitter für die im Satz angegebenen Kugeln ist. Die Packungsdichte folgt damit aus Hilfssatz 30 und der oben durchgeführten Berechnung der Gitterdeterminante $d(\Gamma_C)$. □

Für festes n erhält man also eine besonders dichte Kugelpackung, wenn einerseits w groß ist und andererseits k, also die Informationsrate des Codes. Als *Beispiel* kann erstens der $[4,3,2]$–Code C^* dienen, der bei Erweiterung des trivialen $[3,3,1]$–Codes entsteht und hier auf die Packungsdichte

$$\kappa_4 \cdot 2^{-3} \;=\; \frac{\pi^2}{16}$$

führt. Es ist genau jenes Gitter maximaler Packungsdichte im $\mathbb{R}^4$, das schon am Ende von Abschnitt 7 erwähnt wurde.
Als zweites Beispiel betrachte man den erweiterten Hammingcode H_3^*, der als $[8,4,4]$–Code auf ein Gitter $L = \Gamma_{H_3^*}$ der Packungsdichte

$$\kappa_8 \cdot 2^{-4} \;=\; \frac{\pi^4}{4!\cdot 16} \;=\; \frac{\pi^4}{2^7 \cdot 3} \;\approx\; 0.254$$

führt (wieder einmal weit dichter als die Minkowski–Hlawka–Schranke aus Satz 31). Auch dieses Gitter, *E_8–Gitter* genannt, ist als Gitter maximaler Packungsdichte in seiner Dimension bekannt (Blichfeldt 1934, Vetčinkin 1980). Man mache sich klar, daß trotz des bescheidenen Zahlenwerts 0.254 diese Kugelpackung ganz außerordentlich dicht ist: Der Fundamentalquader des Gitters $2\mathbb{Z}^8 \subset \rho^{-1}(H_3^*)$ enthält nicht nur die Repräsentanten der Punkte von einer, sondern von insgesamt 16 Kugeln vom Radius 1 ohne Überlappungen!

9 Golay–Code und Leech–Gitter

Satz 40 hat die unerfreuliche Einschränkung $w \leq 4$, weil man in der dort vorgenommenen Konstruktion von Kugelpackungen wegen $2\mathbb{Z}^n \subset \rho^{-1}(C)$ keine Kugeln vom Radius > 1 packen kann. Da man sich für größere w eigentlich noch bessere Packungsresultate erhofft, muß man die Konstruktion modifizieren, was hier an einem besonders wichtigen Beispiel durchgeführt werden soll. Wir holen dazu etwas aus, um zunächst eine weitere Methode der Codekonstruktion vorzustellen.

9.1 Zyklische Codes und quadratische Reste

Ein binärer linearer Code $C \subseteq \mathbb{F}_2^n$ heißt **zyklisch**, wenn mit jedem Codewort $(v_0, \ldots, v_{n-1})$ auch $(v_{n-1}, v_1, \ldots, v_{n-2})$ ein Codewort ist. Ordnet man dem ersten Codewort die Polynomrestklasse

$$p(x) \;=\; v_0 + v_1 x + \ldots + v_{n-1}x^{n-1} \;\in\; \mathbb{F}_2[x]/(x^n-1)\mathbb{F}_2[x]$$

zu, so entspricht der zyklischen Vertauschung der Koordinaten gerade die Abbildung $p(x) \mapsto xp(x)$. Die Menge der Polynomrestklassen für die Codewörter führt daher auf ein Ideal I in $\mathbb{F}_2[x]/(x^n-1)\mathbb{F}_2[x]$. Da sein Urbild in $\mathbb{F}_2[x]$ ein Hauptideal ist, gibt es dann auch eine *erzeugende Polynomrestklasse* $g(x) \bmod (x^n-1)$, und zwar für ein Polynom $g(x)|(x^n-1)$. Eine Abzählung der Restklassen $\bmod\, g$ und $\bmod\,(x^n-1)$ ergibt

$$k \;=\; \dim C \;=\; n - \operatorname{grad} p\,.$$

Beispiele. $g(x) = 1$ führt auf den trivialen Code, $g(x) = 1 + x + \ldots + x^{n-1}$ auf den $[n,1,n]$–Wiederholungscode, und $g(x) = 1 + x + x^3 \bmod (x^7-1)$ auf den Hammingcode H_3 — man überzeuge sich an Hand der Kontrollmatrix davon, daß die den Polynomrestklassen $x^j g(x)$, $j = 0,1,2,3$, entsprechenden Codewörter wirklich in H_3 liegen. Allgemeiner kann man den Hammingcode H_r als zyklischen Code folgendermaßen konstruieren: Für $n = 2^r - 1$ hat das Polynom $x^n - 1$ über dem Körper $\mathbb{F}_{2^r}$ die Primfaktorzerlegung

$$x^n - 1 \;=\; \prod_{i=1}^{n} (x - a^i)$$

mit einem erzeugenden Element a der multiplikativen Gruppe $\mathbb{F}_{2^r}^*$. Man wähle nun g als das irreduzible Polynom von a über $\mathbb{F}_2$, nach der Galoistheorie endlicher Körper also

$$g(x) \;:=\; \prod_{j=0}^{r-1} (x - a^{2^j})\,.$$

Daß hierbei tatsächlich wieder die in Abschnitt 8.2 eingeführten Hammingcodes H_r entstehen, ist nicht ganz einfach einzusehen, für das folgende aber auch unerheblich. Wichtiger ist die

Beobachtung, daß für $r = 3$, $n = 7$ die Exponenten von a genau die quadratischen Reste mod 7 durchlaufen, was zu einer anderen Verallgemeinerung von H_3 Anlaß gibt, den **quadratische Reste–Codes**. Hier sei n eine Primzahl, für die 2 mod n quadratischer Rest ist, also $n \equiv \pm 1 \bmod 8$. Sei jetzt a eine primitive n–te Einheitswurzel über $\mathbb{F}_2$; diese existiert in jedem Körper $\mathbb{F}_{2^r}$, für den $2^r \equiv 1 \bmod n$ ist, also z.B. für $r = n-1$ oder (weil 2 quadratischer Rest mod n ist) sogar für $r = (n-1)/2$. Dann ist

$$g(x) \;:=\; \prod_{q \in Q} (x \,-\, a^q)\,, \quad Q \;:=\; \{\, q \bmod n \,|\, (\frac{q}{n}) = 1 \,\}$$

das Polynom unserer Wahl: Wegen $2 \in Q$ ist g invariant unter dem Frobeniusautomorphismus

$$(\, b \mapsto b^2 \,) \;\in\; \mathrm{Gal}\,\mathbb{F}_{2^r}/\mathbb{F}_2 \;,$$

erfüllt also $g \in \mathbb{F}_2[x]$ und führt zu einem zyklischen $[n, \frac{n+1}{2}, w]$–Code mit Minimalgewicht w.

9.2 Bericht über den Golay–Code

Eine allgemeine Formel für dieses Minimalgewicht ist nicht bekannt, wohl aber untere und obere Abschätzungen, die für kleine n sehr nützlich sind. Außerdem besteht natürlich die Möglichkeit, g explizit zu berechnen, damit eine Generatormatrix aufzulisten und das Minimalgewicht auf direktem Weg zu bestimmen. Beides erfordert einigen Aufwand, deswegen sei hier nur das Resultat für den im Fall $n = 23$ entstehenden **Golay–Code** G_{23} angegeben (den man durchaus nicht nur auf die hier angegebene Weise konstruieren kann, vgl. [Eb]).

Satz 41 *Der Golay-Code G_{23} ist ein perfekter 3-korrigierender $[23, 12, 7]$–Code. Sein erweiterter Code $G_{24} := G_{23}^*$ ist ein selbst–dualer und doppelt–gerader $[24, 12, 8]$–Code.* $\square$

Eine Generator– und Kontrollmatrix von G_{24} kann man als $H = (E, M)$ beschreiben, wo E die 12×12–Einheitsmatrix ist und M eine 12×12–Matrix (m_{ij}) ist, deren Einträge charakterisiert sind durch

- $m_{11} = 0$
- $m_{ij} = 0$ für $i, j \geq 2$ und $i + j - 4$ quadratischer Rest mod 11
- $m_{ij} = 1$ in allen anderen Fällen.

9.3 Das Leech–Gitter

Sei nun $\Lambda := \rho^{-1}(G_{24})$. Wie zu Beginn von Abschnitt 8.3 erläutert, ist

$$[\mathbb{Z}^{24} : \Lambda] \;=\; [\Lambda : 2\mathbb{Z}^{24}] \;=\; d(\Lambda) \;=\; 2^{12}\,,$$

und jedes $\mathbf{x} \in \Lambda$ besitzt eine wohldefinierte Beschreibung als

$$\mathbf{x} = \mathbf{c} + 2\mathbf{y}, \quad \mathbf{y} = (y_1, \ldots, y_{24}) \in \mathbb{Z}^{24}$$

mit einem eindeutig bestimmten Codewort $\mathbf{c} \in G_{24}$. Da der Code doppelt–gerade ist, erfüllen die Koordinaten x_i dieser Gitterpunkte $\sum x_i \equiv 2 \sum y_i \bmod 4$, und wir können ein Untergitter K vom Index 2 in Λ definieren als Kern des Homomorphismus

$$h : \Lambda \to \mathbb{F}_2 : \mathbf{x} \mapsto \sum y_i \equiv \frac{1}{2} \sum x_i .$$

Eine der vielen Möglichkeiten, das **Leech–Gitter** L zu definieren, besteht nun darin, L mit dem üblichen Normierungsfaktor $1/\sqrt{2}$ als Index–2–Obergitter von K

$$L := \frac{1}{\sqrt{2}} \left(K + \left(h^{-1}(1) + \frac{1}{2}(1, \ldots, 1) \right) \right)$$

einzuführen. Gewissermaßen wird also Λ in zwei Teile zerlegt und daraus anders wieder zusammengesetzt. Daß dabei eine diskrete Untermenge des $\mathbb{R}^{24}$ entsteht, ist evident; daß es sich außerdem um ein Gitter handelt, sieht man daran, daß die Summe je zweier Punkte aus $h^{-1}(1)$ wieder in K liegt, ebenso wie der Punkt $\mathbf{1} = (1, \ldots, 1)$ (Kontrollmatrix von G_{24} verwenden!). Aus $d(K) = 2^{13}$ folgt $d(L) = 1$. Da G_{24} selbstdual ist, wird $\Gamma_{G_{24}}$ ein unimodulares Gitter, erst recht bleibt also das Untergitter $\frac{1}{\sqrt{2}}K$ ein ganzes Gitter. Man rechnet leicht nach, daß die Erweiterung L ebenfalls ganz ist und sogar wieder unimodular wird. Man beachte dabei $4 | \sum x_i$ für alle $\mathbf{x} \in h^{-1}(0)$ und

$$4 \,|\, \#\{\, i \,|\, x_i \equiv 1 \bmod 2 \,\} \geq w(G_{24}) = 8$$

für alle $\mathbf{x} \in \Lambda$. Für $\mathbf{x} \in K$ erhalten wir

$$8 \leq ||\mathbf{x}||^2 \equiv 0 \bmod 4$$

und ein entsprechendes Ergebnis auch für $h^{-1}(1) + \frac{1}{2}\mathbf{1}$, denn die Beträge aller Koordinaten der letzteren Punkte sind stets $\geq \frac{1}{2}$, der Betrag mindestens einer Koordinate sogar $\geq \frac{3}{2}$. Nach Umnormierung heißt das

$$4 \leq ||\mathbf{x}||^2 \equiv 0 \bmod 2 \quad \text{für alle} \quad \mathbf{x} \in L ,$$

die Gitterpunkte von L haben also voneinander den Mindestabstand 2. Somit ist L als Packungsgitter für Einheitskugeln brauchbar. Wir fassen diese Überlegungen zusammen:

Satz 42 *Das Leech–Gitter $L \subset \mathbb{R}^{24}$ ist ein gerades unimodulares Gitter. Der Mindestabstand seiner Gitterpunkte ist 2, und seine Packungsdichte für Einheitskugeln K_1 ist*

$$\delta(L, K_1) = \kappa_{24} = \frac{\pi^{12}}{12!} \approx 0.00193 .$$

(Die MINKOWSKI–HLAWKA–Schranke würde wenig mehr als $2^{-23} \approx 10^{-7}$ ergeben!) Es handelt sich in der Tat um die dichteste Kugelgitterpackung in der Dimension 24. Das Leech–Gitter kann noch etliche andere Rekorde aufweisen („Kußzahl", Automorphismengruppe, vgl. [CSl], [Eb]) und spielt in noch vielen anderen Gebieten der Mathematik bis hin zur mathematischen Physik eine bemerkenswerte Rolle. Unsere Raumanschauung ist an den Dimensionen 2 und 3 geschult und hat erhebliche Mühe, sich ein Fundamentalparallelotop vorzustellen, dessen Eckpunkte voneinander nur Abstände ≥ 2 besitzen und das trotzdem nur das Volumen 1 hat.

Literaturverzeichnis

[Ca] J.W.S.Cassels: An Introduction to the Geometry of Numbers. Springer Grundlehren 99.

[Eb] W.Ebeling: Lattices and Codes. Vieweg 1994.

[Hu] M.N.Huxley: Exponential sums and lattice points. II. Proc. London Math. Soc. (3) **66** (1993), 279–301.

[Kr] E.Krätzel: Lattice Points. Kluver 1988.

[Le] M.Leppmeier: Kugelpackungen, von Kepler bis heute. Vieweg 1997.

[Oe] J.Oesterlé: Densité maximale des empilements de sphères en dimension 3 [d'après Thomas C. Hales et Samuel P. Ferguson], Sém. Bourbaki 1998/99, Exp. 863, pp. 405–413.

[Ro] C.A.Rogers: Packing and Covering, Cambridge UP 1964.

[Wo] J.Wolfart: Einführung in die Zahlentheorie und Algebra. Vieweg 1996.