Vorlesungen von Prof. Dr. C.P. Schnorr:

Gittertheorie und algorithmische Geometrie, Reduktion von Gitterbasen und Polynomidealen

an der Johann Wolfgang Goethe-Universität Frankfurt/Main
im Sommersemester 1994 und Wintersemester 1994/95

β_4-Version

Mitschrift Roger Fischlin

26. Oktober 2001

Dieses ist eine modifizierte Mitschrift der beiden Vorlesungen

- „Gittertheorie und algorithmische Geometrie“ und
- „Reduktion von Gitterbasen und Polynomidealen“,

die Prof. Dr. C.P. Schnorr im Sommersemester 1994 bzw. im Wintersemester 1994/95 an der Johann-Wolfgang-Goethe-Universität in Frankfurt am Main gehalten hat. Zusätzlich habe ich zahlreiche Zwischenschritte bei Beweisen eingefügt, relevante Übungsaufgaben und Beispiele integriert und Querverweise zwischen beiden Teilen hergestellt.

Meine Mitschrift habe ich während des Wintersemesters 1995/96 und des Sommersemesters 1996 in LaTeX 2_ε gesetzt. Für Beantwortung von Fragen, Korrekturlesen und Literaturhinweise usw. bedanke ich mich bei Philipp Beckmann, Marc Fischlin, Ulrich Mehringer, Harald Ritter, Julia Steinbach und Annegret Weng.

Für die Zukunft, wenn es die Zeit erlaubt, plane ich, die Disserationen von M. Kaib [Kaib94] über den Gauß-Algorithmus (Kapitel 4) und von H. Ritter [Ritter97] über den ENUM-Algorithmus (Kapitel 8) aufzunehmen und genauer auf das Problem des kürzesten und nächsten Gittervektors (insbesondere die Arbeit von L. Babai [Babai86]) einzugehen. Hinweise auf Fehler und Anregungen zum Skript bitte an meine e-mail-Adresse:

`fischlin@rbi.informatik.uni-frankfurt.de`

Roger Fischlin

Inhaltsverzeichnis

Grundlagen

Notation

Mit $M_{m,n}(S)$ bezeichnen wir die Menge aller $m \times n$-Matrizen mit Einträgen aus der Menge S. Zum Beispiel ist $M_{m,n}(\mathbb{Z})$ die Menge aller ganzzahligen $m \times n$-Matrizen. Zur Matrix B bezeichne B^T die transponierte Matrix. Die Elemente aus $\mathbb{Z}^n$, $\mathbb{R}^n$, etc. schreiben wir, sofern nicht anders angegeben, als Spaltenvektoren.

Zur reellen Zahl r bezeichne $\lceil r \rfloor := \left\lceil r - \frac{1}{2} \right\rceil$ die nächste ganze Zahl. Wir schreiben $\mathbb{R}^+ := \{ x \in \mathbb{R} \mid x > 0 \}$ für die Menge der positiven, reellen Zahlen.

Skalarprodukt

Der Vektorraum $\mathbb{R}^n$ sei mit einem beliebigen Skalarprodukt $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ ausgestattet (*Euklidischer Vektorraum*). Das *Skalarprodukt* hat die folgenden Eigenschaften: Für alle $u, v, w \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$ gilt:

- $\langle \cdot, \cdot \rangle$ ist bilinear:

$$\begin{aligned} \langle u + w, v \rangle &= \langle u, v \rangle + \langle w, v \rangle \\ \langle \lambda u, v \rangle &= \lambda \langle u, v \rangle \\ \langle u, v + w \rangle &= \langle u, v \rangle + \langle u, w \rangle \\ \langle u, \lambda v \rangle &= \lambda \langle u, v \rangle \end{aligned}$$

- $\langle \cdot, \cdot \rangle$ ist symmetrisch:

$$\langle u, v \rangle = \langle v, u \rangle$$

- $\langle \cdot, \cdot \rangle$ ist positiv definit:

$$\langle u, u \rangle > 0 \qquad \text{für } u \neq 0$$

Für $u = 0$ folgt aus der Linearität in jeder Komponente $\langle u, u \rangle = 0$. Das *Standard-Skalarprodukt* ist definiert als:

$$\left\langle (u_1, u_2, \ldots, u_n)^\mathsf{T}, (v_1, v_2, \ldots, v_n)^\mathsf{T} \right\rangle := \sum_{i=1}^{n} u_i v_i$$

Den meisten Anwendungen liegt das Standard-Skalarprodukt zugrunde. Jedes Skalarprodukt $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ läßt sich schreiben als:

$$\langle u, v \rangle := u^\mathsf{T} S v$$

mit symmetrischer Matrix $S \in M_{n,n}(\mathbb{R})$. Im Fall des Standard-Skalarprodukts ist die Matrix S die Identität.

Normen

Eine Abbildung $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ heißt *Norm*, falls für alle $u, v \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$ gilt:

$$\begin{aligned} \|\lambda v\| &= |\lambda| \cdot \|v\| && \text{(positive Homogenität)} \\ \|u + v\| &\leq \|u\| + \|v\| && \text{(Dreiecksungleichung)} \\ \|u\| &\geq 0 \qquad \text{für } u \neq 0 && \text{(positive Definitheit)} \end{aligned}$$

Die reelle Zahl $\|u\|$ heißt *Norm* (oder *Länge*) des Vektors $u = (u_1, u_2, \ldots, u_n)$. Aus einem Skalarprodukt erhält man die *Euklidische Norm* durch:

$$\|u\| := \sqrt{\langle u, u \rangle}$$

Die ℓ_1-*Norm* oder auch *Betragsnorm* ist:

$$\left\| (u_1, u_2, \ldots, u_n)^\mathsf{T} \right\|_1 := \sum_{i=1}^{n} |u_i|$$

Die ℓ_2-*Norm* erhält man aus dem Standard-Skalarprodukt:

$$\left\| (u_1, u_2, \ldots, u_n)^\mathsf{T} \right\|_2 := \sqrt{\langle u, u \rangle} = \left(\sum_{i=1}^{n} u_i^2 \right)^{\frac{1}{2}}$$

Allgemein: Die ℓ_p-*Norm* ist:

$$\left\| (u_1, u_2, \ldots, u_n)^\mathsf{T} \right\|_p := \left(\sum_{i=1}^{n} |u_i|^p \right)^{\frac{1}{p}}$$

Die *sup-Norm*, *Maximums-Norm* oder auch ℓ_∞-*Norm* ist:

$$\left\| (u_1, u_2, \ldots, u_n)^\mathsf{T} \right\|_\infty := \max_{i=1,2,\ldots,n} |u_i|$$

Ungleichungen

Für die sup-, Betrags- und 2-Norm eines Vektors $u \in \mathbb{R}^n$ gelten die folgenden Beziehungen:

$$\begin{aligned} \|u\|_2 \leq \|u\|_1 &\leq \sqrt{n} \cdot \|u\|_2 \\ \|u\|_\infty \leq \|u\|_2 &\leq n \cdot \|u\|_\infty \end{aligned}$$

Für die Beziehung Skalarprodukt und zugehörige Norm $\|u\| := \sqrt{\langle u, u \rangle}$ gilt die *Cauchy-Schwarz-Ungleichung* (seien $u, v \in \mathbb{R}^n$):

$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\|$$

Die Gleichheit gilt genau dann, wenn beide Vektoren linear abhängig sind. Seien $b_1, b_2, \ldots, b_n \in \mathbb{R}^n$ die Spaltenvektoren (oder Zeilenvektoren) der Matrix $B \in M_{n,n}(\mathbb{R})$. Die *Hadamard'sche Ungleichung* besagt:

$$\det B \leq \prod_{i=1}^{n} \|b_i\|_2$$

Sind die Vektoren $b_1, b_2, \ldots, b_n$ orthogonal, gilt die Gleichheit.

Kapitel 1

Komplexität, $\mathcal{NP}$-Vollständigkeit

Wir fassen mit Hinblick auf die Gittertheorie die Grundbegriffe der Komplexitätstheorie, speziell die $\mathcal{NP}$-Vollständigkeit, zusammen.

1.1 $\mathcal{NP}$-Vollständigkeit

Wir definieren die Bitlänge endlicher Objekte (das Vorzeichen speichern wir getrennt):

- $\ell(0) := 1$
- $\ell(n) := \lceil \log_2(n+1) \rceil$ für $n \in \mathbb{N}$
- $\ell\left(\frac{p}{q}\right) := \ell(p) + \ell(q)$ mit $p, q \in \mathbb{N}$ und $\mathrm{ggT}(p, q) = 1$
- $\ell(A) = \sum_{i,j} \ell(a_{ij})$ für $A = [a_{ij}] \in M_{m,n}(\mathbb{Q})$

Wir setzen die Laufzeit des Algorithmus' in Beziehung zur Eingabelänge. Wir interessieren uns für Polynomialzeit-Verfahren:

Definition 1.1.1 (Polynomialzeit)
Ein Algorithmus ist in Polynomialzeit, falls die Schrittzahl (Turing-Maschine oder Anzahl Bit-Operationen) polynomiell in der Länge der Eingabe beschränkt ist:

$$\textit{Schrittzahl(Eingabe)} = \mathrm{poly}\,(\ell(\textit{Eingabe}))$$

In der theoretischen Informatik betrachtet man die Polynomialzeit-Algorithmen als effizient.

Definition 1.1.2 (Charakteristische Funktion)
Zu einer Menge $A \subseteq \{0,1\}^*$ ist die charakteristische Funktion $\chi_A : \{0,1\}^* \to \{0,1\}$ definiert durch: $\chi_A(a) = 1$ genau dann, wenn $a \in A$ ist.

Wir definieren mit charakteristischen Funktionen die Klasse der Polynomialzeit-Sprachen:

Definition 1.1.3 (Klasse $\mathcal{P}$ der Polynomialzeit-Sprachen)
Die Klasse $\mathcal{P}$ der Polynomialzeit-Sprachen besteht genau aus den Sprachen $A \subseteq \{0,1\}^*$, für welche die charakteristische Funktion χ_A in Polynomialzeit berechenbar ist.

Die Klasse $\mathcal{NP}$ umfaßt die Sprache, so daß es genau für jedes Wort aus der Sprache einen Bitstring gibt, anhand dessen wir effizient überpüfen können, daß dieses Wort in der Sprache liegt.

Definition 1.1.4 (Klasse $\mathcal{NP}$)
Die Klasse $\mathcal{NP}$ der nichtdeterministischen Polynomialzeit-Sprachen $A \subseteq \{0,1\}^*$ ist erklärt durch:

$$A \in \mathcal{NP} \iff \begin{array}{l} \exists B \in \{0,1\}^* \times \{0,1\}^*, B \in \mathcal{P} : \\ A = \left\{ x \in \{0,1\}^* \;\middle|\; \exists y \in \{0,1\}^{\mathrm{poly}(\ell(x))} \text{ mit } (x,y) \in B \right\} \end{array}$$

Sei $(x,y) \in B$. Dann heißt y Zeuge für $x \in A$.

Die Cook'sche Hypothese ist $\mathcal{P} \neq \mathcal{NP}$, d.h. es gibt Sprachen in der Klasse $\mathcal{NP}$, für die wir nicht in Polynomialzeit einen Zeugen finden können.

Definition 1.1.5 (Cook-Karp-Reduktion)
Seien $A, B \subseteq \{0,1\}^*$:

$$A \leq_{pol} B \iff \begin{array}{l} \exists \text{ Polynomialzeit-Abbildung } h \text{ mit:} \\ \forall x \in \{0,1\}^* : x \in A \Leftrightarrow h(x) \in B \end{array}$$

Aus $A \leq_{\mathrm{pol}} B$ und $B \leq_{\mathrm{pol}} C$ folgt $A \leq_{\mathrm{pol}} C$. Es gilt $A \leq_{\mathrm{pol}} B$, wenn man $x \in A$ in Polynomialzeit mit einmaliger Anwendung eines Orakels für B entscheiden kann.

Definition 1.1.6 ($\mathcal{NP}$-vollständig)
$A \subseteq \{0,1\}^*$ heißt $\mathcal{NP}$-vollständig, wenn:

1. $A \in \mathcal{NP}$
2. $\forall B \in \mathcal{NP} : B \leq_{pol} A$

Falls wir einen Polynomialzeit-Algorithmus zu einem $\mathcal{NP}$-vollständigen Problem finden, folgt $\mathcal{P} = \mathcal{NP}$. Dies würde der Cook'schen Hypothese widersprechen. Daher gelten die $\mathcal{NP}$-vollständigen Probleme als die schwierigsten in $\mathcal{NP}$.

1.2 Schwierige, algorithmische Gitterprobleme

Wir lernen in diesem Abschnitt mit der Gittertheorie verbundene Probleme kennen, die $\mathcal{NP}$-vollständig sind oder für die bisher keine effizienten Algorithmen bekannt sind. Ein solches Problem ist die ganzzahlige, lineare Programmierung (Integer Programming):

Definition 1.2.1 (Ganzzahlige, lineare Programmierung)
Das Problem der ganzzahligen, linearen Programmierung lautet:

- *Gegeben: $m, n \in \mathbb{N}$, $A \in M_{m,n}(\mathbb{Z})$ und $b \in \mathbb{Z}^m$*
- *Finde $x \in \mathbb{Z}^n$ mit $Ax \leq b$ oder zeige, daß kein solcher Vektor existiert.*

Die ganzzahlige, lineare Programmierung ist „schwierig". Wir werden in Satz 1.2.5 sehen, daß das zugehörige Entscheidungsproblem $\mathcal{NP}$-vollständig ist:

Definition 1.2.2 (Entscheidungsproblem der ganzzahligen, linearen Programmierung)
Das Problem der ganzzahligen, linearen Programmierung lautet:

- *Gegeben:* $m, n \in \mathbb{N}$, $A \in M_{m,n}(\mathbb{Z})$ *und* $b \in \mathbb{Z}^m$
- *Entscheide, ob ein* $x \in \mathbb{Z}^n$ *mit* $Ax \leq b$ *existiert.*

Falls $\mathcal{P} \neq \mathcal{NP}$, gibt es keinen Lösungsalgorithmus in Polynomialzeit. Dagegen gibt es zum analogen Problem der rationalen, linearen Programmierung Polynomialzeit-Verfahren:

Definition 1.2.3 (Rationale, lineare Programmierung)
Das Problem der rationalen, linearen Programmierung lautet:

- *Gegeben:* $m, n \in \mathbb{N}$, $A \in M_{m,n}(\mathbb{Z})$ *und* $b \in \mathbb{Q}^m$
- *Finde* $x \in \mathbb{Q}^n$ *mit* $Ax \leq b$ *oder zeige, daß kein solcher Vektor existiert.*

Das erste Polynomialzeit-Verfahren für die lineare Programmierung ist die Ellipsoid-Methode von L.G. Khachiyan [Khach79, Khach80]. Diese Methode ist aber nicht praktikabel. Ein bekannter Polynomialzeit-Algorithmus stammt von M. Karmarkars [Karma84]. Dieser hat zur Entwicklung der Interior-Point-Methoden für die lineare Programmierung geführt. Ein bekannter Interior-Point-Algorithmus stammt von Y. Ye [Ye91]. Ein einfaches, praktisches Verfahren ist der Simplex-Algorithmus [Dantzig63, Schrijver86] von G.B. Dantzig, der allerdings im Wortcase exponentielle Laufzeit haben kann. Weitere Probleme, die man in Polynomialzeit lösen kann, sind:

Satz 1.2.4 (Sieveking 1976)
Folgende Probleme sind zu gegebenen $m, n \in \mathbb{N}$, $A \in M_{m,n}(\mathbb{Z})$ *und* $b \in M_{m,1}(\mathbb{Z})$ *in Polynomialzeit lösbar:*

a) Löse $Ax = b$, $x \in \mathbb{Z}^n$ *oder weise Unlösbarkeit nach.*

b) Finde eine $\mathbb{Z}$*-Basis* $b_1, b_2, \ldots, b_k$ *von* $\{x \in \mathbb{Z}^n \mid Ax = 0\}$*, dem* $\mathbb{Z}$*-Kern. Eine* $\mathbb{Z}$*-Basis besteht aus linear unabhängigen Vektoren* $b_1, b_2, \ldots, b_k$*, so daß:*

$$\{x \in \mathbb{Z}^n \mid Ax = 0\} = \left\{ \sum_{i=1}^{k} t_i b_i \;\middle|\; t_1, t_2, \ldots, t_k \in \mathbb{Z} \right\}$$

Beweis. Modifikation des Gauß-Eliminationsverfahrens (M. Sieveking in [SpStr76]). Alternativer Beweis in [KaBa79]. ■

Im folgenden Satz führen wir weitere mit der Gittertheorie verbundene Aufgaben bzw. Entscheidungsprobleme auf, die $\mathcal{NP}$-vollständig sind.

Satz 1.2.5
Folgende Sprachen sind $\mathcal{NP}$*-vollständig:*

1. *Integer-Programming:*

$$\text{IP} := \left\{ (m, n, A, b) \;\middle|\; \begin{array}{l} A \in M_{m,n}(\mathbb{Z}), b \in \mathbb{Z}^m, \\ \exists x \in \mathbb{Z}^n : Ax \leq b \end{array} \right\}$$

2. Rucksack (Knapsack) oder Subsetsum:

$$\text{SubsetSum} := \left\{ (n, a_1, a_2, \ldots, a_n, b) \in \mathbb{N}^{n+2} \;\middle|\; \exists x \in \{0,1\}^n : \sum_{i=1}^{n} a_i x_i = b \right\}$$

3. $\{0,1\}$*-Integer-Programming:*

$$\{0,1\}\text{-IP} := \left\{ (m, n, A, b) \;\middle|\; \begin{array}{c} A \in M_{n,m}(\mathbb{Z}), b \in \mathbb{Z}^m, \\ \exists x \in \{0,1\}^n : Ax \leq b \end{array} \right\}$$

4. Schwache Zerlegung:

$$\left\{ (n, a_1, a_2, \ldots, a_n) \in \mathbb{N}^{n+1} \;\middle|\; \exists\, (x_1, x_2, \ldots, x_n) \in \{0, \pm 1\}^n \setminus \{0^n\} : \sum_{i=1}^{n} a_i x_i = 0 \right\}$$

Beweis. Für 1,2,3 siehe [GaJo79, SpStr76], für 4 siehe [EmBoas81]. Den Nachweis, daß es für die Sprache Integer-Programming polynomiell lange Zeugen gibt, also IP $\in \mathcal{NP}$, werden wir in Satz 1.2.6 führen. ■

Satz 1.2.6 (von zur Gathen, Sieveking 1978)
IP $\in \mathcal{NP}$.

Beweis. Wir wählen als Zeugen für $(m, n, A, b) \in$ IP ein geeignetes $x \in \mathbb{Z}^n$ mit $Ax \leq b$. Offenbar existiert x genau dann, wenn $(m, n, A, b) \in$ IP. Wir müssen noch zeigen, daß der Zeuge polynomielle Länge hat.

Sei $A =: (a_{ij})_{ij}$ und $b =: (b_1, b_2, \ldots, b_m)^{\mathsf{T}}$. Setze $M := \max_{i,j} \{|a_{ij}|, |b_i|\}$. Nach [GaSi78] gilt:

$$(\exists x \in \mathbb{Z}^n : Ax \leq b) \quad \Longleftrightarrow \quad \left(\exists x \in \mathbb{Z}^n : Ax \leq b,\ \|x\|_\infty \leq (n+1) n^{\frac{n}{2}} M^n\right)$$

Die obere Schranke von $\|x\|_\infty$ impliziert, daß die Länge des Zeugen x polynomiell in der Länge von A und b beschränkt ist. Wegen $\ell(m, n, A, b) \geq nm + \log_2 M$ gilt:

$$\ell(x) = \mathcal{O}\left(n^2(\log n + \log M\right) = \mathcal{O}\left(\ell(m, n, A, b)^3\right)$$

■

Wir definieren die Begriffe, die elementar für die weiteren Kapitel sind:

Definition 1.2.7 (Gitter, Basis, Dimension, Rang)
Seien $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ linear unabhängige Vektoren. Wir nennen die additive Untergruppe

$$L(b_1, b_2, \ldots, b_n) := \sum_{i=1}^{n} b_i \mathbb{Z} = \left\{ \sum_{i=1}^{n} t_i b_i \;\middle|\; t_1, t_2, \ldots, t_m \in \mathbb{Z} \right\}$$

des $\mathbb{R}^m$ ein Gitter mit der Basis $b_1, b_2, \ldots, b_n$. Ist die Reihenfolge der Basisvektoren fest, sprechen wir von einer geordneten Basis. Der Rang oder auch die Dimension des Gitters ist $\text{Rang}(L) := n$.

Betrachten wir ein Beispiel:

Beispiel 1.2.8 (Gitter)
$\mathbb{Z}^m$ ist ein Gitter vom Rang m, die Einheitsvektoren bilden eine Basis. Zur Matrix $A \in M_{m,n}(\mathbb{Z})$ ist $\{x \in \mathbb{Z}^n \mid Ax = 0\}$ ein Gitter vom Rang $n - \text{Rang}(A)$; nach Satz 1.2.4 können wir in Polynomialzeit eine Basis konstruieren. ◇

Wir versuchen, durch Gitterreduktion einen kürzesten, nicht-trivialen Gittervektor zu finden. Im Fall der sup-Norm ist dies unter der Annahme $\mathcal{P} \neq \mathcal{NP}$ nicht immer effizient möglich:

Korollar 1.2.9
Das Problem $\|\cdot\|_\infty$-kürzester Gittervektor

$$L_\infty\text{-SVP} := \left\{ (m, n, b_1, b_2, \ldots, b_n) \;\middle|\; \begin{array}{l} m, n \in \mathbb{N}, b_1, b_2, \ldots, b_n \in \mathbb{Z}^m, \\ \exists x \in L(b_1, b_2, \ldots, b_n) : \; \|x\|_\infty = 1 \end{array} \right\}$$

ist $\mathcal{NP}$-vollständig.

Beweis. Das Problem $\|\cdot\|_\infty$-kürzester Gittervektor liegt in $\mathcal{NP}$: Als Zeugen wählt man einen Vektor $x \in L(b_1, b_2, \ldots, b_n) \setminus \{0\}$ mit $\|x\|_\infty = 1$. Das $\mathcal{NP}$-vollständige Problem „schwache Zerlegung“ aus Satz 1.2.5 kann in Polynomialzeit auf $\|\cdot\|_\infty$-kürzester Gittervektor reduziert werden. ■

Beim Problem des kürzesten Gittervektors in der ℓ_2-Norm soll man zu gegebener Gitterbasis $b_1, b_2, \ldots, b_n$ und k entscheiden, ob es einen Gittervektor $z \in L(b_1, b_2, \ldots, b_n)$ gibt mit $z \neq 0$ und $\|z\|_2 \leq \sqrt{k}$.

Definition 1.2.10 (Shortest Vector Problem SVP)
Die Sprache zum kürzesten Gittervektorproblem (Shortest Vector Problem) für die ℓ_2-Norm lautet:

$$L_2\text{-SVP} := \left\{ (k, m, n, b_1, b_2, \ldots, b_n) \;\middle|\; \begin{array}{l} k, m, n, \in \mathbb{N}, b_1, b_2, \ldots, b_n \in \mathbb{Z}^m, \\ \exists x \in L(b_1, b_2, \ldots, b_n) \setminus \{0\} : \; \|x\|_2^2 \leq k \end{array} \right\}$$

Der Status dieses Problems ist offen. Anstrengungen, die Vermutung, daß L_2-SVP $\mathcal{NP}$-hart ist, nachzuweisen, sind im Gegensatz zur sup-Norm (siehe Korollar 1.2.9) bislang fehlgeschlagen (vergleiche [Kannan87]).

Das Problem des kürzesten Gittervektors ist der homogene Spezialfall des Problems nächster Gittervektor, von dem man aber weiß, daß es (auch) in der ℓ_2-Norm $\mathcal{NP}$-vollständig ist:

Satz 1.2.11 (Closest Vector Problem CVP)
Das Problem ℓ_2-nächster Gittervektor

$$L_2\text{-CVP} := \left\{ (k, m, n, b_1, b_2, \ldots, b_n, z) \;\middle|\; \begin{array}{l} k, m, n, \in \mathbb{N}, b_1, b_2, \ldots, b_n, z \in \mathbb{Z}^m, \\ \exists x \in L(b_1, b_2, \ldots, b_n) : \; \|z - x\|_2^2 \leq k \end{array} \right\}$$

ist $\mathcal{NP}$-vollständig.

Beweis. Siehe [Kannan87, Theorem6.2]. ■

Wir fassen zusammen: Zu gegebener Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$ sind folgende Aufgaben nach heutigem Stand schwierige, algorithmische Gitterprobleme:

- Finde kurze Gittervektoren ungleich dem Nullvektor.
- Finde eine Basis bestehend aus kurzen Gittervektoren.
- Finde zu gegebenem $z \in \operatorname{span}(b_1, b_2, \ldots, b_n)$ einen möglichst nahen Gittervektor.

Dagegen kann man in Polynomialzeit zu einem gegebenen Erzeugendensystem $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$ des Gitters L, $n \geq \operatorname{Rang}(L)$, eine Gitterbasis konstruieren.

Kapitel 2

Einführung in die Gittertheorie

Wir definieren die Hermite-Normalform zu einer Matrix und zeigen für ganzzahlige Matrizen die Eindeutigkeit. Wir charakterisieren Gitter als diskrete, additive Untergruppen des $\mathbb{R}^m$. Wir behandeln die Menge aller Basen eines Gitters, primitive Systeme von Gittervektoren, die Gitterdeterminante sowie das einer Gitterbasis zugehörige Orthogonalsystem. Wir zeigen, daß jedes Gitter eine längenreduzierte und eine gewichtsreduzierte Basis hat.

2.1 Bezeichnungen

Der Vektorraum $\mathbb{R}^m$ sei mit einem beliebigen Skalarprodukt $\langle \cdot, \cdot \rangle : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ ausgestattet. Es bezeichne $\|x\| = \sqrt{\langle x, x \rangle}$ die Länge des Vektors x. Zu linear unabhängigen Vektoren $b_1, b_2, \ldots, b_n$ bezeichne

$$L(b_1, b_2, \ldots, b_n) := \sum_{i=1}^{n} b_i \mathbb{Z}$$

das Gitter mit Basis $b_1, b_2, \ldots, b_n$. Zu beliebigen Vektoren $b_1, b_2, \ldots, b_n$ sei

$$\operatorname{span}(b_1, b_2, \ldots, b_n) := \sum_{i=1}^{n} b_i \mathbb{R}$$

der von den Vektoren $b_1, b_2, \ldots, b_n$ aufgespannte lineare Raum und

$$\operatorname{span}(b_1, b_2, \ldots, b_n)^{\perp} := \{ y \in \mathbb{R}^m \mid \langle y, b_i \rangle = 0 \text{ für } i = 1, 2, \ldots, n \}$$

das orthogonale Komplement im $\mathbb{R}^m$.

Eine ganzzahlige Matrix mit Determinante ± 1 nennen wir *unimodular*. Die Menge aller unimodularen $n \times n$-Matrizen bezeichnen wir mit $\mathrm{GL}_n(\mathbb{Z})$:

Definition 2.1.1 ($\mathrm{GL}_n(\mathbb{Z})$)
$\mathrm{GL}_n(\mathbb{Z})$ *ist die Gruppe der ganzzahligen $n \times n$-Matrizen mit Determinante ± 1:*

$$\mathrm{GL}_n(\mathbb{Z}) := \{ A \in M_{n,n}(\mathbb{Z}) \mid \det A = \pm 1 \}$$

Wir überlegen uns, daß $\mathrm{GL}_n(\mathbb{Z})$ eine Gruppe ist. Die Einheitsmatrix ist unimodular und mit $S, T \in \mathrm{GL}_n(\mathbb{Z})$ ist wegen $\det(ST) = \det S \cdot \det T$ auch das Produkt eine unimodulare Matrix. Sei $T \in \mathrm{GL}_n(\mathbb{Z})$. Da

$$\det\left(T^{-1}\right) = \frac{1}{\det T}$$

ist $\det\left(T^{-1}\right) = \pm 1$, und nach der Cramer'schen Regel steht an der Position (i, j) der Matrix T^{-1}

$$\frac{(-1)^{i+j} \cdot \det T_{ij}}{\det T} = \pm \det T_{ij},$$

wobei T_{ij} aus T durch Streichen der i-ten Zeile und j-ten Spalte gebildet wird. Da T_{ij} eine ganzzahlige Matrix, folgt, daß $\det T_{ij} \in \mathbb{Z}$. Für $T \in \mathrm{GL}_n(\mathbb{Z})$ gilt $T^{-1} \in \mathrm{GL}_n(\mathbb{Z})$. Die folgenden *Spalten-Elementaroperationen* können bei einer Matrix durch Multiplikation von links mit unimodularen Matrizen realisiert werden:

- Vertauschen zweier Spalten
- Multiplikation einer Spalte mit -1
- Addition eines ganzzahligen Vielfachen einer Spalte zu einer anderen

Man kann zeigen, daß jede unimodulare Matrix als Produkt dieser drei Matrizentypen dargestellt werden kann.

2.2 Grundbegriffe und Eigenschaften

In diesem Abschnitt definieren wir Grundbegriffe der Gittertheorie und werden erste, elementare Eigenschaften beweisen.

2.2.1 Diskrete, additive Untergruppen des $\mathbb{R}^m$ und Gitter

Wir definieren:

Definition 2.2.1 (Diskrete Menge)
Eine Menge $S \subseteq \mathbb{R}^m$ heißt diskret, wenn S keinen Häufungspunkt in $\mathbb{R}^m$ hat.

Es gilt:

Lemma 2.2.2
Sei $G \subseteq \mathbb{R}^m$ eine additive Gruppe. Dann sind folgende Aussagen äquivalent:

a) *G ist diskret.*

b) *0 ist kein Häufungspunkt von G.*

c) *$\{x \in G : \|x\| < r\}$ ist endlich für alle $r > 0$.*

d) $\inf\{\|x - y\| : x \neq y, x, y \in G\} > 0$

Der Beweis zu Lemma 2.2.2 ist elementar und sei dem Leser zur Übung überlassen.

Satz 2.2.3
Jede diskrete, additive Untergruppe des $\mathbb{R}^m$ ist ein Gitter, d.h. die Gruppe wird von einer Gitterbasis erzeugt.

Bevor wir den Satz beweisen, wollen wir zuerst eine Folgerung machen:

Korollar 2.2.4
$L \subseteq \mathbb{R}^m$ ist genau dann ein Gitter, wenn L eine diskrete, additive Untergruppe des $\mathbb{R}^m$ ist.

Beweis (zu Korollar 2.2.4). Wir zeigen beide Richtungen:

„⇐“ Behauptung folgt aus Satz 2.2.3.

„⇒“ Zu zeigen: Jedes Gitter $L := L(b_1, b_2, \ldots, b_n) \subseteq \mathbb{R}^m$ ist diskret. Sei $\varphi : \mathbb{R}^n \to \operatorname{span}(L)$ die lineare Abbildung mit

$$\varphi(t_1, t_2, \ldots, t_n) = \sum_{i=1}^{n} t_i b_i$$

φ ist ein Isomorphismus mit $\varphi(\mathbb{Z}^n) = L$. Weil $\mathbb{Z}^n$ diskret und φ^{-1} stetig auf $\operatorname{span}(L)$ ist, folgt, daß L diskret ist.

■

Beweis (zu Satz 2.2.3). Sei $L \subseteq \mathbb{R}^m$ eine diskrete, additive Untergruppe. Sei n die Maximalzahl linear unabhängiger Vektoren in L. Es gilt $n \leq m$. Durch Induktion über n zeigen wir: L ist ein Gitter vom Rang n:

- $n = 1$

  Sei $b \in L$ ein kürzester Vektor mit $b \neq 0$ (ein solcher Vektor existiert, da 0 kein Häufungspunkt von L ist). Dann gilt $L(b) = L$.

- $n > 1$

  Wähle $b_1 \in L \setminus \{0\}$ mit $\frac{1}{k} \cdot b_1 \notin L$ für alle $k \geq 2$. Dann gilt:

  $$L(b_1) = L \cap \operatorname{span}(b_1) \tag{2.1}$$

  Die orthogonale Projektion $\pi : \mathbb{R}^m \to \operatorname{span}(b_1)^{\perp}$ ist erklärt durch:

  $$\pi(b) = b - \frac{\langle b_1, b \rangle}{\langle b_1, b_1 \rangle} \cdot b_1$$

  Die Induktionsbehauptung ergibt sich aus folgenden Aussagen:

  1. $\pi(L)$ ist diskret und ein Gitter vom Rang $n - 1$.
  2. Für jede Basis $\pi(b_2), \pi(b_3), \ldots, \pi(b_n)$ von $\pi(L)$ mit $b_2, b_3, \ldots, b_n \in L$ gilt:

     $$L = L(b_1, b_2, \ldots, b_n)$$

  Beweis der beiden Aussagen:

  1. Wir zeigen, daß 0 kein Häufungspunkt von $\pi(L)$ ist. Angenommen, $\left(y^{(i)}\right)_{i \in \mathbb{N}}$ sei eine Folge in L, so daß die Vektoren $\pi\left(y^{(i)}\right)$ paarweise verschieden sind und $\lim\limits_{i \to \infty} \pi\left(y^{(i)}\right) = 0$. Zu diesen Vektoren

     $$\pi\left(y^{(i)}\right) = y^{(i)} - \frac{\langle y^{(i)}, b_1 \rangle}{\langle b_1, b_1 \rangle} \cdot b_1$$

     erhalten wir kurze π-Urbilder $\overline{y}^{(i)}$ nach der Vorschrift:

     $$\overline{y}^{(i)} := y^{(i)} - \underbrace{\left\lceil \frac{\langle y^{(i)}, b_1 \rangle}{\langle b_1, b_1 \rangle} \right\rfloor}_{\text{ganzzahlig}} b_1$$

Es gilt: $\overline{y}^{(i)} \in L$, $\pi\left(\overline{y}^{(i)}\right) = \pi\left(y^{(i)}\right)$ und:

$$\left\|\overline{y}^{(i)} - \pi\left(y^{(i)}\right)\right\| \leq \tfrac{1}{2}\,\|b_1\|$$

Da $\lim\limits_{i\to\infty} \left\|\pi\left(\overline{y}^{(i)}\right)\right\| = 0$, gibt es unendlich viele Vektoren $\overline{y}^{(i)} \in L$ mit:

$$\left\|\pi\left(\overline{y}^{(i)}\right)\right\| \leq \|b_1\|$$

Dies ist ein Widerspruch, da L diskret ist, sowie nach Annahme $\pi\left(y^{(i)}\right)$ paarweise verschieden sind und $\pi\left(\overline{y}^{(i)}\right) = \pi\left(y^{(i)}\right)$.
Damit ist 0 kein Häufungspunkt von $\pi(L)$, und $\pi(L)$ ist nach Lemma 2.2.2 diskret. Die Maximalzahl der linear unabhängigen Vektoren in $\pi(L)$ ist $n-1$. Nach Induktionsvoraussetzung ist $\pi(L)$ ein Gitter vom Rang $n-1$.

2. Sei $\pi(b_2), \pi(b_3), \ldots, \pi(b_n)$ eine Basis von $\pi(L)$ mit $b_2, b_3, \ldots, b_n \in L$. Wir zeigen, daß $L \subseteq L(b_1, b_2, \ldots, b_n)$. Sei $b \in L$. Wegen

$$\pi(b) \in \pi(L) = L\left(\pi(b_2), \pi(b_3), \ldots, \pi(b_n)\right)$$

gibt es ein $\overline{b} \in L(b_2, b_3, \ldots, b_n)$ mit $\pi(b) = \pi\left(\overline{b}\right)$. Es gilt $b - \overline{b} \in \operatorname{span}(b_1)$. Nach Wahl von b_1 und wegen

$$b - \overline{b} \in (L \cap \operatorname{span}(b_1)) \overset{(2.1)}{=} L(b_1)$$

folgt $b - \overline{b} \in L(b_1)$. Es gilt daher $b \in L(b_1, b_2, \ldots, b_n)$.

■

Im obigen Beweis haben wir insbesondere gezeigt, daß der Rang eines Gitters gleich der Maximalzahl der linear unabhängigen Gittervektoren ist. Es ist bequem, sich bei der Konstruktion mit Gittern stets auf eine beliebige Gitterbasis zu beziehen. Mit folgendem Satz kann man zeigen, daß gewisse Konstruktionen von der Wahl der Gitterbasis unabhängig sind. Sei $[b_1, b_2, \ldots, b_n]$ die Matrix mit Spaltenvektoren $b_1, b_2, \ldots, b_n$. $\mathrm{GL}_n(\mathbb{Z})$ ist die Gruppe der ganzahligen $n \times n$-Matrizen mit Determinante ± 1.

Satz 2.2.5
Die Vektoren $\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n$ bilden genau dann eine Basis des Gitter $L(b_1, b_2, \ldots, b_n)$, wenn es eine Matrix $T \in \mathrm{GL}_n(\mathbb{Z})$ gibt mit:

$$\left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] = [b_1, b_2, \ldots, b_n] \cdot T$$

Nach Satz 2.2.5 bilden die Basen eines Gitters vom Rang n eine Bahn zur Gruppe $\mathrm{GL}_n(\mathbb{Z})$.

Beweis. Wir zeigen beide Richtungen:

„⇒“ Wegen $\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n \in L(b_1, b_2, \ldots, b_n)$ gibt es ein $T \in M_{n,n}(\mathbb{Z})$ mit:

$$\left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] = [b_1, b_2, \ldots, b_n] \cdot T$$

Weil $\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n$ linear unabhängig sind, gilt $\det T \neq 0$. Es folgt:

$$\left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] \cdot T^{-1} = [b_1, b_2, \ldots, b_n]$$

Wegen $b_i \in L\left(\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right)$ für $i = 1, 2, \ldots, n$ ist T^{-1} ganzzahlig. Da $\det T \cdot \det T^{-1} = 1$ und $\det T$, $\det T^{-1}$ ganzzahlig sind, folgt $|\det T| = 1$.

„⇐“ Angenommen, es gilt für ein $T \in \mathrm{GL}_n(\mathbb{Z})$:

$$\left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] = [b_1, b_2, \ldots, b_n] \cdot T$$

Es folgt, daß $\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n \in L(b_1, b_2, \ldots, b_n)$. Hierzu symmetrisch folgt aus

$$\left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] \cdot T^{-1} = [b_1, b_2, \ldots, b_n],$$

daß $b_1, b_2, \ldots, b_n \in L\left(\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right)$. Also gilt:

$$L\left(\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right) = L(b_1, b_2, \ldots, b_n)$$

■

2.2.2 Definition der Hermite-Normalform

Wir definieren im Vorgriff die Hermite-Normalform einer Matrix und fassen wichtige Eigenschaften zusammen. Wir werden uns in Kapitel 11.2 ab Seite 124 intensiv mit der Hermite-Normalform beschäftigen und einen Algorithmus zur Bestimmung der Hermite-Normalform zu einer gegebenen Matrix formulieren.

Definition 2.2.6 (Hermite-Normalform)
Eine Matrix $[a_{ij}] \in M_{m,n}(\mathbb{R})$ mit $n \leq m$ ist in Hermite-Normalform (kurz HNF), falls:

a) $a_{ij} = 0$ *für* $j > i$, *d.h. A ist eine untere Dreiecksmatrix.*

b) $a_{ii} > 0$ *für* $i = 1, 2, \ldots, m$.

c) $0 \leq a_{ij} < a_{ii}$ *für* $j < i$.

Es ist willkürlich zu verlangen, daß die Hermite-Normalform eine untere und keine obere Dreiecksmatrix ist. Man könnte ebenso definieren, daß eine Hermite-Normalform eine obere Dreiecksmatrix sei (vergleiche Korollar 2.2.9). Alternative Definitionen unterscheiden sich in Punkt c). In [DKT87] fordern die Autoren:

$$a_{ij} \leq 0 \text{ und } |a_{ij}| < a_{ii} \quad \text{für } j < i$$

In [PaSchn87] fordern die Autoren A. Paz und C.P. Schnorr, daß die Elemente links der Diagonalen jeweils die betragsmäßig kleinste sind:

$$|a_{ij}| < \tfrac{1}{2}\,|a_{ii}| \quad \text{für } j < i$$

Wir können die verschiedenen Hermite-Normalformen durch Addition von ganzzahligen Vielfachen einer Spalte zu einer anderen Spalte überführen. Die verschiedene Hermite-Normalformen sind damit für unsere Anwendungen äquivalent, da nach Satz 2.2.5 die Spaltenvektoren das gleiche Gitter erzeugen.

Den folgenden Satz hat C. Hermite [Hermite1850] zunächst nur für quadratische Matrizen bewiesen:

Satz 2.2.7 (Hermite 1850)
Zu jeder Matrix $A \in M_{m,n}(\mathbb{Q})$ mit $\mathrm{Rang}\,(A) = m \leq n$ *gibt es eine Matrix $T \in \mathrm{GL}_n(\mathbb{Z})$, so daß AT in Hermite-Normalform ist. Die Hermite-Normalform AT ist eindeutig bestimmt.*

Beweis. Sei a der Hauptnenner der Einträge der Matrix A. Dann ist $aA \in M_{m,n}(\mathbb{Z})$ und $\frac{1}{a}(aA)\,T$ die zugehörige Hermite-Normalform. Wir können uns daher auf den Fall $A \in M_{m,n}(\mathbb{Z})$ beschränken.

Die Existenz der Hermite-Normalform werden wir in Kapitel 11.2 durch Algorithmus 11.2.1 (Seite 125) beweisen. Zu zeigen ist noch die Eindeutigkeit. Angenommen, es gebe zwei Hermite-Normalformen $B, C \in M_{m,n}(\mathbb{Z})$ zu A.

$$B = \begin{bmatrix} b_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ b_{21} & b_{22} & \ddots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & 0 & 0 & \cdots & 0 \\ b_{m1} & b_{m2} & \cdots & b_{mm} & 0 & \cdots & 0 \end{bmatrix} \qquad C = \begin{bmatrix} c_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ c_{21} & c_{22} & \ddots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & 0 & 0 & \cdots & 0 \\ c_{m1} & c_{m2} & \cdots & c_{mm} & 0 & \cdots & 0 \end{bmatrix}$$

Seien $B_1, B_2, \ldots, B_n$ und $C_1, C_2, \ldots, C_n$ die Spaltenvektoren von B und C. Da beide Hermite-Normalformen aus A durch Multiplikation mit unimodularer Matrix hervorgehen, gilt nach Satz 2.2.5, daß $L(B) = L(C)$ ist. Speziell gilt $B_m \in L(C)$ und $C_m \in L(B)$. Da die Diagonalelemente alle ungleich 0 sind, muß B_m ein ganzzahliges Vielfaches von C_m und umgekehrt sein, also $c_{mm} = b_{mm}$.

Sei j der maximale Index mit $B_j \neq C_j$. Weil $c_{mm} = b_{mm}$ ist, gilt $j{+}1 \leq m$. Wegen $L(B) = L(C)$ existieren ganzzahlige Koeffizienten $t_j, t_{j+1}, \ldots, t_m \in \mathbb{Z}$, so daß für jedes i mit $j \leq i \leq m$ gilt (beachte: $c_{ik} = 0$ für $k > i$):

$$b_{ij} = \sum_{k=j}^{i} t_k c_{ik} \tag{2.2}$$

Wie im Fall des m-ten Spaltenvektors erhalten wir $b_{jj} = c_{jj}$, also $t_j = 1$. Für $i = j + 1$ gilt

$$b_{j+1,j} = t_j \cdot c_{j+1,j} + t_{j+1} \cdot c_{j+1,j+1} = c_{j+1,j} + t_{j+1} \cdot c_{j+1,j+1} \tag{2.3}$$

und wir erhalten:

$$c_{j+1,j+1} \mid t_{j+1} \cdot c_{j+1,j+1} = (b_{j+1,j} - c_{j+1,j}) \tag{2.4}$$

Da B eine Hermite-Normalform ist, gilt nach Wahl von j, dem maximalen Index mit $B_j \neq C_j$:

$$0 \leq b_{j+1,j} < b_{j+1,j+1} = c_{j+1,j+1}$$

Weil auch C eine Hermite-Normalform ist, gilt $0 \leq c_{j+1,j} < c_{j+1,j+1}$, und wir erhalten:

$$|b_{j+1,j} - c_{j+1,j}| < c_{j+1,j+1}$$

Aus (2.4) folgt $b_{j+1,j} = c_{j+1,j}$ und wegen (2.3) ist $t_{j+1} = 0$. Induktiv zeigt man, daß ebenfalls

$$t_{j+2} = t_{j+3} = \cdots = t_m = 0$$

und erhält aus (2.2) den Widerspruch $B_j = 0$. ∎

Wir wollen die Hermite-Normalform für einen einfachen Fall berechnen:

Beispiel 2.2.8 (Hermite-Normalform)
Wir betrachten den Fall, daß die Matrix A nur aus einer Zeile besteht

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

und mindestens einer der Einträge ungleich 0 ist. Man überlege sich, daß wir die Operationen des Euklidischen Algorithmus' auf den Werten $a_1, a_2, \ldots, a_n$ durch Multiplikation mit unimodularen

Matrizen von rechts bewerkstelligen können. Das Ergebnis $g := \mathrm{ggT}(a_1, a_2, \ldots, a_n)$ speichern wir links, die übrigen Werte sind am Ende 0. Wir erhalten Hermite-Normalform zu A:

$$\mathrm{HNF}(A) = \begin{bmatrix} g & 0 & \cdots & 0 \end{bmatrix}$$

Beachte, daß die Spaltenvektoren von A und $\mathrm{HNF}(A)$ dasselbe Gitter, nämlich

$$\sum_{i=1}^{n} \mathbb{Z} a_i = \mathbb{Z} \cdot \mathrm{ggT}(a_1, a_2, \ldots, a_n)$$

erzeugen. Wir werden in Satz 2.2.5 sehen, daß dies kein Zufall ist. Der Euklidische Algorithmus bildet die Basis des Verfahrens in Kapitel 11.2 zur Berechnung der Hermite-Normalform. ◇

Wir haben Satz 2.2.7 nur für rationale und insbesondere ganzzahlige Matrizen formuliert. Für reelle Matrizen gilt der Satz im allgemeinen nicht, denn zu

$$A := \begin{bmatrix} 1 & \sqrt{2} \\ 3 & 4 \end{bmatrix} \in M_{2,2}(\mathbb{R})$$

gibt es keine Matrix $T \in \mathrm{GL}_2(\mathbb{Z})$, so daß AT eine Hermite-Normalform ist (Beweis durch Widerspruch).

Korollar 2.2.9
Zu jeder Matrix $A \in M_{m,n}(\mathbb{Q})$ mit $\mathrm{Rang}(A) = m \leq n$ gibt es eine Matrix $T \in \mathrm{GL}_n(\mathbb{Z})$, so daß AT obere Dreicksmatrix ist.

Beweis. Definiere Matrix $U := [u_{ij}] \in \mathrm{GL}_n(\mathbb{Z})$ durch $u_{ij} := \delta_{i,n+1-j}$. U entsteht aus der Einheitsmatrix durch Umdrehen der Spaltenreihenfolge. Es gilt offenbar $U = U^{-1}$. Die Multiplikation einer Matrix mit U

- von links bewirkt das Umdrehen der Zeilenreihenfolge und
- von rechts das Umdrehen der Spaltenreihenfolge.

Wir bestimmen zu UAU nach Satz 2.2.7 die Hermite-Normalform $(UAU) \cdot S$ mit $S \in \mathrm{GL}_n(\mathbb{Z})$. $(UAU) \cdot S$ ist eine untere Dreiecksmatrix. Definiere die Matrix $T := USU \in \mathrm{GL}_n(\mathbb{Z})$. Wir erhalten durch Umdrehen der Zeilen- und Spaltenreihenfolge der Matrix $(UAU) \cdot S$ eine obere Dreiecksmatrix:

$$U \cdot (UAUS) \cdot U = AUSU = AT$$

■

2.2.3 Determinante und Grundmasche

Wir definieren die Determinante eines Gitters:

Definition 2.2.10 (Determinante)
Die Determinante $\det L$ des Gitters $L = (b_1, b_2, \ldots, b_n) \subseteq \mathbb{R}^m$ ist definiert als:

$$\det L = \left(\det \left[\langle b_i, b_j \rangle \right]_{1 \leq i,j \leq n} \right)^{\frac{1}{2}}$$

Für das Skalarprodukt $\langle u, v \rangle = u^{\mathsf{T}} S v$ gilt: $\det L = \det(B \cdot S \cdot B^{\mathsf{T}})^{\frac{1}{2}}$, wobei $B := [b_1, b_2, \ldots, b_n]$.

Satz 2.2.11
Die Gitterdeterminante ist unabhängig von der Wahl der Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$.

Beweis. Seien $B, \overline{B}$ Basismatrizen des Gitters und $T \in \mathrm{GL}_n(\mathbb{Z})$ mit $\overline{B} = B \cdot T$ (der Beweis gilt allgemein für $T \in \mathrm{GL}_n(\mathbb{R})$). Sei $S \in M_{m,m}(\mathbb{R})$ die symmetrische Matrix mit $\langle u, v \rangle = u^{\mathsf{T}} S v$. Wegen $\det T = 1$ gilt:

$$\begin{aligned}
\det L &= \left(\det \left[\langle b_i, b_j \rangle\right]_{1 \le i,j \le n}\right)^{\frac{1}{2}} \\
&= \det \left(B^{\mathsf{T}} \cdot S \cdot B\right)^{\frac{1}{2}} \\
&= \det \left(T^{\mathsf{T}} \cdot B^{\mathsf{T}} \cdot S \cdot B \cdot T\right)^{\frac{1}{2}} \\
&= \det \left((BT)^{\mathsf{T}} \cdot S \cdot (BT)\right)^{\frac{1}{2}}
\end{aligned}$$

Mit $\overline{B} = B \cdot T$ folgt:

$$\begin{aligned}
\det L &= \det \left(\overline{B}^{\mathsf{T}} \cdot S \cdot \overline{B} \right)^{\frac{1}{2}} \\
&= \left(\det \left[\left\langle \overline{b}_i, \overline{b}_j \right\rangle\right]_{1 \le i,j \le n}\right)^{\frac{1}{2}} \\
&= \det \overline{L}
\end{aligned}$$

■

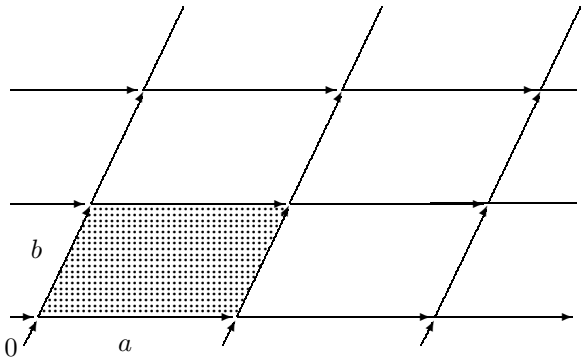

Abbildung 2.2.1: Grundmasche des Gitters $L(a, b)$

Für das Parallelepiped, das von den Basisvektoren des Gitters aufgespannt wird, führen wir den Begriff der Grundmasche ein. Abbildung 2.2.1 zeigt die Grundmasche des Gitters, das von den beiden Vektoren a und b im $\mathbb{R}^2$ aufgespannt wird.

Definition 2.2.12 (Grundmasche)
Sei $b_1, b_2, \ldots, b_n$ eine Basis des Gitters L. Das folgende Parallelepiped heißt Grundmasche zur Basis $b_1, b_2, \ldots, b_n$:

$$\left\{ \sum_{i=1}^{n} t_i b_i \;\middle|\; 0 \le t_1, t_2, \ldots, t_n < 1 \right\}$$

Wir beschäftigen uns meist mit dem Fall, daß die Dimension des Gitters $L = \subseteq \mathbb{R}^m$ gleich m ist.

Definition 2.2.13 (Vollständiges Gitter)
Ein Gitter $L \subseteq \mathbb{R}^m$ heißt vollständig, falls $\mathrm{Rang}(L) = m$.

Das folgende Lemma liefert für das Standard-Skalarprodukt eine anschauliche Interpretation der Gitterdeterminante:

Lemma 2.2.14
Für jedes Gitter $L = L(b_1, b_2, \ldots, b_n) \subseteq \mathbb{R}^m$ und das Standard-Skalarprodukt gilt (Beachte: Volumina implizieren stets das Standard-Skalarprodukt):

$$\det L = \mathrm{vol}_n \left(\left\{ \sum_{i=1}^{n} t_i b_i \;\middle|\; 0 \leq t_1, t_2, \ldots, t_n < 1 \right\} \right)$$

In Worten: Das n-dimensionale Volumen der Grundmasche ist gleich der Gitterdeterminante. Für vollständige Gitter ist die Gitterdeterminante gleich der Determinante der Basismatrix.

Beweis. Wir behandeln zunächst den Fall vollständiger Gitter. In diesem Fall gilt für die Basismatrix $B := [b_1, b_2, \ldots, b_n]$, daß:

$$\det L = |\det B| = \left(\det B^{\mathsf{T}} B\right)^{\frac{1}{2}} = \left(\det \left[\langle b_i, b_j \rangle\right]_{1 \leq i,j \leq n}\right)^{\frac{1}{2}}$$

Im allgemeinen Fall, also $\mathrm{Rang}(L) = n \leq m$, gibt es, wie wir später zeigen, eine isometrische Abbildung $T : \mathrm{span}(L) \to \mathbb{R}^m$, d.h. für alle $u, v \in \mathrm{span}(L)$ gilt, daß $\langle u, v \rangle = \langle T(u), T(v) \rangle$. Wir wenden das Lemma auf das vollständige Gitter $T(L)$ an und benutzen, daß T Volumina und Skalarprodukte erhält:

$$\det L = \det T(L) = \left(\det \left[\langle T(b_i), T(b_j) \rangle\right]_{1 \leq i,j \leq n}\right)^{\frac{1}{2}} = \left(\det \left[\langle b_i, b_j \rangle\right]_{1 \leq i,j \leq n}\right)^{\frac{1}{2}}$$

■

Wie im Beweis von Lemma 2.2.14 können wir uns für geometrische Überlegungen ohne Beschränkung der Allgemeinheit auf vollständige Gitter beschränken. Dabei betreffen geometrische Überlegungen solche Größen, die bei isometrischen Abbildungen erhalten bleiben. Geometrische Invarianten sind zum Beispiel Volumeninhalte, Determinanten, Skalarprodukte und Längen von Vektoren. Der Grund, daß vollständige Gitter für geometrische Betrachtungen ausreichen, liegt darin, daß es zu jedem Gitter L vom Rang n eine isometrische Abbildung gibt, die $\mathrm{span}(L)$ auf $\mathbb{R}^n$ abbildet. Diese isometrische Abbildung erhält im allgemeinen nicht die Ganzzahligkeit von Vektoren. Kombinatorische und algorithmische Untersuchungen darf man daher nicht auf den Fall vollständiger Gitter beschränken.

2.2.4 Untergitter

Wir definieren Untergitter als eine Teilmenge der Gitterpunkte, die ein Gitter mit gleichem Rang bilden:

Definition 2.2.15 (Untergitter)
Seinen L_1, L_2 Gitter vom gleichen Rang mit $L_1 \subseteq L_2$, dann heißt L_1 Untergitter von L_2.

Beispiel 2.2.16 (Untergitter)
Wir betrachen ein Untergitter zum Gitter $\mathbb{Z}^n$ (als Basismatrix wählen wir die 2×2-Einheitsmatrix E_2). Sei

$$A := \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$L(A) \subseteq \mathbb{Z}^2$ ist ein Untergitter von $\mathbb{Z}^2$, denn $\operatorname{Rang}(L(A)) = 2$ (vergleiche Abbildung 2.2.2). Es gibt eine Matrix $T \in M_{2,2}(\mathbb{Z})$ mit $A = E_2 \cdot T$:

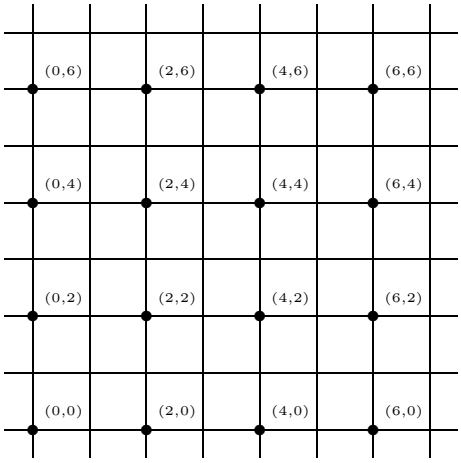

Abbildung 2.2.2: Untergitter $L(A)$ von $\mathbb{Z}^2$

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{=:T}$$

Wir werden in Lemma 2.2.17 sehen, daß stets $\det L_1 = \det L_2 \cdot |\det T|$ gilt. Die Faktorgruppe $\mathbb{Z}^2/L(A)$ besteht aus den vier Äquivalenzklassen

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} + L(A), \quad \begin{bmatrix} 1 \\ 0 \end{bmatrix} + L(A), \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} + L(A), \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} + L(A)$$

Also $[\mathbb{Z}^n : L(A)] = 4$. In Lemma 2.2.17 zeigen wir allgemein, daß der Index gleich $|\det T|$ ist. Wir repräsentieren die Äquivalenzklassen durch den jeweiligen Vertreter, der in der Grundmasche liegt. ◇

Lemma 2.2.17
Sei $L_2 = L(b_1, b_2, \ldots, b_n)$ ein Gitter und $L_1 = L(a_1, a_2, \ldots, a_n)$ ein Untergitter von L_2. Sei $T \in M_{n,n}(\mathbb{Z})$ mit $A = B \cdot T$ für $A := [a_1, a_2, \ldots, a_n]$ und $B := [b_1, b_2, \ldots, b_n]$. Dann ist:

$$\det L_1 = \det L_2 \cdot |\det T|$$

Beweis. Sei $S \in M_{m,m}(\mathbb{R})$ die symmetrische Matrix mit $\langle u, v\rangle = u^{\mathsf{T}} S v$. Es gilt:

$$\begin{aligned}
\det L_1 &= \det \left(A^{\mathsf{T}} \cdot S \cdot A\right)^{\frac{1}{2}} \\
&= \det \left((BT)^{\mathsf{T}} \cdot S \cdot (BT)\right)^{\frac{1}{2}} \\
&= \det \left(T^{\mathsf{T}} B^{\mathsf{T}} \cdot S \cdot BT\right)^{\frac{1}{2}} \\
&= \det \left(T^{\mathsf{T}} T\right)^{\frac{1}{2}} \cdot \left(\det B^{\mathsf{T}} \cdot S \cdot B\right)^{\frac{1}{2}} \\
&= |\det T| \cdot \left(\det B^{\mathsf{T}} \cdot S \cdot B\right)^{\frac{1}{2}} \\
&= |\det T| \cdot \det L_2
\end{aligned}$$

■

Definition 2.2.18 (Index des Untergitters)
Die ganze Zahl $|\det T| = \frac{\det L_1}{\det L_2}$ *aus Lemma 2.2.17 heißt der Index des Untergitters* L_1 *in* L_2.

Der Index hängt nicht von der Wahl der Gitterbasen ab. Insbesondere ist L_1 eine Untergruppe der additiven Gruppe L_2 vom Index $|\det T|$.

Satz 2.2.19
Sei L_1 *ein Untergitter von* L_2.

a) Es gibt zu jeder Basis $a_1, a_2, \ldots, a_n$ *von* L_1 *eine Basis* $b_1, b_2, \ldots, b_n$ *von* L_2 *mit*

$$[a_1, a_2, \ldots, a_n] = [b_1, b_2, \ldots, b_n] \cdot T \tag{2.5}$$

für eine obere Dreiecksmatrix $T \in M_{n,n}(\mathbb{Z})$.

b) Umgekehrt gibt es zu jeder Basis $b_1, b_2, \ldots, b_n$ *von* L_2 *eine Basis* $a_1, a_2, \ldots, a_n$ *von* L_1 *mit Eigenschaft* (2.5).

Bemerkung 2.2.20
Die Eigenschaft (2.5) *mit einer oberen Dreiecksmatrix* T *hat zur Folge, daß:*

$$\operatorname{span}(a_1, a_2, \ldots, a_i) = \operatorname{span}(b_1, b_2, \ldots, b_i) \qquad \text{für } i = 1, 2, \ldots, n$$

Beweis (zu Satz 2.2.19). Wir zeigen beide Aussagen:

a) Sei $\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n$ eine beliebige Basis von L_2. Es gibt eine Matrix $S \in M_{n,n}(\mathbb{Z})$ mit $\det S \neq 0$ und:

$$[a_1, a_2, \ldots, a_n] = \left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] \cdot S \tag{2.6}$$

Anwendung des HNF-Satzes 2.2.7 auf S^{T} liefert ein $U \in \mathrm{GL}_n(\mathbb{Z})$, so daß $S^{\mathsf{T}} U^{\mathsf{T}} = (US)^{\mathsf{T}}$ eine untere Dreiecksmatrix ist. Definiere $T := US$ (eine untere Dreiecksmatrix) und Basismatrix zum Gitter L_2:

$$[b_1, b_2, \ldots, b_n] := \left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] \cdot U^{-1}$$

Aus (2.6) erhalten wir:

$$[a_1, a_2, \ldots, a_n] = \left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n\right] \cdot S = [b_1, b_2, \ldots, b_n] \cdot U \cdot S$$

b) Sei $\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_n$ beliebige Basis von L_1. Es gibt eine Matrix $S \in M_{n,n}(\mathbb{Z})$ mit $\det S \neq 0$ und:

$$[\,\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_n] = [b_1, b_2, \ldots, b_n] \cdot S$$

Nach Korollar 2.2.9 gibt es $U \in \mathrm{GL}_n(\mathbb{Z})$, so daß SU eine obere Dreiecksmatrix ist. Die Behauptung gilt für die Basis

$$[a_1, a_2, \ldots, a_n] := [\,\overline{a}_1, \overline{a}_2, \ldots, \overline{a}_n] \cdot U$$

und $T := SU$.

■

Sei $L_1 = L(a_1, a_2, \ldots, a_n)$ ein Untergitter von $L_2 = L(b_1, b_2, \ldots, b_n)$, so daß die beiden Basen $a_1, a_2, \ldots, a_n$ und $b_1, b_2, \ldots, b_n$ Eigenschaft (2.5) mit $T = [t_{ij}]$ erfüllen. Dann ist L_2/L_1 eine endliche Gruppe, und für den Index $[L_2 : L_1]$ von L_1 in L_2 gilt:

$$[L_2 : L_1] = \frac{\det L_1}{\det L_2} = \det T = \prod_{i=1}^{n} |t_{ii}| \tag{2.7}$$

Insbesondere ist genau dann $L_1 = L_2$, wenn $|\det T| = 1$. Diese Beziehung kann man zum Nachweis der Basiseigenschaft von $a_1, a_2, \ldots, a_n$ nutzen. Sei L_1 das von $a_1, a_2, \ldots, a_n$ erzeugte Gitter. Angenommen, wir kennen $\frac{\det L_1}{\det L_2}$ und eine Basis $b_1, b_2, \ldots, b_n$ von L_2. Ferner erfülle das System von Vektoren $a_1, a_2, \ldots, a_n$ die Beziehung (2.5). Dann gilt:

$$L_2 = L(a_1, a_2, \ldots, a_n) \quad \Longleftrightarrow \quad \prod_{i=1}^{n} |t_{ii}| = \frac{\det L_1}{\det L_2} = 1$$

Aus (2.7) folgt:

Korollar 2.2.21
Sei $a = (a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n \setminus \{0\}$ und $b \in \mathbb{N}$. Für das ganzzahlige Gitter

$$L_{a,b} = \{x \in \mathbb{Z}^n \mid \langle x, a\rangle \equiv 0 \pmod{b}\}$$

(wobei $\langle \cdot, \cdot \rangle$ das Standard-Skalarprodukt ist) gilt:

$$\det L_{a,b} = \frac{b}{\mathrm{ggT}\,(a_1, a_2, \ldots, a_n, b)}$$

Beweis. Wir zeigen zuerst, daß $L_{a,b}$ ein Untergitter von $\mathbb{Z}^n$ ist, d.h. das Gitter $L_{a,b} \subseteq \mathbb{Z}^n$ hat vollen Rang. Sei $v_n = (1, \ldots, 1) \in \mathbb{Z}^n$. Man wähle $n-1$ ganzzahlige Vektoren $v_1, v_2, \ldots, v_{n-1}$, so daß $bv_1, bv_2, \ldots, bv_n \in \mathbb{Z}^n$ linear unabhängig sind (zum Beispiel die ersten $n-1$ Einheitsvektoren). Wegen

$$\langle bv_i, a\rangle \equiv b \cdot \langle v_i, a\rangle \equiv 0 \pmod{b} \qquad \text{für } i = 1, 2, \ldots, n$$

liegen $bv_1, bv_2, \ldots, bv_n \in \mathbb{Z}^n$ im Gitter $L_{a,b}$ und bilden eine Basis. Also ist $L_{a,b}$ ein Untergitter von $\mathbb{Z}^n$. Wegen $\det \mathbb{Z}^n = 1$ folgt aus (2.7):

$$\det L_{a,b} = [\mathbb{Z}^n : L_{a,b}] \tag{2.8}$$

Die Faktorgruppe $\mathbb{Z}^n / L_{a,b}$ besteht aus maximal b Restklassen, nämlich $R_0, R_1, \ldots, R_{b-1}$ mit:

$$R_i := \{x \in \mathbb{Z}^n \mid \langle x, a\rangle \equiv i \pmod{b}\}$$

Wir zeigen, daß die Faktorgruppe genau aus diesen b Restklassen besteht. Wir betrachten zunächst den Fall:

$$\operatorname{ggT}(a_1, a_2, \ldots, a_n, b) = 1 \tag{2.9}$$

Dann existieren ganzzahlige Koeffizienten $t_1, t_2, \ldots, t_{n+1} \in \mathbb{Z}$ mit:

$$\sum_{j=1}^{n} t_j \cdot a_j + t_{n+1} \cdot b = 1$$

Für den Vektor $t := (t_1, t_2, \ldots, t_n) \in \mathbb{Z}^n$ gilt:

$$\langle t, a \rangle \equiv 1 - t_{n+1} \cdot b \equiv 1 \pmod{b}$$

Für die Vektoren $c_i := i \cdot t \in \mathbb{Z}^n$ mit $i = 0, 1, \ldots, b-1$ erhalten wir:

$$\langle c_i, a \rangle \equiv i \cdot \langle t, a \rangle \equiv i \pmod{b}$$

Die Restklassen $R_0, R_1, \ldots, R_{b-1}$ sind deshalb nicht-leer. Wegen der Vorausetzung (2.9) erhalten wir aus (2.8):

$$\det L_{a,b} = [\mathbb{Z}^n : L_{a,b}] = \frac{b}{1} = \frac{b}{\operatorname{ggT}(a_1, a_2, \ldots, a_n, b)}$$

Betrachten wir den anderen Fall:

$$d := \operatorname{ggT}(a_1, a_2, \ldots, a_n, b) > 1$$

Da für alle $x \in \mathbb{Z}^n$ gilt

$$\begin{aligned} \langle x, a \rangle \equiv 0 \pmod{b} \quad &\Longleftrightarrow \quad b \mid \langle x, a \rangle \\ &\Longleftrightarrow \quad \tfrac{b}{d} \mid \langle x, \tfrac{a}{d} \rangle \\ &\Longleftrightarrow \quad \langle x, \tfrac{a}{d} \rangle \equiv 0 \pmod{\tfrac{b}{d}}, \end{aligned}$$

ist $L_{a,b} = L_{\frac{a}{d}, \frac{b}{d}}$. Wegen $\operatorname{ggT}\left(\frac{a_1}{d}, \frac{a_2}{d}, \ldots, \frac{a_n}{d}, \frac{b}{d}\right) = 1$ erhalten wir:

$$\det L_{a,b} = \det L_{\frac{a}{d}, \frac{b}{d}} = \frac{b}{d} = \frac{b}{\operatorname{ggT}(a_1, a_2, \ldots, a_n, b)}$$

■

2.2.5 Primitive Systeme

Wir charakterisieren in Satz 2.2.23 Systeme von Gittervektoren, die man zu einer Basis des Gitters ergänzen kann. Dieses Kriterium ist für die Konstruktion spezieller Gitterbasen nützlich.

Definition 2.2.22 (Primitives System)
Sei $L \subseteq \mathbb{R}^n$ ein Gitter. Die Vektoren $b_1, b_2, \ldots, b_k \in L$ bilden ein primitives System bezüglich L, falls:

1. *$b_1, b_2, \ldots, b_k$ sind linear unabhängig.*
2. *$\operatorname{span}(b_1, b_2, \ldots, b_k) \cap L = L(b_1, b_2, \ldots, b_k)$*

Weil die Inklusion $\text{span}(b_1, b_2, \ldots, b_k) \cap L \supseteq L(b_1, b_2, \ldots, b_k)$ für beliebige Gittervektoren gilt, kommt es nur auf $\text{span}(b_1, b_2, \ldots, b_k) \cap L \subseteq L(b_1, b_2, \ldots, b_k)$ an. Ein einzelner Vektor $b \in L$ bildet ein primitives System, wenn $\frac{1}{k} b \notin L$ für alle $k \in \mathbb{Z}$ mit $|k| \geq 2$, also der größte gemeinsame Teiler der Vektoreinträge gleich 1 ist.

Satz 2.2.23
Sei $L \subseteq \mathbb{R}^n$ ein Gitter mit $\text{Rang}(L) = n$ und $b_1, b_2, \ldots, b_k \in L$. Genau dann können die Vektoren $b_1, b_2, \ldots, b_k$ zu einer Gitterbasis ergänzt werden, wenn sie ein primitives System bezüglich L bilden.

Beweis. Wir zeigen beide Richtungen:

„⇒“ Sei $b_1, b_2, \ldots, b_k, b_{k+1}, \ldots, b_n$ eine Basis des Gitters L. Die Vektoren $b_1, b_2, \ldots, b_k$ sind linear unabhängig. Jeder Vektor $b \in L$ hat eine eindeutige Darstellung $b = \sum_{i=1}^{n} t_i b_i$ mit $t_1, t_2, \ldots, t_n \in \mathbb{Z}$. Es gilt:

$$b \in \text{span}(b_1, b_2, \ldots, b_k) \quad \Longleftrightarrow \quad t_{k+1} = t_{k+2} = \cdots = t_n = 0$$

Also $\text{span}(b_1, b_2, \ldots, b_k) \cap L \subseteq L(b_1, b_2, \ldots, b_k)$.

„⇐“ Sei $b_1, b_2, \ldots, b_k$ ein primitives System und $\pi : \text{span}(L) \to \text{span}(b_1, b_2, \ldots, b_k)^{\perp}$ die orthogonale Projektion. Aus dem Beweis zu Satz 2.2.3 folgt, daß $\pi(L)$ ein Gitter der Dimension $n-k$ ist. Das Gitter $\pi(L)$ habe die Basis $\pi(b_{k+1}), \pi(b_{k+2}), \ldots, \pi(b_n)$ mit $b_{k+1}, b_{k+2}, \ldots, b_n \in L$.

Wir zeigen, daß $L = L(b_1, b_2, \ldots, b_n)$ ist. Sei $b \in L$. Wegen

$$\pi(b) \in L\big(\pi(b_{k+1}), \pi(b_{k+2}), \ldots, \pi(b_n)\big)$$

gibt es ein $\overline{b} \in L(b_{k+1}, b_{k+2}, \ldots, b_n)$ mit $\pi(\overline{b}) = \pi(b)$. Seien

$$\pi(b) = \sum_{i=k+1}^{n} t_i \pi(b_i) \qquad \text{und} \qquad \overline{b} = \sum_{i=k+1}^{n} t_i b_i$$

Also ist $b - \overline{b} \in \text{span}(b_1, b_2, \ldots, b_k)$. Weil $b_1, b_2, \ldots, b_k$ nach Voraussetzung ein primitives System bildet, gilt $b - \overline{b} \in L(b_1, b_2, \ldots, b_k)$. Somit ist $b \in L(b_1, b_2, \ldots, b_n)$.

■

Der folgende Begriff einer Minkowski-reduzierten Basis ist nicht algorithmisch motiviert und soll nur aus Gründen der Vollständigkeit erwähnt werden (für einen Reduktionsalgorithmus verweisen wir auf B. Helfrichs Arbeit [Helfrich85]):

Definition 2.2.24 (Minkowski-reduziert)
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine geordnete Basis zum Gitter L, d.h. die Reihenfolge der Vektoren sei fest. Die geordnete Basis heißt Minkowski-reduziert, wenn für $i = 1, 2, \ldots, n$ gilt:

$$\|b_i\| = \min \left\{ \|b\| \;\middle|\; \begin{array}{l} b \in L \text{ und } (b_1, b_2, \ldots, b_{i-1}, b) \\ \text{ist primitives System bzgl. } L \end{array} \right\}$$

2.2.6 Orthogonalsysteme

Das Ziel der Reduktion von Gitterbasen ist es, eine gegebene Gitterbasis in eine solche zu transformieren, die aus möglichst kurzen Vektoren besteht bzw. deren Vektoren möglichst orthogonal zueinander sind.

Definition 2.2.25 (Orthogonalsystem, orthogonale Projektion π_i)
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine geordnete Gitterbasis. Es bezeichne

$$\pi_i : \mathbb{R}^m \to \operatorname{span}(b_1, b_2, \ldots, b_{i-1})^\perp$$

die orthogonale Projektion, d.h. es gilt für alle $b \in \mathbb{R}^m$:

$$\begin{aligned} \pi_i(b) &\in \operatorname{span}(b_1, b_2, \ldots, b_{i-1})^\perp \\ b - \pi_i(b) &\in \operatorname{span}(b_1, b_2, \ldots, b_{i-1}) \end{aligned}$$

Es bezeichne $\widehat{b}_i := \pi_i(b_i)$. Man berechnet $\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n$ durch das Gram-Schmidt-Verfahren:

$$\widehat{b}_1 := b_1 \tag{2.10}$$

$$\widehat{b}_i := \pi_i(b_i) = b_i - \sum_{j=1}^{i-1} \mu_{i,j} \widehat{b}_j \qquad \text{für } i = 2, 3, \ldots, n \tag{2.11}$$

Dabei sind die Gram-Schmidt-Koeffizienten $\mu_{i,j}$ erklärt durch:

$$\mu_{i,j} := \frac{\left\langle b_i, \widehat{b}_j \right\rangle}{\left\langle \widehat{b}_j, \widehat{b}_j \right\rangle} = \frac{\left\langle b_i, \widehat{b}_j \right\rangle}{\|\widehat{b}_j\|^2}$$

Insbesondere gilt $\mu_{i,i} = 1$ und für $j > i$ ist $\mu_{i,j} = 0$. Es ist $\sum_{j=1}^{k} \mu_{i,j} \widehat{b}_j$ die orthogonalen Projektionen von b_i in $\operatorname{span}(b_1, b_2, \ldots, b_k)$ und $\sum_{j=k+1}^{i} \mu_{i,j} \widehat{b}_j$ die orthogonalen Projektionen von b_i in $\operatorname{span}(b_1, b_2, \ldots, b_k)^\perp$. Nach Definition (2.10) und (2.11) können wir den Vektor b_i mit Hilfe des Orthogonalsystems schreiben als:

$$b_i = \widehat{b}_i + \sum_{j=1}^{i-1} \mu_{i,j} \widehat{b}_j \tag{2.12}$$

Diese Darstellung der Basisvektoren $b_1, b_2, \ldots, b_n$ lautet in Matrizenform:

$$[b_1, b_2, \ldots, b_n] = \left[\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n \right] \cdot [\mu_{i,j}]_{1 \le i,j \le n}^{\mathsf{T}}$$

Man erkennt leicht, daß die Abbildung $T : \operatorname{span}(L) \to \mathbb{R}^m$ mit

$$T(b_i) = \left[\mu_{i,1} \|\widehat{b}_1\|, \quad \mu_{i,2} \|\widehat{b}_2\|, \quad \ldots \quad , \mu_{i,m} \|\widehat{b}_n\| \right]^{\mathsf{T}}$$

eine Isometrie ist. Es gilt:

$$[T(b_1), T(b_2), \ldots, T(b_n)] = \begin{bmatrix} \|\widehat{b}_1\| & 0 & \cdots & 0 \\ 0 & \|\widehat{b}_2\| & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \|\widehat{b}_n\| \end{bmatrix} \cdot \begin{bmatrix} 1 & \mu_{2,1} & \mu_{3,1} & \cdots & \mu_{n,1} \\ 0 & 1 & \mu_{3,2} & & \mu_{n,2} \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mu_{n,n-1} \\ 0 & \ldots & \cdots & 0 & 1 \end{bmatrix}$$

Da $\det[\mu_{i,j}]_{1 \le i,j \le n}^{\mathsf{T}} = 1$, gilt nach Beweis zu Satz 2.2.11:

$$\det L(b_1, b_2, \ldots, b_n) = \prod_{i=1}^{n} \|\widehat{b}_i\| \tag{2.13}$$

Wir definieren den Orthogonalitätsdefekt als Maß, wie weit die Basisvektoren senkreicht aufeinander stehen:

Definition 2.2.26 (Orthogonalitätsdefekt)
Der Orthogonalitätsdefekt einer Basis $b_1, b_2, \ldots, b_n$ des Gitters L ist:

$$\text{OrthDefekt}(L) := \frac{\prod_{i=1}^{n} \|b_i\|}{\det L},$$

Der Orthogonalitätsdefekt ist wegen $\|b_i\| \geq \|\widehat{b}_i\|$ größer oder gleich 1. Der Orthogonalitätsdefekt ist genau dann 1, wenn die Basisvektoren senkrecht aufeinander stehen, also $b_i = \widehat{b}_i$ für $i = 1, 2, \ldots, n$ gilt (siehe (2.13)).

Definition 2.2.27 (Isometrisches Gitter)
Zwei Gitter $L, \overline{L} \subseteq \mathbb{R}^m$ heißen isometrisch, falls es eine isometrische Abbildung $T : \mathbb{R}^m \to \mathbb{R}^m$ mit $T(L) = \overline{L}$ gibt.

Seien L und $\overline{L}$ isometrische Gitter und T eine Isometrie mit $T(L) = \overline{L}$ sowie $b_1, b_2, \ldots, b_n$ eine Basis von L. Dann gilt für die Basis $\overline{b}_i := T(b_i)$, $i = 1, 2, \ldots, n$, daß:

$$\langle b_i, b_j \rangle = \langle \overline{b}_i, \overline{b}_j \rangle \qquad \text{für } 1 \leq i, j \leq n$$

Zwei solche Basen nennen wir *isometrisch*. Für das Skalarprodukt $\langle u, v \rangle = u^{\mathsf{T}} S v$ gilt: Die Isometrieklasse einer Gitterbasis $B = [b_1, b_2, \ldots, b_n]$ ist durch die Matrix $B^{\mathsf{T}} \cdot S \cdot B$ charakterisiert. Insbesondere sind zwei Gitter genau dann isometrisch, wenn sie isometrische Basen besitzen. Beispiel für den Fall des Standard-Skalarprodukts:

$$B := \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3/2} \end{bmatrix} \qquad \overline{B} := \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad B^{\mathsf{T}} \cdot B = \overline{B}^{\mathsf{T}} \cdot \overline{B} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

2.2.7 Quadratische Formen

In diesem Anschnitt sei das Skalarprodukt das Standard-Skalarprodukt. Einer Basismatrix $B = [b_1, b_2, \ldots, b_n]$ ordnen wir folgende *quadratische Form* QF_B in den reellen Variablen $x_1, x_2, \ldots, x_n$ zu:

$$\text{QF}_B(x_1, x_2, \ldots, x_n) := \sum_{1 \leq i,j \leq n} \langle b_i, b_j \rangle \, x_i x_j$$

Diese quadratische Form QF_B ist positiv definit, d.h. es gilt stets $\text{QF}_B(x_1, x_2, \ldots, x_n) \geq 0$ und speziell $\text{QF}_B(x_1, x_2, \ldots, x_n) = 0$ genau dann, wenn $x_1 = x_2 = \ldots x_n = 0$. QF_B ist positiv definit wegen:

$$\text{QF}_B(x_1, x_2, \ldots, x_n) = \left\| \sum_{i=1}^{n} x_i b_i \right\|^2$$

und $\sum_{i=1}^{n} x_i b_i$ ist genau dann der Nullvektor, falls $x_1 = x_2 = \ldots x_n = 0$. QF_B hat für ganzzahlige $x_1, x_2, \ldots, x_n$ die Längenquadrate der Gittervektoren in $L(b_1, b_2, \ldots, b_n)$ als Werte.

Umgekehrt gibt es zu jeder positiv definiten, symmetrischen, quadratischen Form

$$\text{QF} = \sum_{1 \leq i,j \leq n} q_{ij} x_i x_j$$

eine Gitterbasis $b_1, b_2, \ldots, b_n$, so daß $\langle b_i, b_j \rangle = q_{ij}$ für $1 \leq i, j \leq n$. Denn jede positiv definite, symmetrische Matrix $(q_{ij}) \in M_{n,n}(\mathbb{R})$ kann man schreiben als $(q_{ij}) = B^{\mathsf{T}}B$ mit $B \in M_{n,n}(\mathbb{R})$.

Zwei Gitterbasen $B, \overline{B}$ erzeugen genau dann dieselbe quadratische Form $\mathrm{QF}_B = \mathrm{QF}_{\overline{B}}$, wenn es eine isometrische Abbildung gibt, welche B in $\overline{B}$ transformiert. Die quadratischen, positiv definiten Formen entsprechen eindeutig den Isometrieklassen von Gitterbasen. Die Theorie der Gitterbasen und der positiv definiten Formen ist in diesem Sinne äquivalent. Die älteren Beiträge unter anderem von J.L. Lagrange, C.F. Gauß, C. Hermite, A. Korkine, G. Zolotareff und H. Minkowski sind in der Sprache der quadratischen Formen abgefaßt. Zwei quadratische Formen

$$\mathrm{QF}_Q = \sum_{1 \leq i,j \leq n} q_{ij} x_i x_j \qquad \text{und} \qquad \mathrm{QF}_{\overline{Q}} = \sum_{1 \leq i,j \leq n} \overline{q}_{ij} x_i x_j$$

heißen *kongruent*, wenn sie durch eine unimodulare Variablentransformation ineinander überführbar sind. Das heißt, es existiert eine Matrix $U \in \mathrm{GL}_n(\mathbb{Z})$ mit:

$$[q_{ij}]_{1 \leq i,j \leq n} = U^{\mathsf{T}} \left[\overline{q}_{ij}\right]_{1 \leq i,j \leq n} U$$

Zum Beispiel sind für zwei Basen B und $\overline{B}$ desselben Gitters die zugehörigen quadratischen Formen QF_B und $\mathrm{QF}_{\overline{B}}$ kongruent.

2.2.8 Duale und ganze Gitter

In diesem Abschnitt sei das Skalarprodukt das Standard-Skalarprodukt. Wir definieren das duale Gitter:

Definition 2.2.28 (Duales (polares, reziprokes) Gitter)
Sei L ein Gitter. Das duale (polare, reziproke) Gitter ist definiert durch:

$$L^* := \{x \in \mathrm{span}(L) \mid \forall b \in L : \langle x, b \rangle \in \mathbb{Z}\}$$

Es gilt:

Satz 2.2.29
Sei $L \subseteq \mathbb{R}^n$ ein volldimensionales Gitter mit der Basismatrix $B := [b_1, b_2, \ldots, b_n]$. Dann ist $\left(B^{-1}\right)^{\mathsf{T}}$ eine Basismatrix des dualen Gitters L^*. Speziell gilt $(L^*)^* = L$.

Beweis. Sei:

$$B^* := [b_1^*, b_2^*, \ldots, b_n^*\,] := \left(B^{-1}\right)^{\mathsf{T}}$$

Zu zeigen ist $L^* = L(B^*)$. Wir zeigen beide Inklusionen:

- $L(B^*) \subseteq L^*$: Es gilt für die Einheitmatrix E:

$$E = B^{-1} \cdot B = \left(\left(B^{-1}\right)^{\mathsf{T}}\right)^{\mathsf{T}} \cdot B = (B^*)^{\mathsf{T}} \cdot B = \begin{bmatrix} \langle b_1^*, b_1 \rangle & \langle b_1^*, b_2 \rangle & \cdots & \langle b_1^*, b_n \rangle \\ \langle b_2^*, b_1 \rangle & \langle b_2^*, b_2 \rangle & \cdots & \langle b_2^*, b_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle b_n^*, b_1 \rangle & \langle b_n^*, b_2 \rangle & \cdots & \langle b_n^*, b_n \rangle \end{bmatrix}$$

  Also $\langle b_i^*, b_j \rangle \in \{0, 1\}$ für $i, j = 1, 2, \ldots, n$. Für $x = \sum_{i=1}^n t_i b_i^* \in L(B^*)$ mit $t_1, t_2, \ldots, t_n \in \mathbb{Z}$ gilt:

$$\langle x, b_j \rangle = \sum_{i=1}^{n} t_i \cdot \langle b_i^*, b_j \rangle \in \mathbb{Z} \qquad \text{für } j = 1, 2, \ldots, n$$

  Da $b_i^* \in \mathbb{R}^n = \mathrm{span}(L)$ für $i = 1, 2, \ldots, n$, folgt für alle $x \in L(B^*)$, daß $x \in L^*$.

- $L(B^*) \supseteq L^*$: Für jedes $a \in L^*$ ist zu zeigen, daß $a \in L(B^*)$. Da $b_1, b_2, \ldots, b_n \in L$, erhalten wir aus der Definition des dualen Gitters:

$$B^\mathsf{T} \cdot a = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \cdot a = \begin{bmatrix} \langle b_1, a \rangle \\ \langle b_2, a \rangle \\ \vdots \\ \langle b_n, a \rangle \end{bmatrix} \in \mathbb{Z}^n$$

  Ferner ist:

$$a = E \cdot a = \left(B \cdot B^{-1}\right)^\mathsf{T} \cdot a = \left(B^{-1}\right)^\mathsf{T} \cdot B^\mathsf{T} \cdot a = B^* \cdot \underbrace{\left(B^\mathsf{T} \cdot a\right)}_{\in \mathbb{Z}^n}$$

  Setze $t_i := \langle b_i, a \rangle \in \mathbb{Z}$ für $i = 1, 2, \ldots, n$. Dann hat a die Darstellung

$$a = \sum_{i=1}^{n} t_i b_i^*$$

  und $a \in L(B^*)$.

Die Behauptung $(L^*)^* = L$ gilt, da die Reihenfolge für Transponieren und Inverses vertauscht werden können. ■

Definition 2.2.30 (Selbstduales Gitter)
Ein Gitter L heißt selbstdual, wenn $L = L^*$.

Zum Beispiel ist $\mathbb{Z}^n$ ein selbstduales Gitter. Zum Abschluß des Abschnittes definieren wir ganze Gitter:

Definition 2.2.31 (Ganzes Gitter)
Ein Gitter L heißt ganz, wenn $\langle a, b \rangle \in \mathbb{Z}$ für alle $a, b \in L$.

Satz 2.2.32
Für jedes ganze Gitter L gilt: $L \subseteq L^* \subseteq \frac{1}{\det L^2} L$.

Beweis. Nachrechnen. ■

2.3 Längen- und gewichtsreduzierte Gitterbasen

Wir definieren in diesem Abschnitt zwei algorithmisch motivierte Reduktionsbegriffe: Längen- und gewichtsreduzierte Gitterbasen.

Definition 2.3.1 (Längenreduzierte Basis)
Die geordnete Basis $b_1, b_2, \ldots, b_n$ ist längenreduziert, wenn $|\mu_{i,j}| \leq \frac{1}{2}$ für $1 \leq j < i \leq n$.

Wir erhalten:

Satz 2.3.2
Für jede geordnete, längenreduzierte Basis $b_1, b_2, \ldots, b_n$ gilt:

$$\|b_i\|^2 \leq \|\widehat{b}_i\|^2 + \frac{1}{4} \sum_{j=1}^{i-1} \|\widehat{b}_j\|^2 \qquad \text{für } i = 1, 2, \ldots, n$$

Beweis. Nach Gleichung (2.12) von Seite 27 gilt für $i = 1, 2, \ldots, n$:

$$b_i = \widehat{b}_i + \sum_{j=1}^{i-1} \mu_{i,j} \widehat{b}_j$$

Die Vektoren $\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n$ des Orthogonalsystems stehen senkrecht aufeinander, so daß folgt:

$$\|b_i\|^2 = \widehat{b_i}^2 + \sum_{j=1}^{i-1} \mu_{i,j}^2 \|\widehat{b}_j\|^2$$

Da die Basis nach Voraussetzung längenreduziert ist, gilt $|\mu_{i,j}| \leq \frac{1}{2}$, und wir erhalten die Behauptung. ■

Algorithmus 2.3.1 Längenreduktion

EINGABE: Geordnete Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$

1. FOR $i = 2, 3, \ldots, n$ DO

FOR $j = i-1, i-2, \ldots, 1$ DO

$b_i := b_i - \lceil \mu_{i,j} \rfloor \cdot b_j$

END for j

END for i

AUSGABE: Längenreduzierte Basis $b_1, b_2, \ldots, b_n$

Der Algorithmus 2.3.1 transformiert eine gegebene Gitterbasis in eine längenreduzierte Basis desselben Gitters. Für die Korrektheit des Algorithmus' 2.3.1 zur Längenreduktion betrachten wir die Basisvektoren bezüglich der Koordinaten im Orthogonalsystem:

$$\begin{array}{ccccccccccc}
 & & & \widehat{b}_1 & \widehat{b}_2 & \widehat{b}_3 & & \widehat{b}_{n-1} & \widehat{b}_n & \\
b_1 & = & (& 1 & 0 & 0 & \cdots & 0 & 0 &) \\
b_2 & = & (& \mu_{2,1} & 1 & 0 & & 0 & 0 &) \\
b_3 & = & (& \mu_{3,1} & \mu_{3,2} & 1 & & 0 & 0 &) \\
\vdots & = & (& \vdots & & & \ddots & 0 & \vdots &) \\
b_{n-1} & = & (& \mu_{n-1,1} & \mu_{n-1,2} & \mu_{n-1,3} & & 1 & 0 &) \\
b_n & = & (& \mu_{n,1} & \mu_{n,2} & \mu_{n,3} & & \mu_{n,n-1} & 1 &)
\end{array}$$

Wir beobachten:

- Der Schritt $b_i := b_i - \lceil \mu_{i,j} \rfloor b_j$ bewirkt, daß $\mu_{i,j}^{\text{neu}} := \mu_{i,j}^{\text{alt}} - 1 \cdot \lceil \mu_{i,j}^{\text{alt}} \rfloor$.
- Insbesondere gilt dann $\left|\mu_{i,j}^{\text{neu}}\right| \leq \frac{1}{2}$, und die $\mu_{i,\nu}$ bleiben für $\nu > j$ unverändert.

Die Höhen der Basisvektoren bleiben erhalten und die Reihenfolge der Basisvektoren wird nicht verändert. Daher ist Längenreduzierung nur ein schwacher Reduktionsbegriff.

Definition 2.3.3 (Gewichtsreduzierte Basis)
Die geordnete Basis $b_1, b_2, \ldots, b_n$ ist gewichtsreduziert, wenn:

a) $\dfrac{|\langle b_i, b_j\rangle|}{\|b_j\|^2} \leq \dfrac{1}{2}$ *für* $1 \leq j < i \leq n$

b) $\|b_i\| \leq \|b_{i+1}\|$ *für* $i = 1, 2, \ldots, n-1$

Beachte, daß es sich bei der ersten Forderung nicht um den Gram-Schmidt-Koeffizienten

$$\mu_{i,j} = \frac{\left\langle b_i, \widehat{b}_j \right\rangle}{\|\widehat{b}_j\|^2}$$

handelt. Die Eigenschaft b) einer gewichtsreduzierten Basis fordert, daß die Basisvektoren nach aufsteigender Länge angeordnet sind:

$$\|b_1\| \leq \|b_2\| \leq \|b_3\| \leq \cdots \leq \|b_{n-1}\| \leq \|b_n\|$$

Die Eigenschaft a) ist äquivalent zu

$$\|b_i\| \leq \|b_i \pm b_j\| \qquad \text{für } 1 \leq j < i \leq n$$

Denn wegen

$$\|b_i \pm b_j\|^2 = \langle b_i \pm b_j, b_i \pm b_j\rangle = \|b_i\|^2 \pm 2 \cdot \langle b_i, b_j\rangle + \|b_j\|^2$$

gilt:

$$\|b_i\|^2 \leq \|b_i \pm b_j\|^2 \quad \Longleftrightarrow \quad \pm\langle b_i, b_j\rangle \leq \tfrac{1}{2} \cdot \|b_j\|^2 \quad \Longleftrightarrow \quad |\langle b_i, b_j\rangle| \leq \tfrac{1}{2} \cdot \|b_j\|^2$$

Wir werden in Kapitel 4 den Spezialfall $n = 2$, also eine Basis bestehend aus zwei Gittervektoren, betrachten.

Algorithmus 2.3.2 Gewichtsreduktion

EINGABE: Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$

1. F :=true

2. WHILE (F) DO

2.1. Ordne $b_1, b_2, \ldots, b_n$ so, daß $\|b_1\| \leq \|b_2\| \leq \cdots \leq \|b_n\|$

2.2. F :=false

2.3. FOR $i = 1, 2, \ldots, n$ DO

FOR $j = 1, 2, \ldots, i-1$ DO /* Reduktionsschritt */

$r := \langle b_i, b_j\rangle \cdot \|b_j\|^{-2}$

IF $|r| > \frac{1}{2}$ THEN $b_i := b_i - \lceil r \rfloor b_j$ und F :=true

END for j

END for i

END while

AUSGABE: Gewichtsreduzierte Basis $b_1, b_2, \ldots, b_n$

Satz 2.3.4
Jedes Gitter hat eine gewichtsreduzierte Basis.

Zum Beweis des Satzes verweisen wir auf Algorithmus 2.3.2 „Gewichtsreduktion“, der eine gegebene Gitterbasis in eine gewichtsreduzierte Basis desselben Gitters transformiert. Die Korrektheit des Verfahrens folgt aus den beiden Beobachtungen:

- Bei Abbruch des Verfahrens liegt offenbar eine gewichtsreduzierte Basis vor.
- Der Algorithmus terminiert, da bei jedem Reduktionsschritt ein Basisvektor echt verkleinert wird und die übrigen Basisvektoren nicht verändert werden.

Im Gegensatz zu Algorithmus 2.3.1 zur Längenreduzierung vertauscht das Verfahren zur Gewichtsreduktion die Reihenfolge der Vektoren und insbesondere steht der kürzestes Basisvektor an erster Stelle. Der Algorithmus ist aber nicht effizient. Wir lernen in den folgenden Kapiteln stärkere Reduktionsbegriffe kennen: In Kapitel 5 den Begriff der LLL-reduzierten Gitterbasen und in Kapitel 7 den Begriff der HKZ- sowie β-reduzierten Gitterbasen.

2.4 Beispiele

In diesem Abschnitt stellen wir Gitter und zugehörige Basen vor, die in den Übungen behandelt wurden. Das Skalarprodukt ist das Standard-Skalarprodukt. Das erste sukzessive Minima λ_1 ist die minimale Länge eines Gittervektors ungleich 0 (formale Definition folgt später).

Zum Gitter

$$A_n := \left\{ (x_0, x_1, \ldots, x_n) \in \mathbb{Z}^{n+1} \;\middle|\; \sum_{i=0}^{n} x_i = 0 \right\}$$

bilden die folgenden Zeilenvektoren eine Basis:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} := \begin{bmatrix} -1 & +1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & +1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & -1 & +1 & 0 \\ 0 & \cdots & 0 & 0 & -1 & +1 \end{bmatrix}$$

Offenbar ist $L(b_1, b_2, \ldots, b_n)$ ein Untergitter von A_n, da $b_1, b_2, \ldots, b_n \in A_n$ und sie linear unabhängig sind. Sei umgekehrt $x = (x_0, x_1, \ldots, x_n) \in A_n$ beliebig. Ziehe für $i = n, n-1, \ldots, 2$ von x den Vektor b_i ab, bis jeweils die $(i+1)$-te Komponente 0 ist. Der Vektor x' hat nur in den ersten beiden Einträgen Werte ungleich 0, und er liegt in A_n. Also gilt $x'_0 = -x'_1$, und der Vektor ist ein ganzzahliges Vielfaches von b_1.

Zum Gitter

$$D_n := \left\{ (x_1, x_2, \ldots, x_n) \in \mathbb{Z}^n \;\middle|\; \sum_{i=0}^{n} x_i \equiv 0 \pmod 2 \right\}$$

bilden die folgenden Zeilenvektoren eine Basis:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} := \begin{bmatrix} +2 & 0 & 0 & 0 & \cdots & 0 \\ +1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & +1 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & +1 & -1 & 0 \\ 0 & \cdots & 0 & 0 & +1 & -1 \end{bmatrix}$$

Eine Alternative ist $b_1' = (-1, -1, 0, \ldots, 0)$. Wegen $b_1 = -b_1' + b_2$ erhält man dasselbe Gitter. Offenbar ist $L(b_1, b_2, \ldots, b_n)$ ein Untergitter von D_n, da $b_1, b_2, \ldots, b_n \in D_n$ und sie linear unabhängig sind. Sei umgekehrt $x = (x_1, x_2, \ldots, x_n) \in D_n$ beliebig. Ziehe für $i = n, n-1, \ldots, 2$ von x den Vektor b_i ab, bis jeweils die i-te Komponente 0 ist. Der Vektor hat nur im ersten Eintrag einen Wert ungleich 0, und er liegt in D_n. Also ist $x_1 \equiv 0 \pmod 2$, und der Vektor ist ein ganzzahliges Vielfaches von b_1. Es gilt $\det D_n = 2$. Offenbar gibt es keine kürzeren Gittervektoren als die Basisvektoren, so daß $\lambda_1(D_n) = \sqrt{2}$ ist.

Sei $n = 0 \bmod 4$. Zum Gitter

$$E_n := \left\{ (x_1, x_2, \ldots, x_n) \in \mathbb{Z}^n \;\middle|\; \begin{array}{l} \sum_{i=0}^{n} x_i \equiv 0 \pmod 4 \text{ und} \\ x_j \equiv x_{j+1} \pmod 2 \text{ für } 1 \le j < n \end{array} \right\}$$

bilden die folgenden Zeilenvektoren eine Basis:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} := \begin{bmatrix} +4 & 0 & 0 & 0 & \cdots & 0 \\ +2 & -2 & 0 & 0 & \cdots & 0 \\ 0 & +2 & -2 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & +2 & -2 & 0 \\ +1 & \cdots & +1 & +1 & +1 & +1 \end{bmatrix}$$

Offenbar ist $L(b_1, b_2, \ldots, b_n)$ ein Untergitter von E_n, da $b_1, b_2, \ldots, b_n \in E_n$ und sie linear unabhängig sind. Sei umgekehrt $x = (x_1, x_2, \ldots, x_n) \in E_n$ beliebig. Betrachte $x' := x - x_n b_n \in E_n$, dessen letzte Komponente 0 ist. Wegen $x' \in E_n$ sind alle Einträge gerade. Ziehe für $i = n-1, n-2, \ldots, 2$ von x' den Vektor b_i ab, bis jeweils die i-te Komponente 0 ist. Der Vektor hat nur im ersten Eintrag einen Wert ungleich 0, und er liegt in E_n. Also gilt $x_1 \equiv 0 \pmod 4$, und der Vektor ist ein ganzzahliges Vielfaches von b_1. Wir erhalten $\det E_n = 2^n$. Es gilt

$$\lambda_1(E_n)^2 = \begin{cases} 4 & \text{falls } n = 4 \\ 8 & \text{sonst} \end{cases}$$

Für $n = 4$ ist zum Beispiel $u := (1, 1, 1, 1)$ ein Vektor mit $\|u\|^2 = 4$, für $n > 4$ ist zum Beispiel $v := (2, 2, 0, \ldots, 0)$ ein Vektor mit $\|v\|^2 = 8$. Man überlege sich, daß es keine kürzeren Gittervektoren gibt (Beachte: Es gilt für alle Gittervektoren $x_{n-1} \equiv x_n \pmod 2$).

Sei $a = (a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n \setminus \{0\}$. Wir betrachten zum Vektor a das Gitter, dessen Vektoren ganzzahlig sind und senkrecht auf a stehen:

$$L_a := \operatorname{span}(a)^\perp \cap \mathbb{Z}^n = \{ t \in \mathbb{Z}^n \mid \langle a, t \rangle = 0 \}$$

Wir zeigen:

$$\det L_a = \frac{\|a\|}{\operatorname{ggT}(a_1, a_2, \ldots, a_n)}$$

Das Gitter L_a hat offenbar die Dimension $n-1$. Sei $c_2, c_3, \ldots, c_n$ eine Basis von L_a. Wegen

$$\operatorname{span}(L_a) \cap \mathbb{Z}^n = L_a$$

bilden die Vektoren der Basis ein primitives System bezüglich des Gitters $\mathbb{Z}^n$ und nach Satz 2.2.23 existiert ein Vektor $c_1 \in \mathbb{Z}$ mit:

$$L(c_1, c_2, \ldots, c_n) = \mathbb{Z}^n$$

Die Einträge des Vektors c_1 sind teilerfremd, da c_1 ein primitiver Vektor bezüglich $\mathbb{Z}^n$ ist. Der Anteil von c_1, der senkrecht auf L_a steht, ist

$$\frac{\langle c_1, a \rangle}{\|a\|^2} \cdot a$$

und hat die Länge:

$$\left\langle \frac{\langle c_1, a\rangle}{\|a\|^2} \cdot a, \frac{\langle c_1, a\rangle}{\|a\|^2} \cdot a \right\rangle^{\frac{1}{2}} = \frac{\langle c_1, a\rangle}{\|a\|^2} \cdot \langle a, a\rangle^{\frac{1}{2}} = \frac{\langle c_1, a\rangle}{\|a\|}$$

Wir erhalten aus der geometrischen Interpretation der Gitterdeterminanten als Volumen der Grundmasche („Fläche $\det L_a$ mal Höhe $\frac{|\langle c_1, a\rangle|}{\|a\|}$“):

$$\det \mathbb{Z}^n = 1 = \det L_a \cdot \frac{|\langle c_1, a\rangle|}{\|a\|}$$

$|\langle c_1, a\rangle|$ ist die kleinste, positive, ganze Zahl im Hauptideal $\sum_{i=1}^{n} \mathbb{Z}a_i = \mathbb{Z} \cdot \mathrm{ggT}(a_1, a_2, \ldots, a_n)$, denn es gibt ein $\bar{c}_1 \in \mathbb{Z}^n$ mit

$$\langle \bar{c}_1, a\rangle = \mathrm{ggT}(a_1, a_2, \ldots, a_n)$$

und $c_1 = k \cdot \bar{c}_1$ für ein $k \in \mathbb{Z}$. Da aber die Einträge des Vektors c_1 teilerfremd sind, folgt $|k| = 1$. Wir erhalten daher

$$\langle c_1, a\rangle = \mathrm{ggT}(a_1, a_2, \ldots, a_n)\,,$$

und es folgt die Behauptung.

Kapitel 3

Sukzessive Minima, Minkowski-Sätze und Hermite-Konstante

In diesem Kapitel definieren wir die sukzessiven Minima zu einem Gitter. Wir lernen zwei bekannte Sätze von Minkowski kennen, u.a. den in der englichen Literatur als „Minkowski's Convex Body Theorem" bezeichneten zweiten Satz von Minkowski. Wir definieren die Hermite-Konstante γ_n und leiten untere und obere Schranken her. Sofern nichts anderes angegeben ist, sei in diesem Kapitel $\langle \cdot, \cdot \rangle$ das Standard-Skalarprodukt und $\|x\| = \sqrt{\langle x, x \rangle}$ die Euklidische Norm.

3.1 Sukzessive Minima und erster Satz von Minkowski

In Kapitel 2.4 haben wir das erste sukzessive Minimum bereits informell eingeführt. Allgemein sind die sukzessiven Minima definiert als:

Definition 3.1.1 (Sukzessive Minima $\lambda_1, \lambda_2, \ldots, \lambda_n$)
Sei $\|\cdot\|$ eine beliebige Norm. Zu einem Gitter $L \subseteq \mathbb{R}^m$ vom Rang n sind die sukzessiven Minima $\lambda_1, \lambda_2, \ldots, \lambda_n$ bezüglich der Norm $\|\cdot\|$ wie folgt definiert:

$$\lambda_i = \lambda_i(L) := \inf \left\{ r > 0 \;\middle|\; \begin{array}{l} \text{Es gibt } i \text{ linear unabhängige} \\ \text{Vektoren } c_1, c_2, \ldots, c_i \in L \\ \text{mit } \|c_j\| \leq r \text{ für } j = 1, 2, \ldots, i \end{array} \right\} \qquad \text{für } i = 1, 2, \ldots, n$$

Die Definition der sukzessiven Minima geht auf H. Minkowski zurück, der sie ursprünglich für das Gitter $\mathbb{Z}^n$ definiert hat. Abbildung 3.1.1 illustriert den Begriff anhand des Gitters $L(a, b)$: Das erste sukzessiven Minimum bezüglich der Euklidischen Norm realisiert der Vektor x.

Es gilt $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Die sukzessiven Minima sind geometrische Gitterinvarianten, d.h. diese Größen bleiben bei isometrischer Transformation des Gitters erhalten. Für jede Gitterbasis $b_1, b_2, \ldots, b_n$ gilt für $i = 1, 2, \ldots, n$:

$$\max_{j=1,2,\ldots,i} \|b_j\| \geq \lambda_i$$

Die sukzessiven Minima liefern einen Maßstab für die Reduziertheit einer Gitterbasis: Eine Basis gilt als „reduziert", wenn die Größen $\frac{\|b_i\|}{\lambda_i}$ für $i = 1, 2, \ldots, n$ „klein" sind. Für reduzierte Basen sind

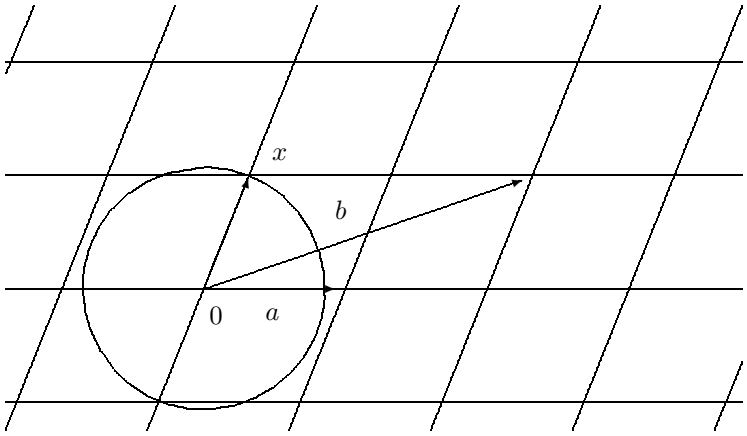

Abbildung 3.1.1: Beispiel zu erstem sukzessiven Minimum $\lambda_1(L(a,b))$

deren Vektoren nahezu orthogonal. Im allgemeinen gibt es keine Basis $b_1, b_2, \ldots, b_n$ mit $\|b_i\| = \lambda_i$ für $i = 1, 2, \ldots, n$. Betrachte beispielsweise das Gitter

$$L := \mathbb{Z}^n + \mathbb{Z}\big(\underbrace{\tfrac{1}{2}, \ldots, \tfrac{1}{2}}_{n\text{-mal}}\big)^{\mathsf{T}}$$

und die Euklidische Norm zum Standard-Skalarprodukt. Für $n \geq 5$ ist $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 1$ und die kanonischen Einheitsvektoren (die einzigen Gittervektoren mit Länge 1) bilden keine Basis.

Die sukzessiven Minima sind abhängig von der zugrundeliegenden Norm. Speziell betrachten wir das erste sukzessive Minimum in der Euklidischen Norm zum Standard-Skalarprodukt

$$\|L\| := \lambda_1(L) = \min\{\|b\| \ : \ b \in L \setminus \{0\}\}$$

und in der sup-Norm:

$$\|L\|_\infty := \lambda_{1,\infty}(L) = \min\{\|b\|_\infty \ : \ b \in L \setminus \{0\}\}$$

Für Gitter $L \subseteq \mathbb{R}^m$ gilt wegen $\|x\|_\infty \leq \|x\| \leq \sqrt{n} \cdot \|x\|_\infty$ für alle $x \in \mathbb{R}^m$:

$$\lambda_{1,\infty}(L) \leq \lambda_1(L) \leq \sqrt{n} \cdot \lambda_{1,\infty}(L)$$

Die Größe $\|L\|_\infty$ ist keine geometrische Invariante. Sie läßt sich aber durch eine scharfe Ungleichung begrenzen, wie folgender Satz zeigt:

Satz 3.1.2 (Minkowski 1896)
Sei $L \subseteq \mathbb{R}^m$ ein Gitter vom Rang n. Dann gilt $\|L\|_\infty \leq (\det L)^{\frac{1}{n}}$.

Diese Schranke ist scharf, denn es gilt $\|\mathbb{Z}^n\|_\infty = 1 = (\det \mathbb{Z}^n)^{\frac{1}{n}}$. Aus Satz 3.1.2 folgt als obere Schranke für die Länge des kürzesten, nicht-trivialen Gittervektors $\lambda_1(L) \leq \sqrt{n} \cdot (\det L)^{\frac{1}{n}}$. Bevor wir den Satz 3.1.2 beweisen, zeigen wir ein Ergebnis von H.F. Blichfeldt [Blich14]:

Lemma 3.1.3 (Blichfeldt 1914)
Sei $L \subseteq \mathbb{R}^m$ ein vollständiges Gitter und sei $Q \subseteq \mathbb{R}^m$ kompakt mit $\mathrm{vol}(Q) \geq \det L$. Dann gibt es ein $b \in L \setminus \{0\}$ mit $Q \cap (Q + b) \neq \emptyset$, d.h. es existieren $x, y \in Q$ mit $x - y \in L \setminus \{0\}$.

Beweis. Zu $i \in \mathbb{N}$ sind die Mengen $\left(1+\frac{1}{i}\right) Q$ und $\left(1+\frac{1}{i}\right) Q + b_i$ mit $b_i \in L \setminus \{0\}$ nicht paarweise disjunkt, weil das Volumen von $\left(1+\frac{1}{i}\right) Q$ das der Grundmasche übersteigt. Zu jedem i gibt es ein $b_i \in L \setminus \{0\}$, so daß der folgende Durchschnitt nicht-leer ist und wir aus ihm y_i wählen können:

$$y_i \in \left(1+\tfrac{1}{i}\right) Q \cap \left[\left(1+\tfrac{1}{i}\right) Q + b_i\right] \qquad\qquad i = 1, 2, \ldots$$

Da Q kompakt ist, hat die Folge $(y_i)_{i\in\mathbb{N}}$ einen Häufungspunkt $y \in Q$, so daß eine Teilfolge $\left(b_{\alpha(i)}\right)_{i\in\mathbb{N}} \subseteq L$ gegen y konvergiert. Die Folge $\left(b_{\alpha(i)}\right)_{i\in\mathbb{N}} \subseteq L$ ist beschränkt und durchläuft nur endlich viele Gitterpunkte. Mindestens ein Gitterpunkt $b \in L \setminus \{0\}$ wird unendlich oft durchlaufen. Es folgt: $y \in Q \cap (Q + b)$. ∎

Beweis (zu Satz 3.1.2). Wir behandeln zunächst den Fall vollständiger Gitter L, also $n = m$. Wir wenden Lemma 3.1.3 auf die Menge

$$Q := \left\{ x \in \mathbb{R}^m \,:\, \|x\|_\infty \leq \tfrac{1}{2} (\det L)^{\frac{1}{m}} \right\}$$

an. Q ist ein m-dimensionaler Würfel mit Kantenlänge $(\det L)^{\frac{1}{m}}$. Es gilt $\mathrm{vol}(Q) = \det L$. Nach Lemma 3.1.3 gibt es ein $b \in L \setminus \{0\}$ und ein $y \in Q \cap (Q + b)$. Wegen $y, y - b \in Q$ gilt:

$$\|y\|_\infty \leq \tfrac{1}{2} (\det L)^{\frac{1}{m}}$$
$$\|y - b\|_\infty \leq \tfrac{1}{2} (\det L)^{\frac{1}{m}}$$

Aus der Dreiecksungleichung folgt:

$$\|b\|_\infty \leq \|y\|_\infty + \|y - b\|_\infty \leq (\det L)^{\frac{1}{m}}$$

Den Fall $n < m$ führen wir auf die Situation $n = m$ zurück. Zu $I = (i_1, i_2, \ldots, i_n)$ mit $1 \leq i_1 < i_2 < \cdots < i_n \leq n$ sei:

$$\varphi_I(x_1, x_2, \ldots, x_m) := (x_{i_1}, x_{i_2}, \ldots, x_{i_n})$$

Es durchlaufe I diejenigen Auswahlen mit $\mathrm{span}\,(\varphi_I(\mathrm{span}\,(L)) = \mathbb{R}^n$. Dann gilt $\det \varphi_I(L) \leq \det L$, und für alle $b \in L$ ist $\|b\|_\infty = \max_I \|\varphi_I(b)\|_\infty$. Weil die Behauptung für $n = m$ bewiesen ist, folgt:

$$\|L\|_\infty = \max_I \|\varphi_I(b)\|_\infty \leq \max_I \big(\det \varphi_I(L)\big)^{\frac{1}{m}} \leq (\det L)^{\frac{1}{n}}$$

∎

Wir definieren:

Definition 3.1.4 (Konvexe, nullsymmetrische Menge)
$S \subseteq \mathbb{R}^m$ heißt konvex, falls mit $x, y \in S$ und $\xi \in [0,1]$ auch $\xi x + (1-\xi) y \in S$ gilt. S heißt nullsymmetrisch, falls mit x auch $-x$ in S liegt.

Satz 3.1.2 mit $n = m$ ist ein Spezialfall des folgenden Satzes, der in der englischsprachigen Literatur „Minkowski's Convex Body Theorem“ genannt wird:

Satz 3.1.5 (Erster Satz von Minkowski 1893)
Sei $L \subseteq \mathbb{R}^m$ ein vollständiges Gitter und $S \subseteq \mathbb{R}^m$ konvex, nullsymmetrisch, kompakt mit $\mathrm{vol}(S) \geq 2^m \cdot \det L$. Dann gilt $|S \cap L| \geq 3$, d.h. S enthält mindestens die Vektoren $\pm y \in L$ ungleich 0.

Beweis. Wir setzen:

$$Q := \tfrac{1}{2}S = \left\{ \tfrac{1}{2} \cdot x \;\middle|\; x \in S \right\} \subseteq \mathbb{R}^m$$

Es gilt $\operatorname{vol}(Q) = 2^{-m} \operatorname{vol}(S) \geq \det L$. Wir wenden Blichfeldts Lemma 3.1.3 auf Q an: Es existiert ein $b \in L \setminus \{0\}$ mit:

$$\left(\tfrac{1}{2}S\right) \cap \left(\tfrac{1}{2}S + b\right) \neq \emptyset$$

Sei y im Durchschnitt. Dann gilt $y \in \frac{1}{2}S$ und $y = x + b$ mit $x \in \frac{1}{2}S$. Es gilt $b = y - x$ mit $x, y \in \frac{1}{2}S$. Weil S nullsymmetrisch und konvex ist, folgt $b \in S$. Insbesondere gilt $\{0, \pm b\} \subseteq S \cap L$. ■

Satz 3.1.5 angewandt auf $S = \{x \in \mathbb{R}^m \,:\, \|x\|_\infty \leq \det L\}$ liefert Satz 3.1.2 für $n = m$, da $\operatorname{vol}(S) = 2^n \cdot \det L$. Als weitere Konsequenz von Satz 3.1.5 erhalten wir einen Beweis des folgenden, bekannten Satzes von G.L. Dirichlet [Di1842] über die simultane Approximation reeller Zahlen durch rationale Zahlen:

Satz 3.1.6 (Dirichlet 1842)
Seien $\alpha_1, \alpha_2, \ldots, \alpha_n$ reelle Zahlen und $\epsilon \in \left]0, \frac{1}{2}\right[$. Dann gibt es ganze Zahlen $p_1, p_2, \ldots, p_n$ und q mit $0 < q \leq \epsilon^{-n}$, so daß gilt:

$$\left| \alpha_i - \frac{p_i}{q} \right| \leq \frac{\epsilon}{q} \qquad \text{für } i = 1, 2, \ldots, n$$

Beweis. Wende Satz 3.1.5 auf $L = \mathbb{Z}^{n+1}$ und

$$S := \left\{ (p_1, p_2, \ldots, p_n, q) \in \mathbb{R}^{n+1} \,:\, |q| \leq \epsilon^{-n}, \quad |p_i - q\alpha_i| \leq \epsilon \text{ für } i = 1, 2, \ldots, n \right\}$$

an. S ist ein Quader mit Kantenlänge 2ϵ in den ersten n Richtungen und $2\epsilon^{-n}$ in der letzten:

$$\operatorname{vol}(S) = (2\epsilon)^n 2\epsilon^{-n} = 2^{n+1} \cdot \det \mathbb{Z}^{n+1}$$

S ist konvex, nullsymmetrisch und kompakt. Nach Satz 3.1.5 gibt es $(p_1, p_2, \ldots, p_n, q) \in S \cap \mathbb{Z}^{n+1}$ ungleich dem Nullvektor. Es ist $q \neq 0$, da sonst wegen $|p_i| \leq \epsilon < \frac{1}{2}$ und $p_i \in \mathbb{Z}$ gilt $(p_1, p_2, \ldots, p_n, q) = 0$. Wir erhalten die Behauptung. ■

3.2 Hermite-Konstante und kritische Gitter

In diesem Abschnitt definieren wir die Hermite-Konstante γ_n, leiten obere und untere Schranke her und lernen kritische Gitter kennen. Die Hermite-Konstante γ_n bezieht sich stets auf die Euklidische Norm:

Definition 3.2.1 (Hermite-Konstante γ_n)
Die Hermite-Konstante γ_n ist für $n \in \mathbb{N}$ definiert:

$$\gamma_n := \sup \left\{ \frac{\lambda_1(L)^2}{(\det L)^{\frac{2}{n}}} \;\middle|\; L \subseteq \mathbb{R}^n \text{ vollständiges Gitter} \right\}$$

Das erste sukzessive Minimum bezieht sich auf die Euklidische Norm zum Standard-Skalarprodukt.

Aus historischen Gründen betrachtet man das Quadrat. Es genügt, das Supremum über die vollständigen Gitter $L \subseteq \mathbb{R}^m$ zu nehmen, denn $\lambda_1(L)$ und $\det L$ sind geometrische Invarianten. Wegen

$$\frac{\lambda_1(L)^2}{(\det L)^{\frac{2}{n}}} = \frac{\lambda_1(\alpha L)^2}{(\det \alpha L)^{\frac{2}{n}}} \qquad \text{für } \alpha \in \mathbb{R} \setminus \{0\}$$

genügt es, das Supremum über vollständige Gitter $L \subseteq \mathbb{R}^m$ mit $\det L = 1$ zu nehmen. Weil die reduzierten Basen dieser Gitter über einem kompakten Bereich des $\mathbb{R}^{n^2}$ variieren, wird das Supremum angenommen, d.h. wir können die Hermite-Konstante als das Maximum der Menge definieren. Insbesondere ist $\gamma_1 = 1$.

Bemerkung 3.2.2
Wegen $\lambda_1(L) \leq \sqrt{n} \cdot \lambda_{1,\infty}(L) \leq \sqrt{n} \cdot (\det L)^{\frac{1}{n}}$ *folgt aus Satz 3.1.2,daß* $\gamma_n \leq n$ *ist.*

Wir verbessern im folgenden diese obere Schranke. Es bezeichne

$$S_n(r) := \{x \in \mathbb{R}^n \ : \ \|x\| \leq r\}$$

die n-dimensionale Kugel mit Radius r mit Mittelpunkt 0. Das Volumen der n-dimensionalen Kugel mit Radius 1 ist:

$$\text{vol}(S_n(1)) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(1+\frac{n}{2}\right)} = \frac{2 \cdot \pi^{\frac{n}{2}}}{n \cdot \Gamma\left(\frac{1}{2}n\right)} \tag{3.1}$$

Für die Gamma-Funktion gilt $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, $\Gamma(n+1) = n!$ für $n \in \mathbb{N}$ und allgemein für $x \in \mathbb{R}^+$ (Stirling'sche Approximation):

$$\Gamma(x+1) = x \cdot \Gamma(x) = \sqrt{2\pi x} \cdot \left(\frac{x}{e}\right)^x \left(1 + \mathcal{O}\left(\frac{1}{x}\right)\right)$$

Wir erhalten als obere Schranke für die Hermite-Konstante γ_n:

Satz 3.2.3
Es gilt:

$$\gamma_n \leq \frac{4}{\pi} \cdot \Gamma\left(1+\frac{n}{2}\right)^{\frac{2}{n}} \leq \frac{2n}{e\pi} + \mathcal{O}(1) \approx 0,2342n + \mathcal{O}(1)$$

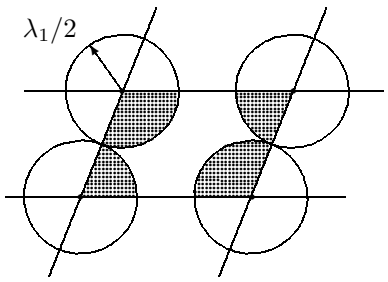

Abbildung 3.2.1: Veranschaulichung von (3.2)

Beweis. Sei L ein volldimensionale Gitter. Aus Abbildung 3.2.1 erhalten wir:

$$r < \tfrac{1}{2}\lambda_1 \quad \Rightarrow \quad \text{vol}(S_n(r)) < \det L. \tag{3.2}$$

Für volldimensionale Gitter L gilt daher folgende Implikation:

$$\text{vol}(S_n(r)) \geq \det L \quad \Rightarrow \quad \lambda_1 \leq 2r.$$

Sei $L \subseteq \mathbb{R}^m$ ein vollständiges Gitter mit $\gamma_n = \frac{\lambda_1(L)^2}{(\det L)^{\frac{2}{n}}}$ und $\det L = 1$. Wir suchen das minimale $r > 0$ mit $\mathrm{vol}(S_n(r)) \geq \det L$. Wegen

$$\mathrm{vol}(S_n(r)) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(1+\frac{n}{2}\right)} \cdot r^n$$

ist dies $r_{\min} := \frac{\Gamma\left(1+\frac{n}{2}\right)^{\frac{1}{n}}}{\pi^{\frac{1}{2}}}$ und wir erhalten aus $\lambda_1 \leq 2r_{\min}$:

$$\gamma_n = \frac{\lambda_1(L)^2}{(\det L)^{\frac{2}{n}}} \leq (2r_{\min})^2 = \frac{4}{\pi} \cdot \Gamma\left(1+\frac{n}{2}\right)^{\frac{2}{n}}.$$

■

Setzt man auf jeden Punkt des vollständigen Gitters $L \subseteq \mathbb{R}^n$ eine Kugel mit Radius $r := \frac{1}{2}\lambda_1$, erhält man die gitterartige Kugelpackung zu L.

Definition 3.2.4 (Gitterartige Kugelpackung)
Sei $M \subseteq \mathbb{R}^n$ eine nicht-leere, diskrete Menge und $r > 0$. Die Kugelpackung zu M und r ist:

$$\{m + S_n(r) \mid m \in M\}$$

Die Kugelpackung heißt gitterartig, wenn zu je zwei verschiedenen Punkten $m_1, m_2 \in M$ die Kugeln $m_1 + S_n(r)$ und $m_2 + S_n(r)$ keinen gemeinsamen, inneren Punkt haben.

Betrachten wir die gitterartige Kugelpackung zu L und $\frac{1}{2}\lambda_1$. Die 2^n Kugelteile in einer Grundmasche des Gitters ergeben zusammen gerade eine Kugel vom Radius $\frac{1}{2}\lambda_1$. Es folgt:

$$\det L \geq \mathrm{vol}\left(S_n\left(\tfrac{1}{2}\lambda_1\right)\right) = \frac{\lambda_1^n \cdot \mathrm{vol}(S_n(1))}{2^n} \tag{3.3}$$

Definition 3.2.5 (Dichte des Gitters)
Die Dichte des vollständigen Gitters $L \subseteq \mathbb{R}^n$ ist die Dichte der gitterartige Kugelpackung zu L, d.h. der Volumenanteil der Kugeln der Kugelpackung bezogen auf $\mathbb{R}^n$.

Die Dichte der Kugelpackung zum Gitter L ist:

$$\frac{\mathrm{vol}\left(S_n\left(\frac{1}{2}\lambda_1\right)\right)}{\det L} = \frac{\lambda_1^n}{\det L} \cdot \frac{\mathrm{vol}(S_n(1))}{2^n}$$

Für festes n die die Dichte maximal, wenn der Faktor $\frac{\lambda_1^n}{\det L}$ größtmöglich ist. Aus der Definition der Hermite-Konstanten

$$\gamma_n = \max\left\{ \frac{\lambda_1(L')^2}{(\det L')^{\frac{2}{n}}} \;\middle|\; L' \subseteq \mathbb{R}^n \text{ vollständiges Gitter} \right\}$$

erhalten wir, daß die Dichte des Gitters $L \subseteq \mathbb{R}^n$ genau dann maximal ist, wenn gilt:

$$\frac{\lambda_1^n}{\det L} \stackrel{!}{=} (\gamma_n)^{\frac{n}{2}} = \max\left\{ \frac{\lambda_1(L')^n}{\det L'} \;\middle|\; L' \subseteq \mathbb{R}^n \text{ vollständiges Gitter} \right\}$$

Wir definieren global extreme bzw. kritische Gitter:

Definition 3.2.6 (Global extremes (oder kritisches) Gitter)
Ein vollständiges Gitter $L \subseteq \mathbb{R}^n$ heißt global extrem (oder kritisch), wenn gilt

$$\frac{\lambda_1(L)^2}{(\det L)^{\frac{2}{n}}} = \gamma_n,$$

d.h. $\frac{\lambda_1(L)^2}{(\det L)^{\frac{2}{n}}}$ *das absolute Maximum für Gitter vom Rang n ist.*

Ein Gitter ist genau dann kritisch, wenn die Kugeln vom Radius $\frac{1}{2}\lambda_1$ um die Gitterpunkte eine dichteste, gitterartige Kugelpackung des $\mathbb{R}^n$ bilden.

Definition 3.2.7 (Lokal extremes Gitter)
Ein vollständiges Gitter $L = L(b_1, b_2, \ldots, b_n) \subseteq \mathbb{R}^n$ heißt lokal extrem, wenn

$$\frac{\lambda_1(L)^2}{(\det L)^{\frac{2}{n}}}$$

bei infinitesimal kleiner Veränderung der Basisvektoren nicht zunimmt.

Diese Eigenschaft hängt nicht von der Wahl der Basis von L ab. Jedes kritische Gitter ist extrem, aber es gibt extreme Gitter, die nicht kritisch sind. Die der Basis $b_1, b_2, \ldots, b_n$ zugeordnete Form $F(x_1, x_2, \ldots, x_n) := \sum_{i=1}^{n} \langle b_i, b_j \rangle x_i x_j$ nennt man dann *Extremform*.

Die obere Schranke $\gamma_n \leq \frac{2n}{e\pi} + \mathcal{O}(1)$ für die Hermite-Konstante aus Satz 3.2.3 auf Seite 41 hat H.F. Blichfeldt [Blich14] verbessert zu:

$$\gamma_n \leq \frac{2}{\pi} \cdot \Gamma\left(2 + \frac{n}{2}\right)^{\frac{2}{n}} \leq \frac{n}{e\pi} + \mathrm{o}(n)\,. \tag{3.4}$$

Diese Schranke berücksichtigt, daß im Beweis von Satz 3.2.3 nur ein sehr kleiner Anteil des Raumes von Kugeln mit Radius $\frac{1}{2}\lambda_1$ überdeckt wird und schätzt diesen Anteil durch $\left(\sqrt{2} + \mathrm{o}(1)\right)^{-n}$ nach oben ab. Es gilt zum Beispiel $\gamma_{10} \leq \frac{2}{\pi}(6!)^{0,2} \approx 2,373$. G.A. Kabatiansky und V.I. Levenshtein [KaLe78] zeigen 1978, daß:

$$\gamma_n \leq \frac{1,744}{2e\pi} n + \mathrm{o}(n) \approx 0,1021n + \mathrm{o}(n)$$

Diese Verbesserung zeigen, daß Minkowskis erster Satz 3.1.5 über konvexe Körper (Seite 39) für die Kugel S_n nicht optimal ist. Als untere Schranke für die Hermite-Konstante γ_n hat man:

$$\gamma_n \geq \frac{1}{\pi}\left(1 + \frac{n}{2}\right) = \frac{n}{2e\pi} + \mathrm{o}(n)$$

Diese untere Schranke erhält man durch Anwendung des folgenden Satzes von H. Minkowski und E. Hlawka [Hlawka44] auf Kugeln $S := S_n(r)$. Der Beweis ist allerdings nicht konstruktiv, es wird nur die Existenz solcher Gitter nachgewiesen, explizite Konstruktionen solcher Gitter sind nicht bekannt.

Satz 3.2.8 (Minkowski, Hlawka)
Sei $S \subseteq \mathbb{R}^n$ mit Jordan-Volumen kleiner als 1. Dann gibt es ein vollständiges Gitter $L \subseteq \mathbb{R}^n$ mit $\det L = 1$ *und* $(L \cap S) \setminus \{0\} \neq \emptyset$.

Beweis. Siehe [GrLek87, Theorem 1,Paragraph 19, Kapitel 3]. ■

Zusammenfassend gelten für die Hermite-Konstante γ_n die beiden folgenden Abschätzungen:

$$\frac{n}{2e\pi} + \mathrm{o}(n) \leq \gamma_n \leq \frac{n}{e\pi} + \mathrm{o}(n)$$

Man vermutet, daß die Hermite-Konstanten γ_n als Funktion in n monoton wachsend sind. In [GrLek87] (Paragraph 38, Kapitel 6) findet sich folgende Abschätzung:

$$\frac{1}{2e\pi} \leq \liminf_{n\to\infty} \frac{\gamma_n}{n} \leq \limsup_{n\to\infty} \frac{\gamma_n}{n} \leq \frac{1}{e\pi}$$

Es ist unbekannt, ob der Grenzwert $\lim_{n\to\infty} \frac{\gamma_n}{n}$ existiert. Die Hermite-Konstanten $\gamma_2, \gamma_3, \gamma_4, \gamma_5$ wurden in der zweiten Hälfte des 19. Jahrhunderts von A. Korkine und G. Zolotareff [KoZo1872, KoZo1873, KoZo1877] bestimmt. 1935 hat H.F. Blichfeldt [Blich35] $\gamma_6, \gamma_7, \gamma_8$ ermittelt:

n	2	3	4	5	6	7	8
$(\gamma_n)^n$	$\frac{4}{3}$	2	4	8	$\frac{2^6}{3}$	$\frac{2^7}{2}$	2^8
$\left(\frac{2}{\pi}\right)^n \cdot \Gamma\left(2+\frac{n}{2}\right)^2$	1,62	2,85	5,9	14,14	38,34	116	388
$\gamma_n \Big/ \frac{2}{\pi} \cdot \Gamma\left(2+\frac{n}{2}\right)^{\frac{2}{n}}$	0,907	0,887	0,907	0,892	0,907	0,918	0,949

Die letzte Zeile zeigt die Relation γ_n dividiert durch Blichfeldts Abschätzung (3.4). Blichfeldts Beweis ist kompliziert und wurde von G.L. Watson [Watson66] und N.M. Vetchinkin [Vetchin82]. bestätigt.

Für $n = 2, 3, \ldots, 8$ sind die kritischen Gitter vom Rang n mit $\lambda_1 = 1$ bis auf Isometrie eindeutig bestimmt (ohne die Forderung $\lambda_1 = 1$ sind sie bis auf Isometrie und Skalierung, also Similarity oder Ähnlichkeit, eindeutig bestimmt). Dies wurde von E.S. Barnes [Barnes59] und N.M. Vetchinkin [Vetchin82] bewiesen. Wir zeigen, daß diese Gitter eine Basis $b_1, b_2, \ldots, b_n$ haben mit der Eigenschaft, daß:

$$\langle b_i, b_j \rangle = \begin{cases} 1 & \text{für } i = j \\ \frac{1}{2} & \text{für } 1 \leq i - j \leq 2 \\ 0 & \text{sonst.} \end{cases} \tag{3.5}$$

Die Skalarprodukte $\langle b_i, b_j \rangle$ bestimmen die zugehörige Gitterbasis $b_1, b_2, \ldots, b_n$ bis auf Isometrie. In der Isometrieklasse von Gitterbasen mit den oben vorgegebenen Skalarprodukten $\langle b_i, b_j \rangle$ gibt es genau eine obere Dreiecksmatrix $[b_1, b_2, \ldots, b_n]$ mit positiven Diagonalelementen. Es gilt:

$$\begin{bmatrix} b_1, & b_2, & b_3, & b_4, & b_5, & b_6, & b_7, & b_8 \end{bmatrix} := \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{3} & \frac{1}{2\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{\sqrt{3}}{2\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2\sqrt{2}} & \frac{1}{\sqrt{2\cdot 3}} & \sqrt{\frac{2}{3}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2\cdot 3}} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

Sei $L^{(n)} := L(b_1, b_2, \ldots, b_n)$ zu $n \leq 8$. Eine andere Basismatrix zu $L^{(4)}$ ist:

$$\begin{bmatrix} b_1, & b_2, & b_3, & b_4 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \sqrt{\frac{2}{3}} & \sqrt{\frac{3}{2}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Wegen $\|b_1\| = 1$ gilt $\lambda_1(L^{(n)}) \leq 1$. Sei $b := \sum_{i=1}^n t_i b_i$ mit $t_1, \ldots, t_n \in \mathbb{Z}$ ein beliebiger Gittervektor aus $L^{(n)}$ mit $b \neq 0$. Wir wollen zeigen, daß $\|b\| \geq 1$ ist. Aus

$$\|b\|^2 = \left\langle \sum_{i=1}^n t_i b_i, \sum_{j=1}^n t_j b_j \right\rangle = \sum_{i=1}^n t_i \left\langle b_i, \sum_{j=1}^n t_j b_j \right\rangle = \sum_{i=1}^n t_i \cdot \sum_{j=1}^n \langle b_i, b_j \rangle \cdot t_j$$

Wegen (3.5) gilt $\langle b_i, b_j \rangle \in \left\{0, \frac{1}{2}, 1\right\}$ und $\|b_i\|^2 = 1$, so daß wir

$$\|b\|^2 = \sum_{i=1}^n \left(t_i^2 \langle b_i, b_i \rangle + \sum_{\substack{j=1 \\ j \neq i}}^n t_i t_j \langle b_i, b_j \rangle \right) = \underbrace{\sum_{i=1}^n \left(t_i^2 \langle b_i, b_i \rangle + 2 \sum_{j<i} t_i t_j \langle b_i, b_j \rangle \right)}_{\in \mathbb{Z} \setminus \{0\}} \geq 1$$

erhalten. Weil für jeden beliebigen Vektor $b \in L \setminus \{0\}$ gilt $\|b\| \geq 1$, ist der Vektor b_1 mit $\|b_1\| = 1$ einer der kürzesten, nicht-trivialen Gittervektoren. Es folgt daher $\lambda_1 = 1$.

Insbesondere sieht man an der Konstruktion der Vektoren $b_1, b_2, \ldots$, daß die dichteste Kugelpackung im $\mathbb{R}^n$, $n \leq 8$, die dichteste Kugelpackung im $\mathbb{R}^{n-1}$ erweitert. Das obige Schema kann nicht für $n \geq 9$ fortgesetzt werden, da $[b_1, b_2, \ldots, b_9]$ singulär ist. Das Gitter $\sqrt{2}L^{(8)}$ ist *selbstdual*, d.h. es gilt:

$$\sqrt{2}L^{(8)} = \left(\sqrt{2}L^{(8)}\right)^*$$

Wir beschreiben die Konstruktion der vorgestellten kritischen Gitter. Zunächst zwei Definitionen:

Definition 3.2.9 (Tiefes Loch)
Sei L ein Gitter. Der Punkt $x \in \operatorname{span}(L)$ heißt tiefes Loch des Gitters L, wenn

$$\min\{\|x - y\| \ : \ y \in L\} = \max_{p \in \operatorname{span}(L)} (\min\{\|p - y\| \ : \ y \in L\})$$

Definition 3.2.10 (Laminated Gitter)
Das Gitter $L(b_1, b_2, \ldots, b_{n+1})$ ist laminated („verklebt“) bezüglich $L(b_1, b_2, \ldots, b_n)$, wenn

$$b_{n+1} - \pi_{n+1}(b_{n+1}) \in \operatorname{span}(b_1, b_2, \ldots, b_n)$$

tiefes Loch des Gitters $L(b_1, b_2, \ldots, b_n)$ ist.

Um die Gitter $L^{(n)}$ zu konstruieren, wählen wir $(1, 0, \ldots, 0) \in \mathbb{R}^n$ und $b_2, b_3, \ldots, b_n$ so, daß jeweils $L(b_1, b_2, \ldots, b_i)$ laminated bezüglich $L(b_1, b_2, \ldots, b_{i-1})$ und $\|b_i\| = 1$ für $i = 2, 3, \ldots, n$ ist. Grafik 3.2.2 zeigt diese Konstruktion für die beiden Gitter $L^{(2)}$ und $L^{(3)}$.

Die folgende Ungleichung von H. Minkowski verschärft die Ungleichung $\lambda_1^2 \leq \gamma_n (\det L)^{\frac{2}{n}}$, welche für beliebige Gitter L vom Rang n gilt:

Satz 3.2.11 (Minkowski'sche Ungleichung)
Für jedes Gitter L vom Rang n gilt:

$$\prod_{i=1}^n \lambda_i(L) \leq (\gamma_n)^{\frac{n}{2}} \cdot \det L$$

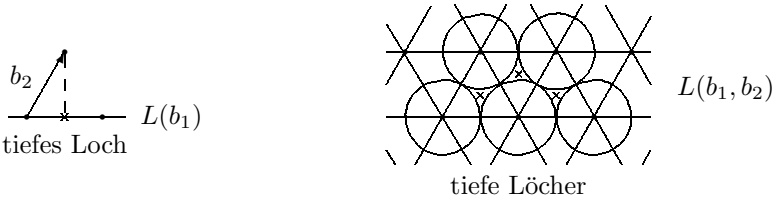

Abbildung 3.2.2: Konstruktion von $L^{(2)}$ und $L^{(3)}$

Bemerkung 3.2.12
Da für kritische Gitter $(\gamma_n)^{\frac{n}{2}} \det L = \lambda_1^n$ *und allgemein* $\prod_{i=1}^n \lambda_i \geq \lambda_1^n$ *gilt, ist für kritische Gitter* $\lambda_1 = \lambda_2 = \cdots = \lambda_n$.

Beweis (zu Satz 3.2.11). Seien $a_1, a_2, \ldots, a_n \in L$ linear unabhängige Vektoren, so daß:

$$\|a_i\| = \lambda_i \qquad \text{für } i = 1, 2, \ldots, n$$

$L_1 = L(a_1, a_2, \ldots, a_n)$ ist ein Untergitter von L. Wähle Basis $b_1, b_2, \ldots, b_n$ von L, so daß für $L_2 := L$ der Satz 2.2.19 (Seite 23) gilt: Es gibt eine obere Dreiecksmatrix $T \in M_{n,n}(\mathbb{Z})$ mit:

$$[b_1, b_2, \ldots, b_n] \cdot T = [a_1, a_2, \ldots, a_m]$$

Für alle $b \in L(b_1, b_2, \ldots, b_n)$ und $s = 1, 2, \ldots, n$ gilt:

$$b \notin L(b_1, b_2, \ldots, b_{s-1}) \quad \Longrightarrow \quad \|b\| \geq \lambda_s \tag{3.6}$$

Denn: Aus $b \notin L(b_1, b_2, \ldots, b_{s-1})$ und $b \in L$ folgt $b \notin \operatorname{span}(b_1, b_2, \ldots, b_{s-1})$, und weil T eine obere Dreicksmatrix ist, gilt nach Bemerkung 2.2.20 auf Seite 23:

$$\operatorname{span}(b_1, b_2, \ldots, b_{s-1}) = \operatorname{span}(a_1, a_2, \ldots, a_{s-1}),$$

so daß $a_1, a_2, \ldots, a_{s-1}, b$ linear unabhängig sind. Wir setzen für $i = 1, 2, \ldots, n$

$$\overline{b}_i := \sum_{j=1}^{i} \frac{\mu_{i,j} \widehat{b}_j}{\lambda_j}$$

und betrachten das Gitter $\overline{L} := L\left(\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_m\right)$. Behauptung:

$$\lambda_1\left(\overline{L}\right) \geq 1 \tag{3.7}$$

Sei $b := \sum_{i=1}^n t_i \overline{b}_i$ ein beliebiger Vektor aus $\overline{L} \setminus \{0\}$ und $s := \max_i \{i \mid t_i \leq 0\}$. Dann gilt

$$\left\| \sum_{i=1}^m t_i \overline{b}_i \right\|^2 = \sum_{j=1}^s \left(\sum_{i=j}^s t_i \mu_{i,j} \right)^2 \frac{\|\widehat{b}_j\|^2}{\lambda_j^2} \geq \sum_{j=1}^s \left(\sum_{i=j}^s t_i \mu_{i,j} \right)^2 \frac{\|\widehat{b}_j\|^2}{\lambda_s^2},$$

so daß wegen (3.6) und $t_s \neq 0$ folgt:

$$\left\| \sum_{i=1}^m t_i \overline{b}_i \right\|^2 \geq \frac{1}{\lambda_s^2} \sum_{j=1}^s \left(\sum_{i=j}^s t_i \mu_{i,j} \cdot \|\widehat{b}_j\| \right)^2 = \frac{1}{\lambda_s^2} \left\| \sum_{i=1}^s t_i b_i \right\|^2 \geq 1$$

Aus $\det \overline{L} = \frac{\det L}{\prod_{i=1}^n \lambda_i}$, Ungleichung (3.7) und der Definition der Hermite-Konstante $\frac{\lambda_1^2}{(\det L)^{\frac{2}{n}}} \leq \gamma_n$ erhalten wir:

$$1 \leq \lambda_1\left(\overline{L}\right)^2 \leq \gamma_n \cdot \left(\det \overline{L}\right)^{\frac{2}{n}} \leq \gamma_n \cdot (\det L)^{\frac{2}{n}} \left(\prod_{i=1}^n \lambda_i \right)^{-\frac{2}{n}}$$

Aus dieser Ungleichung folgt durch Erheben in die Potenz $\frac{n}{2}$ und Multiplikation mit $\prod_{i=1}^{n} \lambda_i$ die Behauptung:

$$\prod_{i=1}^{n} \lambda_i(L) \leq (\gamma_n)^{\frac{n}{2}} \det L$$

■

Eine untere Schranke für das Produkt der sukzessiven Minima ist die Gitterdeterminante:

Satz 3.2.13
Für jedes Gitter L vom Rang n gilt:

$$\prod_{i=1}^{n} \lambda_i \geq \det L$$

Beweis. Seien $a_1, a_2, \ldots, a_n$ linear unabhängige Gittervektoren mit $\|a_i\| = \lambda_i$ für $i = 1, 2, \ldots, n$. Weil $L(a_1, a_2, \ldots, a_n)$ ein Untergitter von L ist, gilt:

$$\det L(a_1, a_2, \ldots, a_n) \geq \det L \tag{3.8}$$

Andererseits gilt nach der Ungleichung von Hadamard:

$$\prod_{i=1}^{n} \|a_i\| \geq \det L(a_1, a_2, \ldots, a_n) \tag{3.9}$$

Aus den Abschätzungen (3.8) und (3.9) folgt die Behauptung $\prod_{i=1}^{n} \lambda_i \geq \det L$. ■

3.3 Gauge-Funktionen und Minkowski-Sätze

Wir führen Gauge-Funktionen ein (siehe u.a. [GrLek87, Siegel89]) und formulieren die beiden Minkowski-Sätze für das Gitter $\mathbb{Z}^n$ und anschließend allgemein. Wir definieren zunächst den Begriff des konvexen Körpers:

Definition 3.3.1 (Konvexer Körper)
Ein konvexer Körper $B \subseteq \mathbb{R}^n$ ist eine beschränkte, konvexe, offene Menge.

Die Menge ∂B zu einem konvexen Körper umfaßt genau alle Punkte $p \in \mathbb{R}^m \setminus B$, so daß in jeder Umgebung von p ein Punkt aus B liegt.

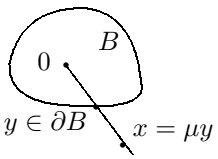

Abbildung 3.3.1: Veranschaulichung der Gauge-Funktion

Definition 3.3.2 (Gauge-Funktion)
Sei $B \subseteq \mathbb{R}^m$ ein konvexer Körper mit $0 \in B$. Die Gauge-Funktion $f : \mathbb{R}^n \to [0, \infty[$ ist wie folgt definiert:

$$f(x) := \begin{cases} 0 & \text{falls } x = 0 \\ 1 & \text{falls } x \in \partial B \\ \mu & \text{falls } x \neq 0,\ x \notin \partial B,\ x = \mu y \text{ mit } y \in \partial B \text{ und } \mu > 0 \end{cases}$$

Eine Gauge-Funktion f heißt gerade, falls für alle $x \in \mathbb{R}^n$ gilt $f(-x) = f(x)$.

Wir erhalten die ℓ_2-Norm mit

$$B := \left\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \ : \ x_1^2 + x_2^2 + \ldots x_n^2 < 1\right\}$$

und die sup-Norm mit:

$$B := \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \ : \ |x_i| < 1 \text{ für } i = 1, 2, \ldots, n\}$$

Es gilt:

Satz 3.3.3
Sei f Gauge-Funktion eines konvexen Körpers $B \subseteq \mathbb{R}^n$ mit $0 \in B$. Dann gilt für $x, y \in \mathbb{R}^n$:

a) $f(\mu \cdot x) = \mu \cdot f(x)$ *für* $\mu > 0$

b) $f(x) > 0$ *für* $x \neq 0$ *und* $f(0) = 0$

c) $f(x + y) \leq f(x) + f(y)$.

Beweis. Siehe [Siegel89, Theoreme 4, 5 und 6]. ■

Satz 3.3.4
Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine Funktion mit:

a) $f(\mu \cdot x) = \mu \cdot f(x)$ *für* $\mu > 0$ *und* $x \in \mathbb{R}^n$

b) $f(x) > 0$ *für* $x \neq 0$

c) $f(x + y) \leq f(x) + f(y)$.

Dann existiert ein konvexer Körper $B \subseteq \mathbb{R}^n$ mit f als Gauge-Funktion.

Beweis. Seieh [Siegel89, Theorem 7] mit $B := \{x \in \mathbb{R}^m \mid f(x) < 1\}$. ■

Sei $B \subset \mathbb{R}^n$ ein konvexer Körper. Der Punkt 0 ist *Zentrum* von B, falls:

$$x \in \partial B \quad \Longleftrightarrow \quad -x \in \partial B$$

Satz 3.3.5
Sei f eine Gauge-Funktion eines konvexen Körpers B. Genau dann ist 0 das Zentrum von B, wenn f gerade ist.

Beweis. Siehe [Siegel89, Theoreme 8 und 9]. ■

Die geraden Gauge-Funktionen entsprechen genau den Normen. Der erste Satz von Minkowski lautet in Form der Gauge-Funktionen (vergleiche Satz 3.1.5 auf Seite 39):

Satz 3.3.6 (Erster Satz von Minkowski)
Sei $f : \mathbb{R}^n \to [0, \infty[$ eine gerade Gauge-Funktion zum konvexen Körper $B \subseteq \mathbb{R}^n$. Falls $\mathrm{vol}(B) \geq 2^n$, existiert ein $g \in \mathbb{Z}^n \setminus \{0\}$ mit $f(g) \leq 1$.

Beweis. Siehe [Siegel89, Theoreme 10 und 11]. ■

Aus dem ersten Satz von Minkowski folgt:

Korollar 3.3.7
Sei $f : \mathbb{R}^n \to [0, \infty[$ eine gerade Gauge-Funktion zum konvexen Körper $B \subseteq \mathbb{R}^n$, $\mu := \min_{x \in \mathbb{Z}^n \setminus \{0\}} f(x)$ und $V := \mathrm{vol}(B)$. Dann ist $\mu^n V \leq 2^n$.

Beweis. Zu $\nu > 0$ bezeichne $B_\nu := \{x \in \mathbb{R}^n \mid f(x) \leq \nu\}$. Es gilt $\mathrm{vol}(B_\nu) = \nu^n V$ und für $0 < \nu_1 < \nu_2$ ist $B_{\nu_1} \subset B_{\nu_2}$. Setze

$$\nu_0 := \sup \{\nu > 0 \mid B_\nu \cap \mathbb{Z}^n = \{0\}\}$$

B_{ν_0} ist offen und enthält außer dem 0-Punkt keinen ganzzahligen Punkt. Aus dem ersten Satz von Minkowski 3.3.6 folgt:

$$(\nu_0)^n V \leq 2^n$$

Wir zeigen $\mu \leq \nu_0$: Angenommen, es sei $\mu > \nu_0$, d.h. es gäbe ein $\epsilon > 0$ mit $\nu_0 + \epsilon = \mu$. Aus der Definition von ν_0 folgt, daß ein Punkt $g \in \mathbb{Z}^n \setminus \{0\}$ mit $g \in B_{\nu_0 + \epsilon}$ existiert. Dies ist ein Widerspruch:

$$f(g) < \nu_0 + \epsilon = \mu = \min_{x \in \mathbb{Z}^n \setminus \{0\}} f(x) \leq f(g)$$

Bemerkung: Aus der Definition von ν_0 folgt, da $\mu < \nu_0$ nicht möglich ist, daß $\nu_0 = \mu$ gilt. ■

Die Aussage von Satz 3.3.7 kann man verschärfen und erhält den zweiten Satz von Minkowski:

Satz 3.3.8 (Zweiter Satz von Minkowski 1907)
Seien $\lambda_1, \lambda_2, \ldots, \lambda_n$ die sukzessiven Minima des Gitters $\mathbb{Z}^n$ bezüglich der geraden Gauge-Funktion $f : \mathbb{R}^n \to [0, \infty[$. Sei V das Volumen des konvexen Körpers $B := \{x \in \mathbb{R}^n \mid f(x) < 1\}$. Dann gilt:

$$\frac{2^n}{n!} \leq V \cdot \lambda_1 \cdot \lambda_2 \ldots \lambda_n \leq 2^n$$

Beweis. Siehe Paragraph 9.1 in Kapitel 2 in [GrLek87]. Für die obere Schranke siehe auch Theorem 16 aus [Siegel89] mit Beweis in Lecture IV. ■

Als zweiten Satz von Minkowski bezeichnet man in der Literatur im allgemeinen nur die obere Schranke.

Satz 3.3.9 (Zweiter Satz von Minkowski für allgemeine Gitter)
Seien $\lambda_1, \lambda_2, \ldots, \lambda_n$ die sukzessiven Minima des vollständigen Gitters $L \subseteq \mathbb{R}^n$ bezüglich der geraden Gauge-Funktion $f : \mathbb{R}^n \to [0, \infty[$. Sei V das Volumen des konvexen Körpers $B := \{x \in \mathbb{R}^n \mid f(x) < 1\}$. Dann gilt:

$$\frac{\det L}{n!} \leq \frac{V}{2^n} \cdot \prod_{i=1}^{n} \lambda_i \leq \det L$$

Für den Fall, daß f die sup-Norm ist, also

$$B := \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \ : \ |x_i| < 1 \text{ für } i = 1, 2, \ldots, n\},$$

gilt $V = 2^n$. Für jedes vollständige Gitter $L \subseteq \mathbb{R}^n$ erhalten wir aus dem zweiter Satz von Minkowski:

$$\prod_{i=1}^{n} \lambda_{i,\infty} \leq \det L$$

Kapitel 4

Gauß'sches Reduktionsverfahren

In diesem Kapitel stellen wir das Gauß'sche Reduktionsverfahren für zweidimensionale Gitter vor. Das Verfahren stellt eine Verallgemeinerung des (zentrierten) Euklidischen Algorithmus' dar. Wir werden das Reduktionsverfahren für den Spezialfall der Euklidischen Norm und den allgemeinen Fall einer beliebigen Norm kennenlernen.

Während wir in Kapitel 3.1 an einem Beispiel gesehen habe, daß es im allgemeinen keine Basis mit Vektoren der Länge der sukzessiven Minima gibt, existiert im Fall eines zweidimensionalen Gitters stets eine solche Basis.

4.1 Reduzierte Basis

Wir führen einen Reduktionsbegriff für Basen mit zwei Vektoren ein:

Definition 4.1.1 ((Gauß-)reduzierte Basis)
Eine geordnete Gitterbasis $a, b \in \mathbb{R}^n$ ist (Gauß-)reduziert bezüglich der Norm $\|\cdot\|$, wenn:

$$\|a\| \leq \|b\| \leq \|a - b\| \leq \|a + b\|$$

Betrachten wir den Fall, daß die Norm durch das Skalarprodukt gegeben ist: $\|x\| = \sqrt{\langle x, x \rangle}$. Für den Gram-Schmidt-Koeffizienten $\mu_{2,1} = \frac{\langle a,b \rangle}{\|a\|^2}$ gilt:

$$\begin{array}{rcccrcl} \mu_{2,1} & \leq & \frac{1}{2} & \iff & \|b\| & \leq & \|a - b\| \\ \mu_{2,1} & \geq & 0 & \iff & \|a - b\| & \leq & \|a + b\| \end{array}$$

Damit ist die Basis a, b genau dann reduziert, wenn:

a) $\|a\| \leq \|b\|$

b) $0 \leq \mu_{2,1} \leq \frac{1}{2}$

Im Vergleich zur gewichtsreduzierten Basis fordern wir, daß nicht nur der Betrag von $\mu_{2,1}$ zwischen 0 und $\frac{1}{2}$ liegt, sondern der Wert selbst. Dies können wir stets durch den Übergang von b zu $-b$ erreichen. Grafik 4.1.1 zeigt die Reduktionsbedingung im Fall des Standard-Skalarproduktes. Der Winkel ϕ zwischen den beiden Gittervektoren der reduzierten Basis liegt im Bereich von 60° bis 90°, denn:

$$\cos \phi = \frac{\langle a, b \rangle}{\|a\| \cdot \|b\|} = \mu_{2,1} \cdot \frac{\|a\|}{\|b\|}$$

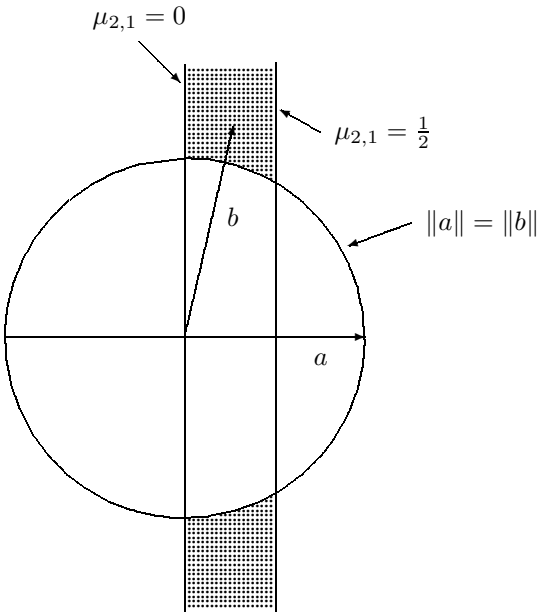

Abbildung 4.1.1: Reduzierte Basis für Standard-Skalarprodukt

Wegen $0 \leq \mu_{2,1} \leq \frac{1}{2}$ und $\|a\| \leq \|b\|$ ist:

$$0 \leq \cos\phi \leq \tfrac{1}{2}$$

Im Fall $\mu_{2,1} = 0$ ist mit a, b auch $-a, b$ reduziert. Im Fall $\mu_{2,1} = \frac{1}{2}$ ist mit a, b auch $a, a - b$ reduziert. Im Fall $\|a\| = \|b\|$ ist mit a, b auch b, a reduziert. In den übrigen Fällen gibt es nur die reduzierte Basis $\pm a, \pm b$.

Satz 4.1.2
Für eine reduzierte Basis $a, b \in \mathbb{R}^n$ sind $\|a\|$ und $\|b\|$ die beiden sukzessiven Minima des Gitters $L = \mathbb{Z}a + \mathbb{Z}b$.

Beweis. O.B.d.A. sei $\|a\| \leq \|b\|$. Die Behauptung lautet:

$$\begin{aligned} &\|a\| \leq \|ra + sb\| && \forall (r,s) \in \mathbb{Z}^2 \setminus \{(0,0)\} \\ &\|b\| \leq \|ra + sb\| && \forall r \in \mathbb{Z} \text{ und } s \in \mathbb{Z} \setminus \{0\} \end{aligned}$$

Diese Ungleichungen folgen in Verbindung mit der Reduktionsbedingung $\|b\| \leq \|a \pm b\|$ aus:

$$\begin{aligned} &\|a\| \leq \|b\| && \\ &\|a\| \leq \|ra\| && \forall r \in \mathbb{Z} \setminus \{0\} \\ &\|b\| \leq \|\xi a + \eta b\| && \forall \xi,\eta \in \mathbb{R} \text{ mit } |\xi|, |\eta| \geq 1 \end{aligned} \tag{4.1}$$

Wir zeigen die Ungleichungen (4.1). Betrachten wir Grafik 4.1.2: Die Norm nimmt ihr Minimum in jedem der vier gepunkteten Bereiche in den Punkten $\pm a \pm b$ an. Auf jeder der dicken Linien liegen die Gitterpunkte, von denen der mittlere Punkt die minimale Norm hat, d.h.:

$$\begin{array}{ccccc} \|\pm a - b\| & \geq & \|\pm a\| & \leq & \|\pm a + b\| \\ \|-a \pm b\| & \geq & \|\pm b\| & \leq & \|a \pm a\| \end{array}$$

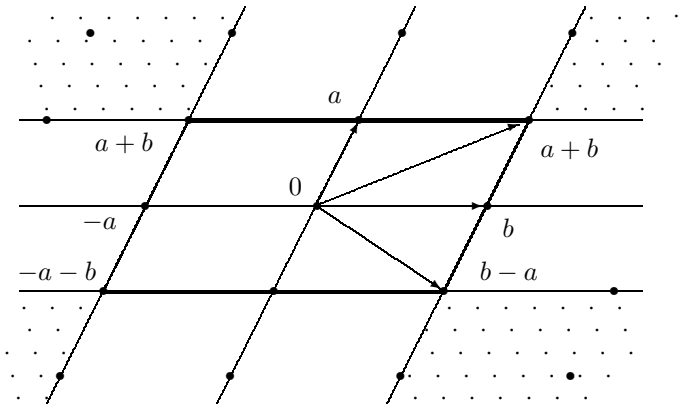

Abbildung 4.1.2: Reduzierte Basis a, b

Die Konvexität der Norm sichert für $|\xi| \geq 1$:

$$\begin{array}{lllll} \|\pm a \pm \xi b\| & \geq & \|\pm a \pm b\| & \geq & \|\pm a\| \\ \|\pm \xi a \pm b\| & \geq & \|\pm a \pm b\| & \geq & \|\pm b\| \end{array}$$

Damit haben die Punkte $\pm a \pm b$ minimale Norm auf der dicken Linie. Wegen der Konvexität nimmt die Norm ihr Minimum in den gepunkteten Bereichen am Rand an, also auf den dicken Linien. ■

Definition 4.1.3 (Wohlgeordnete, reduzierte Basis)
Eine geordnete Gitterbasis $a, b \in \mathbb{R}^n$ heißt wohlgeordnet reduziert bezüglich der Norm $\|\cdot\|$, wenn:

$$\|a\| \leq \|a - b\| \leq \|b\|$$

4.2 Algorithmen

Wir lernen Algorithmen zur Reduktiom zweidimensionaler Gitter kennen. Wir stellen ein Verfahren für allgemeine Normen und für den Spezialfall der Euklidischen Norm vor.

4.2.1 Reduktionsverfahren für Euklidische Norm

Der Algorithmus 4.2.1 reduziert eine Basis bezüglich der Euklidischen Norm. Eine Iteration des Algorithmus' 4.2.1 reduziert b gemäß $b := b - \lceil \mu_{2,1} \rfloor a$ und vertauscht anschließend a und b. Offenbar ist die Ausgabe korrekt.

Satz 4.2.1
Bei Eingabe von a, b mit $\|a\| \leq \|b\|$ endet Algorithmus 4.2.1 nach höchstens

$$\left\lceil \log_{1+\sqrt{2}} \left(\frac{\|a\|}{\lambda_2} \right) \right\rceil + 3$$

vielen Iterationen.

Beweis. Siehe Satz 4.4 von [Schnorr94b]. ■

Algorithmus 4.2.1 Gauß'sches Reduktionsverfahren für Euklidische Norm

EINGABE: Gitterbasis $a, b \in \mathbb{R}^n$ mit $\|a\| \leq \|b\|$

1. WHILE $|\mu_{2,1}| > \frac{1}{2}$ DO

1.1. $[a, b] := [a, b] \cdot \begin{bmatrix} -\lceil \mu_{2,1} \rfloor & 1 \\ 1 & 0 \end{bmatrix}$

1.2. IF $\|a\| > \|b\|$ THEN vertausche a und b

END while

2. $b := b \cdot \text{sign}(\mu_{2,1})$ /* wir erreichen $\mu_{2,1} \geq 0$ */

AUSGABE: Reduzierte Basis a, b

Algorithmus 4.2.2 Gauß'sches Reduktionsverfahren für beliebige Norm

EINGABE: Gitterbasis $a, b \in \mathbb{R}^n$ mit $\|a\| \leq \|b\|$

1. WHILE $\|b\| > \|a - b\|$ DO

1.1. $b := b - \mu a$, $\mu \in \mathbb{Z}$ so gewählt, daß $\|b - \mu a\|$ minimal

1.2. IF $\|a + b\| < \|a - b\|$ THEN $b := -b$

1.3. Vertausche a und b

END while

AUSGABE: Reduzierte Basis a, b

4.2.2 Reduktionsverfahren für beliebige Norm

Das Gauß'sche Reduktionsverfahren wurde auf beliebige Normen verallgemeinert (Algorithmus 4.2.2). Eine ausführliche Analyse findet sich in der Originalarbeit [KaSchn96] von M. Kaib und C.P. Schnorr sowie in M. Kaibs Dissertation [Kaib94]. Dort werden effiziente Algorithmen für den Schritt 1.1 in der l_1- und sup-Norm vorgestellt.

Kapitel 5

LLL-reduzierte Gitterbasen

Der folgende Reduktionsbegriff für geordnete Gitterbasen $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ beliebigen Ranges n wurde 1982 von A.K. Lenstra, H.W. Lenstra und L. Lovász [LLL82] vorgeschlagen. Er bezieht sich auf die Euklidische Norm.

5.1 Definition und Eigenschaften

Wir führen den Begriff der LLL-reduzierten Basis ein und beweisen Eigenschaften einer LLL-reduzierten Basis, inbesondere wie gut die Länge des ersten Basisvektors das erste sukzessive Minimum des Gitters approximiert. Seien $\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n$ das der Basis $b_1, b_2, \ldots, b_n$ zugeordnete Orthogonalsystem und $\mu_{i,j}$ $(1 \leq i, j \leq n)$ die Gram-Schmidt-Koeffizienten.

Definition 5.1.1 (LLL-reduzierte Basis)
Eine geordnete Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ heißt LLL-reduziert (L^3-reduziert) zum Parameter δ mit $\frac{1}{4} < \delta \leq 1$, wenn:

a) $|\mu_{i,j}| \leq \frac{1}{2}$ *für* $1 \leq j < i \leq n$

b) $\delta \cdot \|\widehat{b}_{k-1}\|^2 \leq \|\widehat{b}_k\|^2 + \mu_{k,k-1}^2 \cdot \|\widehat{b}_{k-1}\|^2$ *für* $k = 2, 3, \ldots, n$

Die erste Eigenschaft ist das Kriterium der Längenreduktion (vergleiche Definition 2.3.1 auf Seite 30). Der Parameter δ beeinflußt die Güte der reduzierten Basis: Je größer der Wert, desto stärker ist die Basis reduziert. A.K. Lenstra, H.W. Lenstra und L. Lovász [LLL82] haben die LLL-Reduktion ursprünglich nur für den Parameter $\delta = \frac{3}{4}$ definiert. Mit der orthogonalen Projektion

$$\pi_k : \mathbb{R}^m \to \operatorname{span}(b_1, b_2, \ldots, b_{k-1})^{\perp}$$

kann man die zweite Bedingung wie folgt schreiben:

$$\delta \cdot \|\pi_{k-1}(b_{k-1})\|^2 \leq \|\pi_{k-1}(b_k)\|^2 \quad \text{für } k = 2, 3, \ldots, n$$

Falls die Basis aus zwei Vektoren besteht, erhalten wir mit $\delta = 1$ eine Gauß-reduzierte Basis. Es sei $T : \operatorname{span}(b_1, b_2, \ldots, b_n) \to \mathbb{R}^m$ die Isometrie mit

$$T(\widehat{b}_i) = \|\widehat{b}_i\| \cdot e_i \quad \text{für } i = 1, 2, \ldots, n,$$

wobei e_i der i-te Einheitsvektor im $\mathbb{R}^m$ sei. Die Basismatrix $[T(b_1), T(b_2), \ldots, T(b_m)]$ des zu L isometrischen Gitters $T(L)$ ist eine obere Dreiecksmatrix:

$$[T(b_1),\ldots,T(b_m)] = \begin{bmatrix} \|\widehat{b}_1\| & 0 & \cdots & 0 \\ 0 & \|\widehat{b}_2\| & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \|\widehat{b}_n\| \end{bmatrix} \begin{bmatrix} 1 & \mu_{2,1} & \mu_{3,1} & \cdots & \mu_{n,1} \\ 0 & 1 & \mu_{3,2} & & \mu_{n,2} \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mu_{n,n-1} \\ 0 & \ldots & \cdots & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \|\widehat{b}_1\| & * & & * & * & * \\ & \ddots & & * & * & * \\ & & \begin{bmatrix} \|\widehat{b}_{k-1}\| & \mu_{k,k-1}\|\widehat{b}_{k-1}\| \\ 0 & \|\widehat{b}_k\| \end{bmatrix} & & * & * \\ & & & & \ddots & * \\ & & & & & \|\widehat{b}_n\| \end{bmatrix}$$

Die Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ ist genau dann LLL-reduziert mit δ, wenn:

a) die Basis $b_1, b_2, \ldots, b_n$ längenreduziert ist;

b) wenn die 2×2-Matrizen auf der Diagonalen:

$$\begin{bmatrix} \|\widehat{b}_{k-1}\| & \mu_{k,k-1}\|\widehat{b}_{k-1}\| \\ 0 & \|\widehat{b}_k\| \end{bmatrix}$$

für $k = 2, 3, \ldots, n$ LLL-reduziert zum Parameter δ sind.

Sei $b_1, b_2, \ldots, b_n$ LLL-reduziert mit δ, dann ist $\pi_k(b_k), \pi_k(b_{k+1}), \ldots, \pi_k(b_j)$ LLL-reduziert mit δ für $1 \le k < j \le n$. Wir untersuchen im weiteren Eigenschaften LLL-reduzierter Gitterbasen.

Lemma 5.1.2
Sei $b_1, b_2, \ldots, b_n$ LLL-reduziert zum Parameter δ. Dann gilt für $\alpha = \frac{1}{\delta - \frac{1}{4}}$:

$$\|\widehat{b}_i\|^2 \le \alpha^{j-i} \cdot \|\widehat{b}_j\|^2 \quad \textit{für } 1 \le i \le j \le n$$

Speziell mit $\delta = \frac{3}{4}$, $\alpha = 2$ ist $\|b_1\|^2 \le 2^{j-1} \cdot \|\widehat{b}_j\|^2$, so daß die Längenquadrate für großes j nicht beliebig klein werden können.

Beweis. Aus den Eigenschaften a) und b) der LLL-Reduktion folgt:

$$\delta \cdot \|\widehat{b}_i\|^2 \overset{b)}{\le} \|\widehat{b}_{i+1}\|^2 + \mu_{i+1,i}^2 \|\widehat{b}_i\|^2 \overset{a)}{\le} \|\widehat{b}_{i+1}\|^2 + \tfrac{1}{4}\|\widehat{b}_i\|^2$$

und somit:

$$\underbrace{\left(\delta - \tfrac{1}{4}\right)}_{=1/\alpha} \cdot \|\widehat{b}_i\|^2 \le \|\widehat{b}_{i+1}\|^2$$

Die Behauptung folgt durch Induktion über $j - i$. ∎

Lemma 5.1.3
Sei $b_1, b_2, \ldots, b_n$ eine Basis des Gitters L. Dann gilt für $i = 1, 2, \ldots, n$:

$$\lambda_j \ge \min_{i=j,j+1,\ldots,n} \|\widehat{b}_i\|$$

Beweis. Es gibt linear unabhängige Vektoren $a_1, a_2, \ldots, a_n \in L$, so daß $\|a_j\| = \lambda_j(L)$ für $j = 1, 2, \ldots, n$. Sei

$$a_k = \sum_{i=1}^{n} t_{ik} b_i = \sum_{i=1}^{n} \overline{t}_{ik} \widehat{b}_i \quad \text{für } k = 1, 2, \ldots, n$$

Dabei sind die Koeffizienten t_{ik} ganzzahlig und die $\overline{t}_{ik}$ reell. Sei

$$\mu(k) := \max \{i \ : \ t_{ik} \neq 0\}$$

Da die Vektoren $b_1, b_2, \ldots, b_{\mu(k)}$ linear unabhängig sind, gilt $\overline{t}_{\mu(k),k} = t_{\mu(k),k} \in \mathbb{Z}$. Wegen der linearen Unabhängigkeit der Vektoren $a_1, a_2, \ldots, a_j$ gibt es zu jedem j ein $k \leq j$ mit $\mu(k) \geq j$. Denn aus der Annahme $\mu(k) < j$ für $k = 1, 2, \ldots, j$ folgt

$$a_1, a_2, \ldots, a_j \in \operatorname{span}(b_1, b_2, \ldots, b_{j-1}),$$

so daß $a_1, a_2, \ldots, a_j$ linear abhängig sind — Widerspruch. Wir erhalten:

$$\lambda_j^2 \geq \lambda_k^2 = \|a_k\|^2 \geq \overline{t}_{\mu(k),k}^2 \|\widehat{b}_{\mu(k)}\|^2 \geq \|\widehat{b}_{\mu(k)}\|^2 \geq \min_{i=j,j+1,\ldots,n} \|\widehat{b}_i\|^2$$

■

Während die untere Schranke zu λ_j in Lemma 5.1.3 für beliebige Basen gilt, zeigt folgender Satz, daß die Längen $\|b_j\|$ im Fall LLL-reduzierter Basen „grobe“ Approximationen der sukzessiven Minima λ_j sind.

Satz 5.1.4 (Lenstra, Lenstra, Lovász 1982)
Sei $b_1, b_2, \ldots, b_n$ mit Parameter δ LLL-reduzierte Gitterbasis. Dann gilt für $\alpha = \frac{1}{\delta - \frac{1}{4}}$:

a) $\alpha^{1-j} \leq \dfrac{\|\widehat{b}_j\|^2}{\lambda_j^2}$ *für* $j = 1, 2, \ldots, n$

b) $\dfrac{\|b_j\|^2}{\lambda_j^2} \leq \alpha^{n-1}$ *für* $j = 1, 2, \ldots, n$

c) $\|b_k\|^2 \leq \alpha^{j-1} \cdot \|\widehat{b}_j\|^2$ *für* $k \leq j$

Beweis. Wir zeigen zunächst die erste und dritte Aussage. Es gibt ein k mit $1 \leq k \leq j$ und $\lambda_j \leq \|b_k\|$. Es folgt:

$$\begin{aligned}
\lambda_j^2 &\leq \|b_k\|^2 \\
&\leq \|\widehat{b}_k\|^2 + \tfrac{1}{4} \sum_{i=1}^{k-1} \|\widehat{b}_i\|^2 && \text{(wegen beiden LLL-Eigenschaften)} \\
&\leq \|\widehat{b}_j\|^2 \left(\alpha^{j-k} + \tfrac{1}{4} \sum_{i=1}^{k-1} \alpha^{j-i} \right) && \text{(nach Lemma 5.1.2)} \\
&\leq \|\widehat{b}_j\|^2 \alpha^{j-1} \left(\alpha^{1-k} + \tfrac{1}{4} \sum_{i=1}^{k-1} \alpha^{1-i} \right)
\end{aligned}$$

Die obere Schranke für $\|b_k\|^2$ gilt für alle k und j mit $k \leq j$ (3. Aussage). Für die erste Aussage der Behauptung zeigen wir:

$$\alpha^{1-k} + \tfrac{1}{4} \sum_{i=1}^{k-1} \alpha^{1-i} \leq 1$$

Für $k = 1$ gilt offenbar die Ungleichung. Für $k \geq 2$ gilt wegen $\alpha^{-1} = \delta - \frac{1}{4} \leq \frac{3}{4}$:

$$\alpha^{1-k} + \frac{1}{4} \underbrace{\sum_{i=1}^{k-1} \alpha^{1-i}}_{\text{geom. Reihe}} \leq \left(\frac{3}{4}\right)^{k-1} + \frac{1}{4} \cdot \frac{1 - \left(\frac{3}{4}\right)^{k-1}}{1 - \frac{3}{4}} = \frac{1}{4} \cdot \frac{1}{1 - \frac{3}{4}} = 1$$

Damit sind die erste und die dritte Behauptung gezeigt. Nach Lemma 5.1.3 gibt es ein $k \geq j$, so daß $\lambda_j \geq \|\widehat{b}_k\|$. Aus Lemma 5.1.2 folgt:

$$\begin{aligned}
\lambda_j^2 &\geq \|\widehat{b}_k\|^2 && \text{(wegen Lemma 5.1.3)} \\
&\geq \alpha^{-k+j} \cdot \|\widehat{b}_j\|^2 && \text{(wegen Lemma 5.1.2)} \\
&\geq \alpha^{-k+1} \cdot \|b_j\|^2 && \text{(wegen 3. Aussage des Satzes mit } k = j) \\
&\geq \alpha^{-n+1} \cdot \|b_j\|^2 && \text{(wegen } k \leq n \text{ und } \alpha \geq 1)
\end{aligned}$$

■

Wir folgern:

Korollar 5.1.5
Sei $b_1, b_2, \ldots, b_n$ mit Parameter δ LLL-reduzierte Basis des Gitters L. Dann gilt für $\alpha = \frac{1}{\delta - \frac{1}{4}}$:

a) $\|b_1\|^2 \leq \alpha^{\frac{n-1}{2}} (\det L)^{\frac{2}{n}}$

b) $\prod_{i=1}^{n} \|b_i\|^2 \leq \alpha^{\binom{n}{2}} (\det L)^2$

Beweis. Es ist $\prod_{i=1}^{n} \|\widehat{b}_i\|^2 = (\det L)^2$. Wegen der dritten Aussage des Satzes 5.1.4 gilt:

$$\|b_1\|^2 \leq \|\widehat{b}_i\|^2 \cdot \alpha^{i-1}$$

Es folgt:

$$\|b_1\|^{2n} \leq \alpha^1 \alpha^2 \cdots \alpha^{n-1} \prod_{i=1}^{n} \|\widehat{b}_i\|^2 = \alpha^{\binom{n}{2}} \cdot (\det L)^2$$

Somit gilt die erste Aussage: $\|b_1\|^2 \leq \alpha^{\frac{n-1}{2}} (\det L)^{\frac{2}{n}}$. Die zweite Aussage folgt aus $\prod_{i=1}^{n} \|\widehat{b}_i\|^2 = (\det L)^2$ und $\|b_i\|^2 \leq \|\widehat{b}_i\|^2 \alpha^{i-1}$, der dritten Aussage von Satz 5.1.4. ■

5.2 Lovász-Verfahren zur LLL-Reduktion

Wir geben ein Verfahren an, um eine Gitterbasis in eine LLL-reduzierte Basis desselben Gitters zu überführen. Wir analysieren die Laufzeit und die Größe der Koeffizienten während der Berechnung. Wir verallgemeinern das Verfahren auf Erzeugendensysteme, deren Vektoren nicht linear unabhängig sein müssen.

5.2.1 Algorithmus

Das Lovász-Verfahren, Algorithmus 5.2.1, transformiert zu gegebenem δ, $\frac{1}{4} < \delta < 1$, eine ganzzahlige Gitterbasis in eine mit Parameter δ LLL-reduzierte Basis desselben Gitters. Bei jedem

Algorithmus 5.2.1 Lovász-Verfahren zur LLL-Reduktion

EINGABE: ▷ Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$
▷ Parameter δ mit $\frac{1}{4} < \delta < 1$

1. $k := 2$ /∗ k ist die Stufe ∗/

2. Berechne $\mu_{i,j}$ für $1 \le j < i \le n$ und $\|\widehat{b}_i\|^2$ für $i = 1, 2, \ldots, n$

3. WHILE $k \le n$ DO

/∗ Invariante: $b_1, b_2, \ldots, b_{k-1}$ ist LLL-reduziert ∗/

3.1. Längenreduziere b_k und korrigiere $\mu_{k,j}$ für $j = 1, 2, \ldots, k-1$

3.2. IF $\delta \cdot \|\widehat{b}_{k-1}\|^2 > \|\widehat{b}_k\|^2 + \mu_{k,k-1}^2 \|\widehat{b}_{k-1}\|^2$ THEN

3.2.1. $b_{k-1} \leftrightarrow b_k$, d.h. vertausche b_{k-1} und b_k

3.2.2. $k := \max(k-1, 2)$

ELSE $k := k + 1$

END while

AUSGABE: Mit δ LLL-reduzierte Basis $b_1, b_2, \ldots, b_n$

Austausch $b_{k-1} \leftrightarrow b_k$ müssen die Größen $\|\widehat{b}_k\|^2$, $\|\widehat{b}_{k-1}\|^2$ und $\mu_{i,\nu}$, $\mu_{\nu,i}$ für $\nu = k-1, k$ und $i = 1, 2, \ldots, n$ neu berechnet werden (siehe Beweis Lemma 5.2.3). Die Korrektheit folgt aus der Invariante: Bei Eintritt in Stufe k ist die Basis $b_1, b_2, \ldots, b_{k-1}$ LLL-reduziert mit δ. Am Ende ist $k = n + 1$ und die gesamte Basis $b_1, b_2, \ldots, b_n$ ist reduziert.

Wir analysieren die Laufzeit des Lovász-Verfahrens, Algorithmus 5.2.1, zur LLL-Reduktion. Wir betrachten die Determinanten von Teilgittern:

$$D_i := \det L(b_1, b_2, \ldots, b_i)^2 = \det\left[\langle b_s, b_t\rangle\right]_{1 \le s,t \le i} = \prod_{j=1}^{i} \|\widehat{b}_j\|^2 \tag{5.1}$$

Wir arbeiten mit den Quadraten der Gitterdeterminanten, weil diese für ganzzahlige Basisvektoren ebenfalls ganzzahlig sind. Wir setzen:

$$D := \prod_{j=1}^{n-1} D_j \tag{5.2}$$

Lemma 5.2.1
Für ganzzahlige Eingaben $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$ stoppt das Lovász-Verfahren nach maximal

$$\left\lfloor \log_{1/\delta}\left(D^{Start}\right) \right\rfloor$$

vielen Austauschen $b_{k-1} \leftrightarrow b_k$. Für $M := \max_i \left\|b_i^{Start}\right\|^2$ gilt:

$$\#Austausche \le \binom{n}{2} \log_{1/\delta} M$$

Beweis. Für $j = 1, 2, \ldots, n$ ist D_j stets (d.h. im Lauf des Verfahrens) ganzzahlig und positiv. Wir zeigen, daß jeder Austausch $D^{\text{neu}} \le \delta \cdot D^{\text{alt}}$ bewirkt. Wegen $D^{\text{Ende}} \in \mathbb{N}$ impliziert dies:

$$D^{\text{Start}} \ge D^{\text{Ende}} \cdot \left(\frac{1}{\delta}\right)^{\#\text{Iterationen}} \ge \left(\frac{1}{\delta}\right)^{\#\text{Iterationen}} \tag{5.3}$$

Und wir erhalten den ersten Teil der Behauptung. Die Gitter $L(b_1, b_2, \ldots, b_j)$ mit $j \neq k-1$ werden beim Austausch $b_{k-1} \leftrightarrow b_k$ nicht verändert. Also bleiben die Determinanten D_j mit $j \neq k-1$ erhalten. Aufgrund der Bedingung wird nur dann ausgetauscht, falls:

$$\delta \cdot \|\widehat{b}_{k-1}^{\text{alt}}\|^2 \;>\; \|\widehat{b}_k^{\text{alt}}\|^2 + \underbrace{\left(\mu_{k,k-1}^{\text{alt}}\right)^2 \cdot \|\widehat{b}_{k-1}^{\text{alt}}\|^2}_{=\|\widehat{b}_{k-1}^{\text{neu}}\|^2} \;\geq\; \|\widehat{b}_{k-1}^{\text{neu}}\|^2$$

Wegen $D_{k-1} = \prod_{i=1}^{k-1} \|\widehat{b}_i\|^2$ bewirkt der Austausch $b_{k-1} \leftrightarrow b_k$, daß:

$$D_{k-1}^{\text{neu}} \leq \delta \cdot D_{k-1}^{\text{alt}}$$

Es folgt aus (5.2), daß $D^{\text{neu}} \leq \delta \cdot D^{\text{alt}}$. Wir erhalten die erste Behauptung aus (5.3). Die zweite Behauptung folgt aus $D_i^{\text{Start}} \leq M^i$ für $i = 1, 2, \ldots, n-1$ und $D^{\text{Start}} \leq M^{\binom{n}{2}}$. ■

Welche Laufzeit hat der Algorithmus für reelle Gitterbasen? Für $\delta \leq 1$ stoppt das Lovász-Verfahren, allerdings weiß man bisher nur für $\delta < 1$, daß die Laufzeit polynomiell ist.

Lemma 5.2.2
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine reelle Eingabebasis und $M := \max_i \|\widehat{b}_i\|^2$. Dann stoppt für $\delta < 1$ das Lovász-Verfahren nach höchstens

$$\log_{1/\delta}\left[\prod_{j=1}^{n-1}\left(\frac{\|\widehat{b}_j\|}{\lambda_j}\right)^{2(n-j)} \cdot (\gamma_j)^j\right] \leq \binom{n}{2}\log_{1/\delta}\left(\frac{M}{\lambda_1^2}\right) + \binom{n}{2}\log_{1/\delta} n$$

vielen Austauschen $b_{k-1} \leftrightarrow b_k$ (γ_j ist die Hermite-Konstante der Dimension j)

Beweis. Es gilt

$$D^{\text{Start}} = \prod_{j=1}^{n-1} \|\widehat{b}_j\|^{2(n-j)} \leq M^{\binom{n}{2}}$$

und $D^{\text{Ende}} = \prod_{j=1}^{n-1} D_j$, wobei wir D_j durch Minkowskis Ungleichung nach unten abschätzen durch:

$$D_j = \prod_{i=1}^{j} \|\widehat{b}_i\|^2 \geq (\gamma_j)^{-j} \prod_{i=1}^{j} \lambda_i^2$$

Mit $M = \max_i \|\widehat{b}_i\|^2$ und der einfachen Abschätzung $\gamma_j \leq j$ aus der Bemerkung 3.2.2 auf Seite 41 folgt:

$$\begin{aligned}
\#\text{ Austausche} &\leq \log_{1/\delta}\left(\frac{D^{\text{Start}}}{D^{\text{Ende}}}\right) \\
&\leq \log_{1/\delta}\left(\prod_{j=1}^{n-1}\left(\frac{\|\widehat{b}_j\|}{\lambda_j}\right)^{2(n-j)} \cdot (\gamma_j)^j\right) \\
&\leq \log_{1/\delta}\left(\prod_{j=1}^{n-1}\left(\frac{M}{\lambda_1^2}\right)^{n-j}\right) + \log_{1/\delta}\left(\prod_{j=1}^{n-1} n^j\right)
\end{aligned}$$

Wir erhalten:

$$\begin{aligned} \#\text{ Austausche} &\leq \binom{n}{2} \log_{1/\delta}\left(\frac{M}{\lambda_1^2}\right) + \sum_{j=1}^{n-1} \log_{1/\delta} n^j \\ &\leq \binom{n}{2} \log_{1/\delta}\left(\frac{M}{\lambda_1^2}\right) + \binom{n}{2} \log_{1/\delta} n \end{aligned}$$

■

Nachdem wir die Anzahl der Austausch $b_{k-1} \leftrightarrow b_k$ im im Lovász-Verfahren analysiert haben, wollen wir im folgenden die Anzahl der arithmetischen Schritte für einen Austausch im Lovász-Verfahren abschätzen:

Lemma 5.2.3
Der Austausch $b_{k-1} \leftrightarrow b_k$ bewirkt mit $\mu := \mu_{k,k-1}$ und $\mu^{\text{neu}} := \mu_{k,k-1}^{\text{neu}}$, daß:

a) $\mu_{\text{neu}} = \mu \cdot \dfrac{\|\widehat{b}_{k-1}\|^2}{\|\widehat{b}_{k-1}^{\text{neu}}\|^2}$

b) $\left[\mu_{k,i}^{\text{neu}}, \quad \mu_{k-1,i}^{\text{neu}}\right] = \left[\mu_{k-1,i}, \quad \mu_{k,i}\right]$ *für* $i = 1, 2, \ldots, k-2$.

Beweis. Wir beweisen die vier Behauptungen einzeln (beachte: $\widehat{b}_k$ und $\widehat{b}_{k-1}$ stehen senkrecht aufeinander):

a) Wegen

$$\mu_{\text{neu}} = \frac{\left\langle b_k^{\text{neu}}, \widehat{b}_{k-1}^{\text{neu}} \right\rangle}{\|\widehat{b}_{k-1}^{\text{neu}}\|^2} = \frac{\left\langle b_{k-1}, \widehat{b}_k + \mu\widehat{b}_{k-1} \right\rangle}{\|\widehat{b}_{k-1}^{\text{neu}}\|^2}$$

erhalten wir:

$$\mu_{\text{neu}} = \frac{\left\langle b_{k-1}, \widehat{b}_k \right\rangle + \left\langle b_{k-1}, \mu\widehat{b}_{k-1} \right\rangle}{\|\widehat{b}_{k-1}^{\text{neu}}\|^2} = \frac{\mu\|\widehat{b}_{k-1}\|^2}{\|\widehat{b}_{k-1}^{\text{neu}}\|^2}$$

b) offensichtlich.

■

Wir schätzen die Anzahl der arithmetischen Schritte für einen Austausch im Lovász-Verfahren nach oben ab durch:

Satz 5.2.4
Ein Austausch $b_{k-1} \leftrightarrow b_k$ geht in $\mathcal{O}(k)$ arithmetischen Schritten. Die Längenreduktion von b_k gelingt in $\mathcal{O}(nk)$ arithmetischen Schritten.

Beweis. Die erste Aussage folgt aus Lemma 5.2.3. Die zweite Behauptung berücksichtigt, daß der Schritt $b_k = b_k - \mu b_j$ die Gram-Schmidt-Koeffizienten wie folgt verändert:

$$\mu_{k,i} := \mu_{k,i} - \mu \cdot \mu_{j,i} \qquad \text{für } i = 1, 2, \ldots, j$$

■

Wir haben die die Anzahl der arithmetischen Schritte im Lovász-Verfahren analysiert. Wie groß können aber die während des Lovász-Verfahrens auftretenden Zahlen werden? Wir untersuchen diese Fragestellung und versuchen, die Größe der Koeffizienten während der Berechnung nach oben abzuschätzen.

Lemma 5.2.5
Für eine ganzzahlige Eingabebasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$ gilt:

a) $D_{i-1} \cdot \widehat{b}_i \in \mathbb{Z}^m$

b) $D_j \cdot \mu_{i,j} \in \mathbb{Z}$

Dabei ist $D_j = \det L(b_1, b_2, \ldots, b_j)^2 = \prod_{i=1}^{j} \|\widehat{b}_i\|^2$ *ganzzahlig.*

Beweis. Wir werden die zweite Aussage aus der ersten folgern.

a) Aus $[b_1, b_2, \ldots, b_n] = \left[\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n\right] \cdot [\mu_{i,j}]^{\mathsf{T}}$ folgt für $[\nu_{ij}] := [\mu_{i,j}]^{-1}$:

$$\left[\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n\right] \;=\; [b_1, b_2, \ldots, b_n] \cdot [\nu_{ij}]^{\mathsf{T}}$$

Dabei ist $[\nu_{ij}]^{\mathsf{T}}$ wie $[\mu_{i,j}]$ eine obere Dreiecksmatrix mit Einsen auf der Diagonalen. Wegen $\left\langle \widehat{b}_i, b_j \right\rangle = 0$ für $j = 1, 2, \ldots, i-1$ folgt aus $\widehat{b}_i = b_i + \sum_{t=1}^{i-1} \nu_{it} b_t$ und $\nu_{ii} = 1$:

$$-\langle b_i, b_j \rangle = \sum_{t=1}^{i-1} \nu_{it} \langle b_t, b_j \rangle \qquad\qquad \text{für } j = 1, 2, \ldots, i-1$$

Diese $i-1$ Gleichungen definieren $\nu_{i1}, \nu_{i2}, \ldots, \nu_{i,i-1}$. Die Determinante des Gleichungssystems ist:

$$D_{i-1} = \det\left[\langle b_j, b_k \rangle\right]_{1 \le j,k \le i-1}$$

Da $D_{i-1} \neq 0$, folgt aus der Cramer'schen Regel:

$$D_{i-1}\nu_{ij} \in \mathbb{Z} \qquad \text{für } j = 1, 2, \ldots, i-1$$

Da $\widehat{b}_i = b_i + \sum_{j=1}^{i-1} \nu_{ij} b_j$ und $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$, folgt die Behauptung $D_{i-1} \cdot \widehat{b}_i \in \mathbb{Z}^m$.

b) Nach Definition (5.1) der Determinanten $D_j = \prod_{s=1}^{j} \|\widehat{b}_s\|^2$ gilt:

$$D_j \cdot \mu_{i,j} = D_j \cdot \frac{\left\langle b_i, \widehat{b}_j \right\rangle}{\|\widehat{b}_j\|^2} = D_{j-1} \cdot \left\langle b_i, \widehat{b}_j \right\rangle = \left\langle b_i, D_{j-1} \cdot \widehat{b}_j \right\rangle$$

Aus der ersten Aussage, $\widehat{b}_j \cdot D_{j-1} \in \mathbb{Z}^m$, folgt $\left\langle b_i, D_{j-1} \cdot \widehat{b}_j \right\rangle \in \mathbb{Z}$. Wir erhalten die Behauptung $D_j \cdot \mu_{i,j} \in \mathbb{Z}$.

■

Die Größe $\max_i \|\widehat{b}_i\|^2$ wächst nicht und die Größe $\min_i \|\widehat{b}_i\|^2$ fällt nicht im Laufe des Lovász-Verfahrens. Die erste Aussage gilt, weil für jeden Austausch $b_{k-1} \leftrightarrow b_k$ gilt:

a) $\|\widehat{b}_{k-1}^{\text{neu}}\|^2 \le \delta \cdot \|\widehat{b}_{k-1}^{\text{alt}}\|^2$

b) $\|\widehat{b}_k^{\text{neu}}\|^2 \leq \|\widehat{b}_{k-1}^{\text{alt}}\|^2$

Die zweite Aussage folgt aus:

a) $\|\widehat{b}_{k-1}^{\text{neu}}\|^2 \geq \|\widehat{b}_k^{\text{alt}}\|^2$

b) $\|\widehat{b}_k^{\text{neu}}\|^2 \geq \delta^{-1} \cdot \|\widehat{b}_k^{\text{alt}}\|^2$

Wir geben Schranken für die Zähler der im Lovász-Verfahren auftretenden, rationalen Zahlen $\mu_{i,j}$ an. Zur ganzzahligen Eingabebasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$ sei im folgenden

$$M := \max_{i=1,2,\ldots,n} \|b_i\|^2$$

Lemma 5.2.6
Im Lovász-Verfahren gilt bei Eintritt in Stufe k mit $\alpha := \frac{1}{\delta - \frac{1}{4}}$ für $i = 1, 2, \ldots, n$:

a) $\|b_i\|^2 \leq \dfrac{i+3}{4} \cdot M$

b) $|\mu_{i,j}|^2 \leq \dfrac{i+3}{4} \cdot M \cdot \alpha^{j-1}$ *für* $j < k$

Beweis. Wir zeigen beide Behauptungen:

1. Da b_i längenreduziert ist, gilt für $i < k$:

$$\|b_i\|^2 = \sum_{j=1}^{i} \mu_{i,j}^2 \|\widehat{b}_j\|^2 \leq \left\|\widehat{b}_i\right\|^2 + \frac{i-1}{4} \max_{j=1,2,\ldots,i-1} \|\widehat{b}_j\|^2$$

   Mit $M = \max_{i=1,2,\ldots,n} \|b_i\|^2$ folgt:

$$\|b_i\|^2 \leq M + \frac{i-1}{4} \cdot M = \frac{i+3}{4} \cdot M$$

   Für $i \geq k$ schließt man durch Induktion über die Iterationen, daß die Ungleichung $\|b_i\|^2 \leq \frac{i+3}{4} M$ erhalten bleibt:

   - Für $k = i$ gilt die Ungleichung, da b_{k-1}, b_k beim Austausch $b_{k-1} \leftrightarrow b_k$ längenreduziert sind.
   - Für $i > k$ gilt die Ungleichung nach Induktionsannahme, da der Vektor b_i nicht verändert wird.

2. Allgemein gilt nach Definition der Gram-Schmidt-Koeffizienten und der Cauchy-Schwarz-Ungleichung:

$$|\mu_{i,j}|^2 = \frac{\left|\left\langle b_i, \widehat{b}_j \right\rangle\right|^2}{\|\widehat{b}_j\|^4} \leq \frac{\|b_i\|^2 \cdot \|\widehat{b}_j\|^2}{\|\widehat{b}_j\|^4} \leq \frac{\|b_i\|^2}{\|\widehat{b}_j\|^2}$$

   Aus der ersten Aussage, Lemma 5.1.2 ($b_1, b_2, \ldots, b_{k-1}$ ist LLL-reduziert), und $b_1 \in \mathbb{Z}^m$ erhalten wir:

$$\begin{aligned}
|\mu_{i,j}|^2 &\leq \frac{i+3}{4} \cdot M \cdot \|\widehat{b}_j\|^{-2} && \text{(wegen 1. Aussage: } \|b_i\|^2 \leq \tfrac{n+3}{4} M\text{)} \\
&\leq \frac{i+3}{4} \cdot M \cdot \alpha^{j-1} \cdot \|\widehat{b}_1\|^{-2} && \text{(wegen Lemma 5.1.2)} \\
&\leq \frac{i+3}{4} \cdot M \cdot \alpha^{j-1} && \text{(wegen } \|b_1\| = \|\widehat{b}_1\| \in \mathbb{Z}\text{)}
\end{aligned}$$

■

Lemma 5.2.7
Im Verlauf von Stufe k gilt stets für $j = 1, 2, \ldots, k-1$:

$$|\mu_{k,j}|^2 \leq \frac{k+3}{4} \cdot M \cdot \left(\frac{9\alpha}{4}\right)^{k-1}$$

Beweis. Auf Stufe k bewirkt der Reduktionsschritt

$$b_k := b_k - \lceil \mu_{k,i} \rfloor \cdot b_i$$

der Längenreduktion, daß für $j = 1, 2, \ldots, k-1$ gilt:

$$\mu_{k,j} := \mu_{k,j} - \lceil \mu_{k,i} \rfloor \underbrace{\mu_{i,j}}_{|\mu_{i,j}| \leq 1/2} \tag{5.4}$$

Jeder der $k-1$ Schritte (5.4) verändert $M_k := \max_{j=1,2,\ldots,k-1} |\mu_{k,j}|$ so, daß wir wegen $\lceil \mu_{k,i} \rfloor \leq M_k + \frac{1}{2}$ das neue M_k nach oben abschätzen können durch:

$$M_k^{\text{neu}} \leq M_k^{\text{alt}} + \tfrac{1}{2}\left(M_k^{\text{alt}} + \tfrac{1}{2}\right) \leq \tfrac{3}{2} \cdot M_k^{\text{alt}} + \tfrac{1}{4} \tag{5.5}$$

Nach Lemma 5.2.6 gilt bei Eintritt in die Stufe k, daß

$$M_k \leq \sqrt{\frac{k+3}{4} \cdot M \cdot \alpha^{k-1}}$$

Wegen Abschätzung (5.5) erhöht sich die Größe M_k im Verlauf der Stufe k höchstens um den Faktor $\left(\frac{3}{2}\right)^{k-1}$ (der Summand $\frac{1}{4}$ kann vernachlässigt werden). Also:

$$|\mu_{k,j}|^2 \leq \left(\frac{3}{2}\right)^{2(k-1)} \cdot \left(\frac{k+3}{4} \cdot M \cdot \alpha^{k-1}\right) = \frac{k+3}{4} \cdot M \cdot \left(\frac{9\alpha}{4}\right)^{k-1}$$

Daraus folgt die Behauptung. ■

Beim Eintritt in die Stufe k können die Größen $\mu_{i,j}$ mit $j > k$ sehr groß sein und das Verfahren ist nicht mehr stabil. Für $j > k$ gilt anstelle der zweiten Aussage von Lemma 5.2.6 nur die Schranke

$$|\mu_{i,j}|^2 \leq \frac{n+3}{4} \cdot M^j,$$

denn mit der Cauchy-Schwarz-Ungleichung erhalten wir

$$|\mu_{i,j}|^2 \leq \frac{\left\langle b_i, \widehat{b}_j \right\rangle^2}{\|\widehat{b}_j\|^4} \leq \frac{\|b_i\|^2 \cdot \|\widehat{b}_j\|^2}{\|\widehat{b}_j\|^4}, = \frac{\|b_i\|^2}{\|\widehat{b}_j\|^2},$$

so daß wegen $D_j = \prod_{i=1}^{j} \|\widehat{b}_i\|^2$ und wegen Lemma 5.2.6 auf Seite 63 gilt:

$$|\mu_{i,j}|^2 \leq \|b_i\|^2 \frac{D_{j-1}}{D_j} \leq \frac{n+3}{4} \cdot M \cdot D_{j-1} \leq \frac{n+3}{4} \cdot M^j$$

Das Lovász-Verfahren mit iterativer Orthogonalisierung (Algorithmus 5.2.2) vermeidet die Größen $\mu_{i,j}$ mit $j > k$ und rechnet auf Stufe k nur mit den Größen $\mu_{i,j}$ mit $1 \leq j < i \leq k$. Die Formeln von Schritt 2 des Algorithmus' 5.2.2 sind geeignet, wenn $\mu_{i,j}$ und $\|\widehat{b}_i\|^2$ Gleitkommazahlen sind. Weil die Basis $b_1, b_2, \ldots, b_{k-1}$ schon LLL-reduziert ist, gilt nach Lemma 5.1.2 auf Seite 56 und $\widehat{b_1} = b_1$:

$$\|\widehat{b}_j\|^2 \geq \|b_1\|^2 \alpha^{1-j} \qquad \text{für } j = 1, 2, \ldots, k$$

Die Divisoren $c_j = \|\widehat{b}_j\|^2$ bei der Berechnung von $\mu_{k,j}$ in Schritt 2 sind daher nicht beliebig klein. Dies ist wichtig für die Begrenzung von Gleitkommafehlern.

Algorithmus 5.2.2 Lovász-Verfahren mit iterativer Orthogonalisierung

EINGABE: ▷ Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$
▷ δ mit $\frac{1}{4} < \delta < 1$

1. $c_1 := \|b_1\|^2$, $k := 2$ /∗ k ist die Stufe ∗/

/∗ Bei Eintritt in die Stufe k liegen vor:

- $\mu_{i,j}$ für $1 \leq j < i < k$
- $c_i = \|\widehat{b}_i\|^2$ für $1 \leq i < k$

∗/

2. WHILE $k \leq n$ DO

2.1. IF $k = 2$ THEN $c_1 := \|b_1\|^2$

2.2. FOR $j = 1, 2, \ldots, k-1$ DO

$$\mu_{k,j} := \frac{\langle b_k, b_j \rangle - \sum_{i=1}^{j-1} \mu_{j,i}\mu_{k,i}c_i}{c_j}$$

END for

2.3. $c_k := \langle b_k, b_k \rangle - \sum_{j=1}^{k-1} \mu_{k,j}^2 c_j$

2.4. Längenreduziere b_k und aktualisiere $\mu_{k,1}, \mu_{k,2}, \ldots, \mu_{k,k-1}$

2.5. IF $\delta c_{k-1} \geq c_k + \mu_{k,k-1}^2 c_{k-1}$ THEN

2.5.1. $b_{k-1} \leftrightarrow b_k$, d.h. vertausche b_{k-1} und b_k

2.5.2. $k := \max(k-1, 2)$

ELSE $k := k + 1$

END while

AUSGABE: Mit δ LLL-reduzierte Basis $b_1, b_2, \ldots, b_n$

Satz 5.2.8
Bei ganzzahliger Eingabebasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$ mit $M := \max_{i=1,2,\ldots,n} \|b_i\|^2$ macht das Lovász-Verfahren mit iterativer Orthogonalisierung (Algorithmus 5.2.2)

$$\mathcal{O}\left(n^2 m \left(1 + n \log_{1/\delta} M\right)\right)$$

arithmetische Schritte durch auf den rationalen Zahlen $|\mu_{i,j}| \leq \sqrt{\frac{n+3}{4} \cdot M^j}$ und $|\mu_{k,j}| \leq \sqrt{\frac{n+3}{4} \cdot M \left(\frac{9\alpha}{4}\right)^{k-1}}$, $\|\widehat{b}_j\|^2 \leq M$, den Vektor b_i mit $\|b_i\|^2 \leq \frac{n+3}{4} \cdot M$. Die Zähler und Nenner dieser rationalen Zahlen sind absolut durch

$$M^{n-\frac{1}{2}} \sqrt{\frac{n+3}{4} \cdot \left(\frac{9\alpha}{4}\right)^n}$$

beschränkt.

Beweis. Nach Lemma 5.2.1 gilt:

$$\#\text{Austausche} \leq \binom{n}{2} \cdot \log_{1/\delta} M = \mathcal{O}\left(n^2 \log_2 M\right)$$

Weil die Stufe k zu Beginn des Algorithmus' 2 und am Ende $n+1$ ist, gilt offenbar:

$$\#\text{Iterationen} \leq n - 1 + 2 \cdot \#\text{Austausche}$$

Zu jeder Iteration mit Austausch und Stufenerniedrigung gibt es höchstens eine Iteration ohne Austausch, welche die Erniedrigung der Stufe kompensiert. Jede Iteration erfordert $\mathcal{O}(nm)$ arithmetische Schritte. Es folgt die behauptete Schranke für die Schrittzahl.

Nach Lemma 5.2.5 sind die Nenner der Zahlen μ_{ij} für $j < n$ durch $D_j \leq M^{n-1}$ beschränkt. Nach der zweiten Aussage des Lemmas 5.2.6 und nach Lemma 5.2.7 gilt:

$$|\mu_{i,j}|^2 \leq \frac{n+3}{4} \cdot M \cdot \left(\frac{9\alpha}{4}\right)^{n-1}$$

Damit sind die Zähler von $\mu_{i,j}$ absolut durch

$$\sqrt{\frac{n+3}{4}} \cdot M^{n-\frac{1}{2}} \cdot \left(\frac{9\alpha}{4}\right)^{\frac{n-1}{2}}$$

beschränkt. Zähler und Nenner von $\|\widehat{b}_j\|^2 = \frac{D_j}{D_{j-1}}$ sind durch M^j beschränkt. Die Koeffizienten der Vektoren b_i sind durch $\|b_i\|$ und somit nach Lemma 5.2.6 auf Seite 63 durch $\sqrt{nM}$ beschränkt. Damit sind alle im Verfahren auftretenden, ganzen Zahlen durch

$$\sqrt{\frac{n+3}{4}} M^{n-\frac{1}{2}} \left(\frac{9\alpha}{4}\right)^{\frac{n-1}{2}}$$

absolut beschränkt. ■

5.2.2 Lovász-Verfahren für linear abhängige Erzeugendensysteme

Wir haben das Lovász-Verfahren für Gitterbasen $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ beschrieben. In diesem Abschnitt wollen wir die L^3-Reduktion verallgemeinern auf den Fall linear abhängige Erzeugendensysteme. Seien $b_1, b_2, \ldots, b_n \in \mathbb{R}^m \setminus \{0\}$ beliebige Vektoren, so daß die Untergruppe

$$L(b_1, b_2, \ldots, b_n) = \left\{ \sum_{i=1}^{n} t_i b_i \;\middle|\; t_1, t_2, \ldots, t_n \in \mathbb{Z} \right\} \subseteq \mathbb{Z}^m$$

diskret, also ein Gitter, ist. Wir fordern im Vergleich zu Gitterbasen nicht, daß die Vektoren linear unhängig sind. Es sei

$$\widehat{b}_i := \pi_i(b_i) \in \operatorname{span}(b_1, b_2, \ldots, b_{i-1})^{\perp}$$

die orthogonale Projektion von b_i. Die Gram-Schmidt-Koeffizienten $\mu_{i,j}$ erweitern wir, daß die Identität

$$[b_1, b_2, \ldots, b_n] = \left[\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n\right] \cdot [\mu_{i,j}]_{1 \leq i,j \leq n}^{\mathsf{T}}$$

wie im Fall von Basisvektoren gilt. Falls $\widehat{b}_j = 0$, muß $\mu_{i,j} = \left\langle b_i, \widehat{b}_j \right\rangle \cdot \|\widehat{b}_j\|^{-2}$ sein und sonst können wir $\mu_{i,j}$ beliebig wählen. Wir setzen:

$$\mu_{i,j} := \begin{cases} \frac{\langle b_i, \widehat{b}_j \rangle}{\|\widehat{b}_j\|^2} & \text{falls } \widehat{b}_j \neq 0 \\ 0 & \text{sonst} \end{cases}$$

Algorithmus 5.2.3 Lovász-Verfahren für linear abhängige Erzeugendensysteme

EINGABE: ▷ Erzeugendensystem $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m \setminus \{0\}$
▷ δ mit $\frac{1}{4} < \delta < 1$

1. $c_1 := \|b_1\|^2$, $k := 2$ /∗ k ist die Stufe ∗/

/∗ Bei Eintritt in die Stufe k liegen vor:

- $\mu_{i,j}$ für $1 \leq j < i < k$
- $c_i = \|\widehat{b}_i\|^2$ für $1 \leq i < k$

∗/

2. WHILE $k \leq n$ DO

2.1. IF $k = 2$ THEN $c_1 := \|b_1\|^2$

2.2. FOR $j = 1, 2, \ldots, k-1$ DO

$$\mu_{k,j} := \frac{\langle b_k, b_j \rangle - \sum_{i=1}^{j-1} \mu_{j,i}\mu_{k,i}c_i}{c_j}$$

END for

2.3. $c_k := \langle b_k, b_k \rangle - \sum_{j=1}^{k-1} \mu_{k,j}^2 c_j$

2.4. Längenreduziere b_k und aktualisiere $\mu_{k,1}, \mu_{k,2}, \ldots, \mu_{k,k-1}$

2.5. IF $b_k \neq 0$ THEN

IF $\delta c_{k-1} \geq c_k + \mu_{k,k-1}^2 c_{k-1}$ THEN

$b_{k-1} \leftrightarrow b_k$, d.h. vertausche b_{k-1} und b_k

$k := \max(k-1, 2)$

ELSE $k := k + 1$

ELSE Entferne b_k aus der Menge der Vektoren und setze $n := n - 1$.

END if

END while

AUSGABE: Mit δ LLL-reduziertes Erzeugendensystem $b_1, b_2, \ldots, b_n$

Ist der Rang der Matrix $[b_1, b_2, \ldots, b_n]$ gleich r, sind genau $n - r$ der Vektoren $\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n$ gleich dem Nullvektor. Im Lovász-Verfahren mit iterativer Orthogonalisierung muß man auf Stufe k erkennen, ob $\widehat{b}_k = 0$ ist, da die Division durch $c_k = \|\widehat{b}_k\|^2$ bei der Berechnung der $\mu_{k+1,j}$ auf Stufe $k+1$ nicht durchgeführt werden darf. Es genügt, die im Lovász-Verfahren entstehenden Nullvektoren zu eliminieren. Der Nullvektor kann nur durch Längenreduktion von b_k auf Stufe k entstehen.

Wir wollen zunächst zeigen, daß $c_1, c_2, \ldots, c_{k-1}$ zu Beginn jeder Iteration der Schleife in Schritt 2 stets ungleich 0 sind und wir somit nie versuchen, durch 0 zu dividieren.

Lemma 5.2.9
Wenn auf Stufe k $\widehat{b}_k = 0$ ist, werden in Schritt 2.5 die Vektoren b_{k-1} und b_k vertauscht, sofern der Vektor b_k nicht im ELSE-Teil entfernt wird.

Beweis. Wegen $\widehat{b}_k = 0$ ist $b_k \in \operatorname{span}(b_1, b_2, \ldots, b_{k-1})$ und wir erhalten eine Darstellung des Vektors b_k:

$$b_k = \sum_{i=1}^{k-1} \mu_{k,i} \widehat{b}_i$$

Nach der Längenreduktion von b_k in Schritt 2.5 gilt für die Längenquadrate der Vektoren wegen $c_k = \|\widehat{b}_k\|^2 = 0$:

$$c_k + \mu_{k,k-1}^2 c_{k-1} = \mu_{k,k-1}^2 c_{k-1} \leq \tfrac{1}{4} c_{k-1}$$

Wegen $\delta \geq \frac{1}{4}$ gilt

$$\delta c_{k-1} \geq c_k + \mu_{k,k-1}^2 c_{k-1}$$

und wir vertauschen, sofern wir in den THEN-Teil gehen, die beiden Vektoren b_{k-1} und b_k. ■

Da die Vektoren $b_1, b_2, \ldots, b_n$ ungleich dem Nullvektor sind, gilt insbesondere $c_1 = \|\widehat{b}_1\|^2 \neq 0$. Wenn auf Stufe k das Längenquadrat $c_k = \|\widehat{b}_k\|^2 = 0$ ist, entfernen wir den Vektor und wir dekrementieren k bzw. setzen es auf 2. Induktiv erhalten wir, daß $c_1, c_2, \ldots, c_{k-1}$ zu Beginn jeder Iteration der Schleife in Schritt 2 stets ungleich 0 sind.

Satz 5.2.10
Sei $b_1, b_2, \ldots, b_n$ ein Erzeugendensystem aus ganzzahligen Vektoren, $r := \operatorname{Rang}([b_1, b_2, \ldots, b_n])$ und $M := \max_{i=1,2,\ldots,n} \|b_i\|^2$. Dann ist die Anzahl der Austausche $b_{k-1} \leftrightarrow b_k$ bzw. Iterationen von Algorithmus 5.2.3 beschränkt durch:

$$\left\lfloor \binom{r}{2} \log_{1/\delta} M \right\rfloor \leq n - 1 + n(n-1) \log_{1/\delta} M$$

Beweis. Die Determinantenquadrate

$$D_i := \det L(b_1, b_2, \ldots, b_i)^2 = \prod_{j=1}^{i} \|\widehat{b}_i\|^2 \qquad \text{für } i = 1, 2, \ldots, n$$

sind ganzzahlig. Wir setzen:

$$D := \prod_{\substack{i=1 \\ \widehat{b}_i \neq 0}}^{n} D_i$$

Dieser Wert ist ganzzahlig und ungleich 0. Für den Anfangswert gilt $D^{\text{Start}} \leq M^{\binom{r}{2}}$. Falls B eine Basis ist, stimmt dieser Wert mit dem aus Kapitel 5.2.1 überein. Wir zeigen, daß jeder Austausch $b_{k-1} \leftrightarrow b_k$ bewirkt:

$$D^{\text{neu}} \leq \begin{cases} \delta \cdot D^{\text{alt}} & \text{falls } \widehat{b}_{k-1}^{\text{neu}} \neq 0 \\ \frac{1}{4} \cdot D^{\text{alt}} & \text{sonst} \end{cases}$$

Für $i \neq k-1$ bleibt das Gitter $L(b_1, b_2, \ldots, b_i)$ unverändert und somit auch D_i beim Austausch $b_{k-1} \leftrightarrow b_k$. Wir unterscheiden die zwei Fälle:

- Im Fall $\widehat{b}_{k-1}^{\text{neu}} \neq 0$ gilt $\widehat{b}_{k-1}^{\text{alt}} \neq 0$ und wir erhalten aus

  $$\frac{D^{\text{neu}}}{D^{\text{alt}}} = \frac{D_{k-1}^{\text{neu}}}{D_{k-1}^{\text{alt}}} = \frac{\|\widehat{b}_{k-1}^{\text{neu}}\|^2}{\|\widehat{b}_{k-1}^{\text{alt}}\|^2} \leq \delta$$

  die Behauptung $D^{\text{neu}} \leq \delta D^{\text{alt}}$.

- Im Fall $\widehat{b}_{k-1}^{\text{neu}} = 0$ gilt $\widehat{b}_k^{\text{alt}} = 0$ und

$$\frac{D^{\text{neu}}}{D^{\text{alt}}} = \frac{D_{k-1}^{\text{neu}}}{D_{k-1}^{\text{alt}}}$$

Wegen $\widehat{b}_k^{\text{alt}} = 0$ hat das Gitter $L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{alt}}\right)$ denselben Rang wie das Gitter:

$$L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{alt}}, b_k^{\text{alt}}\right) = L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{neu}}, b_k^{\text{neu}}\right)$$

Da b_k längenreduziert und ungleich dem Nullvektor ist, muß das Gitter $L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{alt}}\right)$ echt enthalten sein in dem Gitter L^{neu}. Da beide Gitter den gleichen Rang haben, ist $L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{alt}}\right)$ ein echtes Untergitter von $L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{neu}}, b_k^{\text{neu}}\right)$ mit Index mindestens 2. Es folgt

$$\det L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{alt}}\right) \geq 2 \cdot \det L\left(b_1, \ldots, b_{k-2}, b_{k-1}^{\text{neu}}, b_k^{\text{neu}}\right)$$

und wir erhalten wegen $D_i = \det L(b_1, b_2, \ldots, b_i)^2$:

$$D_k^{\text{neu}} \leq \tfrac{1}{4} \cdot D_{k-1}^{\text{alt}}$$

Wegen $\delta \geq \frac{1}{4}$ folgt, daß jeder Austausch $b_{k-1} \leftrightarrow b_k$ bewirkt:

$$D_k^{\text{neu}} \leq \delta \cdot D_{k-1}^{\text{alt}}$$

Aus $D^{\text{Ende}} \geq 1$ erhalten wir die folgende Abschätzung für die Anzahl der Austausche:

$$\#\text{Austausche} \leq \log_{1/\delta} D^{\text{Start}}$$

Da die Anzahl der Austausche ganzzahlig ist, folgt:

$$\#\text{Austausche} \leq \left\lfloor \log_{1/\delta} D^{\text{Start}} \right\rfloor \leq \left\lfloor \binom{r}{2} \log_{1/\delta} M \right\rfloor \tag{5.6}$$

Jede Iteration ohne Austausch erhöht die Stufe k, sofern der Vektor b_k nicht der Nullvektor ist. Der Nullvektor b_k tritt genau $(n-r)$-mal auf und in diesen Iterationen wird die Stufe nicht verändert. Weil der Algorithmus mit Stufe $k = 2$ beginnt und mit $k = r + 1$ endet, folgt

$$\begin{aligned} \#\text{Iterationen} &\leq (r-1) + (n-r) + 2 \cdot \#\text{Austausche} \\ &\leq n - 1 + 2 \cdot \#\text{Austausche} \end{aligned}$$

und wir erhalten mit Abschätzung (5.6) die Behauptung:

$$\#\text{Iterationen} \leq n - 1 + 2 \cdot \left\lfloor \binom{r}{2} \log_{1/\delta} M \right\rfloor$$

■

Die Schranke aus Satz 5.2.10 ist scharf: Wenn alle Eingabevektoren die Länge 1 haben, gilt $M = 1$ und Algorithmus 5.2.3 macht $n - 1$ Iterationen. Die Schranken für die auftretenden Werte aus Lemma 5.2.5, 5.2.6 und 5.2.5 gelten offenbar auch für das Lovász-Verfahren für linear abhängige Erzeugendensysteme.

5.2.3 Praktisches Verfahren zur LLL-Reduktion

Im Gleitkomma-Algorithmus 5.2.4 für LLL-Reduktion werden die Basisvektoren $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$ in ganzzahliger Darstellung und die Werte $\mu_{i,j}, \|\widehat{b}_i\|^2$ in Gleitkommazahlen gespeichert. Die Basisvektoren in Gleitkommadarstellung bezeichnen wir mit $b'_1, b'_2, \ldots, b'_n$. Die Basis muß exakt vorliegen, da Abweichungen das Gitter verändern und nicht mehr korrigiert werden können. Die Abweichungen in der Gleitkommadarstellung der übrigen Werte können durch eine exakte Basis berichtigt werden. Der folgende Algorithmus versucht, die Fehler durch Gleitkomma-Arithmetik zu minimieren (sei τ die Anzahl der Precision-Bits):

1. Bei jedem Eintritt in die Stufe k werden durch die Basisvektoren $b_1, b_2, \ldots, b_k$ die Werte $\mu_{k,j}$ für $j = 1, 2, \ldots, k-1$ und außerdem $c_k := \|\widehat{b}_k\|^2$ neu berechnet.

2. Tritt bei der Längenreduktion von b_k ein sehr großer Koeffizient $|\lceil \mu_{k,j} \rfloor| > 2^{\frac{\tau}{2}}$ auf, gehen wir auf Stufe $k-1$ zurück. Dies berichtigt die Koeffizienten $\mu_{k-1,j}$ und $\mu_{k,j}$ (für $j = 1, 2, \ldots, k$) sowie c_{k-1}, c_k und b'_{k-1}, b'_k.

3. Falls $\left|\langle b'_k, b'_j \rangle\right| < 2^{-\frac{\tau}{2}} \|b'_k\| \cdot \left\|b'_j\right\|$, berechnen wir mit Hilfe der ganzzahligen Darstellung der beiden Basisvektoren $\langle b_k, b_j \rangle$ anstelle der Gleitkommawerte. Damit versuchen wir zu verhindern, daß der kleinere Absolutbetrag von der Gleitkomma-Arithmetik wegen großer Koeffizienten der Basisvektoren relativ stark vom korrekten Resultat abweicht.

Es ist nicht bewiesen, daß das obige Verfahren stets terminiert. In der Praxis hat es sich allerdings bewährt (siehe [SchnEu91, LARIFARI]). In [SchnEu91] wird eine Modifikation, sogenannte *tiefe Einfügungen* (Deep Insertions), vorgeschlagen.

Algorithmus 5.2.4 L^3FP (LLL-Reduktion für Gleitkomma-Arithmetik)

EINGABE: ▷ Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^m$
▷ δ mit $\frac{1}{4} < \delta < 1$

1. $c_1 := \|b_1\|^2$, $k := 2$, F :=false

/∗ Bei Eintritt in die Stufe k liegen vor:

$\mu_{i,j}$ für $1 \le j < i \le k$ und $c_i := \|\widehat{b}_i\|^2$ für $i = 1, 2, \ldots, k-1$ ∗/

2. FOR $i = 1, 2, \ldots, n$ DO $b'_i := \mathrm{Float}(b_i)$

3. WHILE $k \le n$ DO

/∗ Berechne $\mu_{k,1}, \mu_{k,2}, \ldots, \mu_{k,k-1}$ und $c_i = \|\widehat{b}_i\|^2$ für $i = 1, 2, \ldots, k-1$ ∗/

3.1. $c_k := \|b'_k\|^2$

3.2. IF $k = 2$ THEN $c_1 := \|b_1\|^2$

3.3. FOR $j = 1, 2, \ldots, k-1$ DO

3.3.1. IF $\left|\langle b'_k, b'_j\rangle\right| < 2^{-\frac{\tau}{2}} \|b'_k\| \cdot \left\|b'_j\right\|$ THEN $s := \mathrm{Float}(\langle b_k, b_j\rangle)$
ELSE $s := \langle b'_k, b'_j\rangle$

3.3.2. $\mu_{k,j} := \left(s - \sum_{i=1}^{j-1} \mu_{j,i}\mu_{k,i}c_i\right)/c_k$

3.3.3. $c_k := c_k - \mu_{k,j}^2 c_j$

END for j

3.4. FOR $j = k-1, k-2, \ldots, 1$ DO /∗ Längenreduktion von b_k ∗/

3.4.1. IF $|\mu_{k,j}| > 1/2$ THEN

3.4.1.1. $\mu := \lceil \mu_{k,j} \rfloor$

3.4.1.2. IF $|\mu| > 2^{\frac{\tau}{2}}$ THEN f :=true

3.4.1.3. FOR $i = 1, 2, \ldots, j-1$ DO $\mu_{k,i} := \mu_{k,i} - \mu\mu_{j,i}$

3.4.1.4. $\mu_{k,j} := \mu_{k,j} - \mu$; $b_k := b_k - \mu b_j$; $b'_k := \mathrm{Float}(b_k)$

END if

END for

3.5. IF (F) THEN F :=false; $k := \max(k-1, 2)$; GOTO 3

/∗ Vertausche b_{k-1} und b_k oder erhöhe k ∗/

3.6. IF $\delta c_{k-1} > c_k + \mu_{k,k-1}^2 c_{k-1}$ THEN

3.6.1. $b_{k-1} \leftrightarrow b_k$ und $b'_{k-1} \leftrightarrow b'_k$, d.h. vertausche Vektoren

3.6.2. $k := \max(k-1, 2)$

ELSE $k := k+1$

END while

AUSGABE: Mit δ LLL-reduzierte Basis $b_1, b_2, \ldots, b_n$

Kapitel 6

Lösen von Subsetsum-Problemen durch Gitterreduktion

In diesem Kapitel lernen wir die erste Anwendung der Gitterreduktion kennen: Wir nehmen an, es gäbe ein Gitterorakel, daß uns den kürzesten, nichttrivialen Gittervektor liefert und reduzieren das Subsetsum-Problem auf das Finden eines kürzesten Gittervektors. Wir stellen zunächst die Lagarias-Odlyzko- und anschließend die verbesserte CJLOSS-Gitterbasis vor. Für kleine Dimensionen können wir das Gitterorakel durch Reduktionsalgorithmen annährend ersetzen.

6.1 Einleitung

Wir versuchen, die folgende Aufgabe mit Hilfe eines Gitterorakels bzw. durch Gitterreduktion zu lösen:

Definition 6.1.1 (Subsetsum-Problem)
Das Subsetsum-Problem lautet:

- *Gegeben:* $n \in \mathbb{N}$*, Gewichte* $a_1, a_2, \ldots, a_n \in \mathbb{N}$ *und* $s \in \mathbb{N}$
- *Finde* $e \in \{0,1\}^n$ *mit* $\sum_{i=1}^{n} a_i e_i = s$ *oder zeige, daß kein solcher Vektor existiert.*

Das Subsetsum-Problem nennt man in der Literatur auch Knapsack- bzw. Rucksack-Problem. Nach Satz 1.2.5 auf Seite 9 ist das Subsetsum-Problem $\mathcal{NP}$-vollständig.

Sei $A \in \mathbb{N}$ eine beliebige Konstante. Wir betrachten Gewichte $(a_1, a_2, \ldots, a_n)$, welche über dem Bereich $[1, A]^n$ variieren. Zusätzlich setzen wir voraus, daß stets eine Lösung existiert: Sei $e = (e_1, e_2, \ldots, e_n) \in \{0,1\}^n \setminus \{0^n\}$ beliebig, aber fest. Setze

$$s := \sum_{i=1}^{n} a_i e_i$$

Die Wahrscheinlichkeiten und die „fast alle“-Aussagen in diesem Kapitel beziehen sich auf rein zufällig gewählte Tupel aus $[1, A]^n$. Zur Subsetsum-Aufgabe formulieren wir das inverse Problem:

Definition 6.1.2 (Inverses Subsetsum-Problem)
Die inverse Aufgabe zu einem Subsetsum-Problem lautet:

- *Gegeben:* $n \in \mathbb{N}$, *Gewichte* $a_1, a_2, \ldots, a_n \in \mathbb{N}$ *und* $s \in \mathbb{N}$
- *Finde* $\overline{e} \in \{0,1\}^n$ *mit* $\sum_{i=1}^{n} a_i \overline{e}_i = \sum_{i=1}^{n} a_i - s =: \overline{s}$ *oder zeige, daß kein solcher Vektor existiert.*

Sei e eine Lösung zum Subsetsum-Problem und $\overline{e}$ zum Inversen. Dann gilt:

$$\overline{e}_i := 1 - e_i \qquad i = 1, 2, \ldots, n$$

Aus der Lösung zum inversen Problem erhalten wir unmittelbar eine Lösung des Ausgangsproblems. Durch den möglichen Übergang zum inversen Problem können wir stets erreichen, daß die Summe der Einsen im Lösungsvektor e bzw. $\overline{e}$ maximal $\frac{n}{2}$ beträgt.

Um die Aufgabe zu lösen, setzen wir ein Gitterorakel voraus. Das Gitterorakel liefert zu gegebener, ganzzahliger Basis zum Gitter L einen Vektor $x \in L$ mit $\|x\| = \lambda_1(L)$ (Euklidische Norm). Wir werden zeigen, daß wir mit dem Gitterorakel die Aufgabe fast immer lösen können (also die Wahrscheinlichkeit, daß wir es nicht lösen können, fällt mit n gegen unendlich gegen 0), wenn die Dichte niedrig ist:

Definition 6.1.3 (Dichte eines Subsetsum-Problems)
Zu einem Subsetsum-Problem mit Gewichten $a_1, a_2, \ldots, a_n \in \mathbb{N}$ *definieren wir die Dichte* d *als:*

$$d := \frac{n}{\log_2 \left(\max\limits_{i=1,2,\ldots,n} a_i \right)}$$

Für Dichte $d \gg 1$ gibt es bei zufälliger Wahl der Gewichte $a_1, a_2, \ldots, a_n$ und der Summe s „in der Regel“ viele Lösungen, als am schwierigsten gelten zufällige Subsetsum-Problem mit Dichte etwa 1. Aus gegebener Dichte d und der Anzahl n erhalten wir eine untere Schranke für die Gewichte $a_1, a_2, \ldots, a_n$:

$$\max_{i=1,2,\ldots,n} a_i \geq 2^{\frac{n}{d}} \tag{6.1}$$

In der Praxis versucht man, statt durch Fragen an das Orakels, einen der kürzesten Gittervektor mit Hilfe der Gitterbasenreduktion zu finden (siehe u.a. [SchnHö95, SchnEu91, Hörner94]). Dies bietet auch eine Angriffsmöglichkeit auf Kryptographie-Schemata, die auf Subsetsum-Problemen basieren. C.P. Schnorr und H.H. Hörner [SchnHö95, Hörner94] haben das Chor-Rivest-System [CR88] mittels Gitterreduktion angegriffen.

6.2 Lagarias-Odlyzko-Gitterbasis

J.C. Lagarias und A.M. Odlyzko [LaOd85] haben 1985 eine Gitterbasis vorgestellt, um Subsetsum-Aufgaben mit Hilfe eines Gitterorakels zu lösen. Unsere Darstellung orientiert sich an [CJLOSS92]. Die Lagarias-Odlyzko-Gitterbasis besteht aus folgenden $n+1$ ganzzahligen Zeilenvektoren, wobei N eine hinreichend große Zahl ist:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ b_{n+1} \end{bmatrix} := \begin{bmatrix} 1 & 0 & \cdots & 0 & Na_1 \\ 0 & 1 & & 0 & Na_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & & 1 & Na_n \\ 0 & 0 & \cdots & 0 & Ns \end{bmatrix} \tag{6.2}$$

Im Beweis wählen wir $N > \sqrt{\frac{1}{2}n}$. Die Motivation: Ein kurzer Gittervektor hat dann in der letzten Komponente den Wert 0, und wir erhalten aus den ersten n Komponenten eine Lösung des Subsetsum-Problems. Sei

$$L_{\mathrm{LO}} := L(b_1, b_2, \ldots, b_{n+1})$$

Der Lösung e des Subsetsum-Problems, die nach Voraussetzung existiert, ist der folgende Lösungsvektor zugeordnet:

$$\widehat{e} := \left(\sum_{i=1}^{n} e_i b_i\right) - b_{n+1} = (e_1, e_2, \ldots, e_n, 0) \tag{6.3}$$

Satz 6.2.1
Sei $A > 0$ beliebig, aber fest. Die Subsetsum-Aufgabe wird für hinreichend große n für fast alle ganzzahligen Gewichte $(a_1, a_2, \ldots, a_n) \in_{\mathrm{R}} [1, A]^n$ mit Dichte $d < 0,6463$ durch zweifache Anwendung des Gitterorakels auf die Lagarias-Odlyzko-Basis effizient gelöst.

Beweis. Wir wenden das Orakel auf das Problem und sein inverses Problem an: Zuvor entfernen wir diejenigen Gewichte, die wir im voraus einer Lösung zuordnen können.

Sei $t := \sum_{i=1}^{n} a_i$. Wir reduzieren das Problem: Solange ein a_i mit $a_i > \min(s, t-s)$ existiert, entferne dieses Gewichte aus der Liste, vermindere n um 1 und aktualisiere s und t. Ein a_i mit $a_i > t - s$ muß Summand in $\sum_{i=1}^{n} e_i a_i$ sein. Ein a_i mit $a_i > s$ muß Summand in $\sum_{i=1}^{n}(1-e_i)a_i$ sein.

Jede Lösung des reduzierten Problems liefert eine Lösung der Aufgabe. Für die reduzierte Aufgabe gilt

$$\tfrac{1}{n} \cdot t \le s \le t - \tfrac{1}{n} \cdot t, \tag{6.4}$$

weil alle kleineren und größeren Gewichte entfernt wurden. Die Ungleichung bleibt bestehen, wenn man auf das inverse Problem übergeht. Die folgende Analyse bezieht sich auf dasjenige Problem mit $\sum_{i=1}^{n} e_i \le \frac{1}{2}n$.

Es bleibt die Wahrscheinlichkeit abzuschätzen, daß das Orakel einen kürzesten Gittervektor $\widehat{x} = (x_1, x_2, \ldots, x_{n+1}) \neq \pm\widehat{e}$ ausgibt, da wir dann aus der Antwort nicht die gesuchte Subsetsum-Lösung erhalten. Für den Vektor $\widehat{x}$ gilt:

$$\begin{aligned} &\|\widehat{x}\| \le \|\widehat{e}\| \le \sqrt{\tfrac{1}{2}n} \\ &\widehat{x} \in L_{\mathrm{LO}} = L(b_1, b_2, \ldots, b_{n+1}) \\ &\widehat{x} \notin \{0, \pm\widehat{e}\} \end{aligned} \tag{6.5}$$

Für $N > \sqrt{\frac{1}{2}n}$ folgt, daß $x_{n+1} = 0$ ist. Setze

$$x := (x_1, x_2, \ldots, x_n) \tag{6.6}$$

Definiere:

$$y := \frac{1}{s}\sum_{i=1}^{n} x_i a_i \tag{6.7}$$

Dann gilt

$$|y| \le n \cdot \sqrt{\tfrac{1}{2}n}, \tag{6.8}$$

da aus der Cauchy-Schwarz-Ungleichung und $a \in \mathbb{N}^n$

$$|y| = \frac{|\langle x, a\rangle|}{s} \leq \frac{\|x\| \cdot \|a\|}{s} \leq \frac{\|x\| \, \|a\|_1}{s} \leq \frac{\|x\|}{s} \cdot \sum_{i=1}^{n} a_i$$

folgt und wir wegen $t = \sum_{i=1}^{n} a_i$ sowie (6.4),(6.5) und $x_{n+1} = 0$ erhalten:

$$|y| \leq \frac{t \cdot \|x\|}{s} \leq n \cdot \sqrt{\tfrac{1}{2}n}$$

Es bezeichne im weiteren

$$P(n) := \mathrm{Ws}[\text{Es existiert ein } \widehat{x} \text{ mit (6.5)}]$$

die Wahrscheinlichkeit bezüglich zufälliger, gleichverteilter und unabhängiger $(a_1, a_2, \ldots, a_n)$ aus $[1, A]^n$. Zu zeigen: Für Dichte $d < 0,6463$ gilt $\lim_{n\to\infty} P(n) = 0$. Es ist:

$$P(n) = \mathrm{Ws}\left[\begin{array}{l} \exists \widehat{x} \in L_{\mathrm{LO}}, \exists y \in \mathbb{Z}, \text{ so daß:} \\ \|\widehat{x}\| \leq \|\widehat{e}\|, \quad |y| \leq n \cdot \sqrt{\frac{1}{2}n}, \quad \widehat{x} \notin \{0, \pm\widehat{e}\}, \quad \sum_{i=1}^{n} a_i x_i = ys \end{array}\right]$$

$$\leq \overbrace{\mathrm{Ws}\left[\sum_{i=1}^{n} a_i x_i = ys, \quad x \in \mathbb{Z}^n \text{ und } y \in \mathbb{Z} \text{ fest}, \quad \|x\| \leq \|\widehat{e}\|, \quad |y| \leq n \cdot \sqrt{\tfrac{1}{2}n}, \quad x \notin \{0, \pm\widehat{e}\}\right]}^{\text{Faktor 1}}$$

$$\cdot \underbrace{\left|\left\{x \in \mathbb{Z}^n \,:\, \|x\| \leq \sqrt{\tfrac{1}{2}n}\right\}\right|}_{\text{Faktor 2}} \cdot \underbrace{\left|\left\{y \in \mathbb{Z} \,:\, |y| \leq n \cdot \sqrt{\tfrac{1}{2}n}\right\}\right|}_{\text{Faktor 3}}$$

Wir schätzen die drei Faktoren nach oben ab:

1. Seien $x \in \mathbb{Z}^n$ und $y \in \mathbb{Z}$ beliebig, aber fest mit den angegebenen Eigenschaften. Mit $z_i := x_i - ye_i$ für $i = 1, 2, \ldots, n$ gilt wegen $\sum_{i=1}^{n} e_i a_i = s$:

   $$\sum_{i=1}^{n} a_i x_i = ys \qquad \Longleftrightarrow \qquad \sum_{i=1}^{n} a_i z_i = 0$$

   Der Vektor $z = (z_1, z_2, \ldots, z_n)$ ist fest. Es gilt $z \neq 0$, da sonst aus $x = y\widehat{e}$ und $x \notin \{0, \pm\widehat{e}\}$ folgt $|y| \geq 2$ und $\|x\| \geq 2\|\widehat{e}\|$ — Widerspruch zu $\|x\| \leq \|\widehat{e}\|$.

   Sei $z_j \neq 0$ für ein festes j. Für fest gewählte a_i mit $i \neq j$ ist die Gleichung $\sum_{i=1}^{n} a_i z_i = 0$ für höchstens ein $a_j \in [1, A]$ erfüllt. Es folgt:

   $$\text{Faktor 1} \leq \mathrm{Ws}\left[\sum_{i=1}^{n} a_i z_i = 0\right] \leq \frac{1}{A}$$

2. J.C.Lagarias und A.M. Odlyzko [LaOd85] haben gezeigt, daß für hinreichend große n gilt:

   $$\left|\left\{x \in \mathbb{Z}^n \,:\, \|x\| \leq \sqrt{\tfrac{1}{2}n}\right\}\right| \leq 2^{c_0 n} \qquad \text{mit } c_0 = 1,54725$$

3. Es gilt:

   $$\left|\left\{y \in \mathbb{Z} \,:\, |y| \leq n \cdot \sqrt{\tfrac{1}{2}n}\right\}\right| \leq 1 + 2\left(n \cdot \sqrt{\tfrac{1}{2}n}\right)$$

Damit ergibt sich:

$$P(n) \leq \frac{2^{c_0 n}}{A} \cdot \left(1 + 2n \cdot \sqrt{\tfrac{1}{2}n}\right)$$

Wegen $\frac{1}{d} > 1,547269 > c_0$ und der unteren Schranke (6.1) auf Seite 74

$$A \geq \max_{i=1,2,\ldots,n} a_i \geq 2^{\frac{n}{d}},$$

gilt $\lim_{n\to\infty} P(n) = 0$. ■

6.3 CJLOSS-Gitterbasis

1992 haben M.J. Coster, A. Joux, B.A. LaMacchina, A.M. Odlyzko, C.P. Schnorr und und J. Stern [CJLOSS92] durch Modifikation des Vektors b_{n+1} der Lagarias-Odlyzko-Basis die Grenzdichte auf $0,9408$ erhöht. Die CJLOSS-Basis erhält man aus der Lagarias-Odlyzko-Gitterbasis (6.2), indem der Zeilenvektor b_{n+1} durch $b'_{n+1} := (\frac{1}{2}, \ldots, \frac{1}{2}, Ns)$ ersetzt wird:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ b'_{n+1} \end{bmatrix} := \begin{bmatrix} 1 & 0 & \cdots & 0 & Na_1 \\ 0 & 1 & & 0 & Na_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & & 1 & Na_n \\ \frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} & Ns \end{bmatrix} \tag{6.9}$$

N ist eine hinreichend große Zahl, im Beweis wählen wir $N > \frac{1}{2}\sqrt{n}$. Sei

$$L_{\text{CJLOSS}} := L(b_1, b_2, \ldots, b_{n+1})$$

Der Lösung e des Subsetsum-Problems, die nach Voraussetzung existiert, ist der folgende Lösungsvektor zugeordnet:

$$\widehat{e}' := \left(\sum_{i=1}^{n} e_i b_i\right) - b'_{n+1} = \left(e_1 - \tfrac{1}{2}, e_2 - \tfrac{1}{2}, \ldots, e_n - \tfrac{1}{2}, 0\right) \tag{6.10}$$

Da $e_i \in \{0,1\}$, gilt $\|\widehat{e}'\| = \frac{1}{2}\sqrt{n}$. Im Vergleich zum Lösungsvektor e der Lagarias-Odlyzko-Basis (6.3) ist $e' = e_i - \frac{1}{2}$. Der Vorteil der CJLOSS-Basis liegt darin, daß ihr Lösungsvektor bis zu einem Faktor $\sqrt{2}$ kleiner als der Lösungsvektor $\widehat{e}$ der Lagarias-Odlyzko-Basis ist.

Satz 6.3.1
Sei $A > 0$ beliebig, aber fest. Die Subsetsum-Aufgabe wird für hinreichend große n für fast alle ganzzahligen Gewichte $(a_1, a_2, \ldots, a_n) \in_{\text{R}} [1, A]^n$ mit Dichte $d < 0,9408$ durch zweifache Anwendung des Gitterorakels auf die CJLOSS-Basis effizient gelöst.

Beweis. Wir schätzen die Wahrscheinlichkeit ab, daß das Orakel einen kürzesten Gittervektor $\widehat{x} = (x_1, x_2, \ldots, x_{n+1}) \neq \pm\widehat{e}'$ liefert. Für den Vektor $\widehat{x}$ gilt:

$$\begin{aligned} &\|\widehat{x}\| \leq \|\widehat{e}'\| \leq \tfrac{1}{2}\sqrt{n} \\ &\widehat{x} \in L_{\text{CJLOSS}} = L(b_1, b_2, \ldots, b_n, b'_{n+1}) \\ &\widehat{x} \notin \{0, \pm\widehat{e}'\} \end{aligned} \tag{6.11}$$

Seien $y_1, y_2, \ldots, y_n, y \in \mathbb{Z}$ die Koeffizienten der Darstellung von $\widehat{x}$ als Linearkombination der Basisvektoren:

$$\widehat{x} = \sum_{i=1}^{n} y_i b_i + y b'_{n+1}$$

Betrachten wir die letzte Komponente: Wegen $\sum_{i=1}^{n} y_i a_i + ys \in \mathbb{Z}$ gilt für $N > \frac{1}{2}\sqrt{n}$, daß $x_{n+1} = 0$ ist. Ferner gilt:

$$x_i = y_i + \tfrac{1}{2}y \qquad \text{für } i = 1, 2, \ldots, n \tag{6.12}$$

$$x_{n+1} = N\left(\sum_{i=1}^{n} a_i y_i + ys\right) \tag{6.13}$$

Wegen $x_{n+1} = 0$ folgt aus (6.13):

$$\sum_{i=1}^{n} a_i y_i = -ys \tag{6.14}$$

Wir erhalten mit $t := \sum_{i=1}^{n} a_i$:

$$\begin{aligned}
\sum_{i=1}^{n} x_i a_i &= \sum_{i=1}^{n} a_i \left(y_i + \tfrac{1}{2}y\right) && \text{(wegen (6.12))} \\
&= \sum_{i=1}^{n} y_i a_i + \tfrac{1}{2}y \sum_{i=1}^{n} a_i \\
&= -ys + \tfrac{1}{2}yt && \text{(wegen (6.14))} \\
&= \tfrac{1}{2}y\,(t - 2s)
\end{aligned}$$

Aus diesem Resultat folgt mit $\alpha := \max_{i=1,2,\ldots,n} a_i$:

$$\begin{aligned}
|y(t-2s)| &\le 2 \cdot \sum_{i=1}^{n} |x_i a_i| && \text{(Dreiecksungleichung)} \\
&\le 2\alpha \cdot \|\widehat{x}\|_1 && \text{(wobei } \|\cdot\|_1 \text{ die 1-Norm ist)} \\
&\le 2\alpha\sqrt{n} \cdot \|\widehat{x}\|_2 && \text{(wobei } \|\cdot\|_2 \text{ die Euklidische Norm ist)} \\
&\le \alpha n && \text{(wegen (6.11): } \|\widehat{x}\| \le \|\widehat{e}'\| = \tfrac{1}{2}\sqrt{n}\text{)}
\end{aligned} \tag{6.15}$$

Um eine geeignete Schranke für $|y|$ zu erhalten, reduzieren wir die Subsetsum-Aufgabe, so daß dann für das Problem gilt

$$|y| \le 2n \tag{6.16}$$

Falls $|t - 2s| \ge \frac{1}{2}\alpha$, wird nicht reduziert, da wegen (6.15) bereits $|y| \le \frac{\alpha n}{|t-2s|} \le 2n$ gilt. Falls $|t - 2s| < \frac{1}{2}\alpha$, dann entferne ein Gewicht a_i mit $a_i = \alpha$ aus der Aufgabe. Wir können zwei Probleme lösen: Eines mit a_i in der Teilmenge, die sich zu s summiert, und ein anderes mit a_i in der Teilmenge, die sich zu $t - s$ summiert.

Für das erste Problem gilt mit $s_{\text{neu}} = s - \alpha$ und $t_{\text{neu}} = t - \alpha$:

$$|t_{\text{neu}} - 2s_{\text{neu}}| = |t - \alpha - 2s + 2\alpha| > \tfrac{1}{2}\alpha$$

Für das andere Problem gilt mit $s_{\text{neu}} = s$ und $t_{\text{neu}} = t - \alpha$:

$$|t_{\text{neu}} - 2s_{\text{neu}}| = |t - \alpha - 2s - 2\alpha| > \tfrac{1}{2}\alpha$$

Beide reduzierte Probleme erfüllen die Ungleichung $|t - 2s| \ge \frac{1}{2}\alpha$. Aus (6.15) folgt, daß die Forderung (6.16), also $|y| \le 2n$, erfüllt ist.

Wir wenden das Orakel, sofern reduziert wurde, auf beide Probleme an. Es bleibt die Wahrscheinlichkeit abzuschätzen, daß das Orakel einen kürzesten Gittervektor $\widehat{x}$ mit Eigenschaften (6.11) liefert. Es bezeichne im weiteren

$$P(n) := \text{Ws}[\,\text{Es existiert ein } \widehat{x} \text{ mit (6.11)}]$$

die Wahrscheinlichkeit bezüglich zufälliger und gleichverteilter $(a_1, a_2, \ldots, a_n)$ aus $[1, A]^n$. Zu zeigen: Für Dichte $d < 0,94080$ gilt $\lim_{n\to\infty} P(n) = 0$. Aus

$$P(n) = \mathrm{Ws}\left[\begin{array}{l} \exists \widehat{x} \in L_{\mathrm{CJLOSS}}, \exists y \in \mathbb{Z}, \text{ so daß:} \\ \\ \|\widehat{x}\| \leq \|\widehat{e}'\|, \quad |y| \leq 2n, \quad \widehat{x}' \notin \{0, \pm\widehat{e}'\}, \quad \sum_{i=1}^n a_i x_i = \frac{1}{2} y(t-2s) \end{array}\right]$$

folgt:

$$P(n) \leq \overbrace{\mathrm{Ws}\left[\begin{array}{l} \sum_{i=1}^n a_i x_i = \frac{1}{2} y(t-2s),\ x \in \mathbb{Z}^n + \mathbb{Z}\left(\frac{1}{2}, \ldots, \frac{1}{2}\right),\ y \in \mathbb{Z} \text{ fest}, \\ \|x\| \leq \|e\|,\ |y| \leq n \cdot \sqrt{2n},\ x \notin \{0, \pm e\} \end{array}\right]}^{\text{Faktor 1}}$$

$$\cdot \underbrace{\left|\left\{x \in \mathbb{Z}^n + \left(\tfrac{1}{2}, \ldots, \tfrac{1}{2}\right)\mathbb{Z} \,:\, \|x\| \leq \tfrac{1}{2}\sqrt{n}\right\}\right|}_{\text{Faktor 2}} \cdot \underbrace{\left|\{y \in \mathbb{Z} \,:\, |y| \leq 2n\}\right|}_{\text{Faktor 3}}$$

Wir schätzen die drei Faktoren nach oben ab:

1. Seien $x \in \mathbb{Z}^n + \mathbb{Z}\left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$ und $y \in \mathbb{Z}$ beliebig, aber fest mit den angegebenen Eigenschaften. Für

   $$z_i := x_i + y\left(e_i - \tfrac{1}{2}\right) = x_i + y e'_i \qquad \text{für } i = 1, 2, \ldots, n$$

   gilt:

   $$\sum_{i=1}^n a_i x_i = \tfrac{1}{2} \cdot y \cdot (t-2s) \qquad \Longleftrightarrow \qquad \sum_{i=1}^n a_i z_i = 0$$

   Der Vektor $z = (z_1, z_2, \ldots, z_n)$ ist fest. Es gilt $z \neq 0$, da sonst aus $x = ye'$ und $x \notin \{0, \pm e'\}$ folgt $|y| \geq 2$ und $\|x\| \geq 2\,|e'|$ — Widerspruch zu $\|x\| \leq \|e'\|$.

   Sei $z_j \neq 0$ für ein festes j. Für fest gewählte a_i mit $i \neq j$ ist die Gleichung $\sum_{i=1}^n a_i z_i = 0$ für höchstens ein $a_j \in [1, A]$ erfüllt. Es folgt:

   $$\text{Faktor 1} \leq \mathrm{Ws}\left[\sum_{i=1}^n a_i z_i = 0\right] \leq \frac{1}{A}$$

2. J.C.Lagraias und A.M. Odlyzko [LaOd85] haben die Anzahl der Gitterpunkte von $\mathbb{Z}^n$ in einer Kugel mit Radius $\sqrt{\alpha n}$ um den Ursprung

   $$N(n, \alpha) := \left|\left\{x \in \mathbb{Z}^n \,:\, \|x\|^2 \leq \alpha \cdot n\right\}\right|$$

   untersucht. Für hinreichend große n zeigen sie, daß für jedes $u > 0$ gilt

   $$N(n, \alpha) \leq 2^{(\log_2 e)\cdot\delta(\alpha,u)\cdot n}$$

   mit $\delta(\alpha, u) = \alpha u + \ln\theta(e^{-u})$ und der Theta-Funktion $\theta(z) = 1 + 2\sum_{i=1}^\infty z^{i^2}$. Für festes α kann man die Minimalstelle u_0 von $\delta(\alpha, u)$ numerisch annähern. Für $\alpha = \frac{1}{4}$ erhalten wir $u_0 \approx 1,8132$. Es folgt:

   $$\min_{u>0} \delta\left(\tfrac{1}{4}, u\right) \leq \delta\left(\tfrac{1}{4}; 1,8132\right) \approx 0,7367$$

   Wir können daher den zweiten Faktor nach oben abschätzen durch:

   $$\left|\left\{x \in \mathbb{Z}^n + \left(\tfrac{1}{2}, \ldots, \tfrac{1}{2}\right)\mathbb{Z} \,:\, \|x\| \leq \tfrac{1}{2} \cdot \sqrt{n}\,\right\}\right| \leq 2^{c'_0 n} \qquad \text{mit } c'_0 = 1,0629$$

3. Es gilt:

$$|\{y \in \mathbb{Z} \,:\, |y| \leq 2n\}| \leq 1 + 4n$$

Damit ergibt sich:

$$P(n) \leq (4n+1) \cdot \frac{2^{c_0' n}}{A}$$

Wegen $\frac{1}{d} > 1,062925 > c_0'$ und $A \geq \max\limits_{i=1,2,\ldots,n} a_i \geq 2^{\frac{n}{d}}$ gilt: $\lim\limits_{n\to\infty} P(n) = 0$. ■

Kapitel 7

HKZ- und β-reduzierte Gitterbasen

Im Kapitel 5 haben wir den Reduktionsbegriff der LLL-reduzierten Basis $b_1, b_2, \ldots, b_n$ kennengelernt. Zwar arbeitet der Reduktionsalgorithmus 5.2.1 in Polynomialzeit, aus der Eigenschaft „LLL-reduziert“ haben wir in Satz 5.1.4 nur zeigen können, daß die Länge des kürzesten Basisvektors maximal ein in n exponentielles Vielfaches des ersten sukzessiven Minimums ist. Aus diesem Grund benötigen wir stärkere Reduktionsbegriffe. In diesem Kapitel lernen wir Begriff einer HKZ-reduzierten Gitterbasis kennen. Die Verallgemeinerung, β-reduzierte Gitterbasen, umfaßt sowohl LLL- als auch HKZ-reduzierte Basen.

7.1 HKZ-reduzierte Gitterbasen

Zu einer Basis $b_1, b_2, \ldots, b_n$ des Gitters L sind die die projizierten Gitter L_i, $i = 1, 2, \ldots, n$, erklärt durch:

$$L_i = \pi_i(L) := L\left(\pi_i(b_i), \pi_i(b_{i+1}), \ldots, \pi_i(b_n)\right)$$

C. Hermite [Hermite1850] sowie unabhängig A. Korkine und G. Zolotareff [KoZo1873, KoZo1877] haben die folgende Definition einer reduzierten Basis in der Sprache quadratischer Formen formuliert:

Definition 7.1.1 (HKZ-reduzierte Basis)
Eine geordnete Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ ist nach Hermite und Korkine-Zolotareff reduziert (eine HKZ-reduzierte Basis), wenn:

a) $|\mu_{i,j}| \leq \frac{1}{2}$ *für* $1 \leq j < i \leq n$

b) $\|\widehat{b}_i\| = \lambda_1(L_i)$ *für* $i = 1, 2, \ldots, n$

Insbesondere ist $\|b_1\| = \lambda_1(L)$, d.h. der Vektor b_1 ist ein kürzester, nichttriviale Gittervektor. Für eine HKZ-reduzierte Basis $b_1, b_2, \ldots, b_n$ ist auch für $1 \leq j \leq n$

$$\pi_j(b_j), \pi_j(b_{j+1}), \ldots, \pi_j(b_n)$$

eine HKZ-reduzierte Basis. Wie gut approximieren der Basisvektoren die kürzesten, nicht-trivialen Gittervektoren?

Satz 7.1.2
Für jede $b_1, b_2, \ldots, b_n$ HKZ-reduzierte Basis von L gilt für $i = 1, 2, \ldots, n$:

$$\frac{4}{i+3} \leq \frac{\|b_i\|^2}{\lambda_i(L)^2} \leq \frac{i+3}{4}$$

Zum Vergleich: Für eine LLL-reduzierte Basis $b_1, b_2, \ldots, b_n$ gilt nach Satz 5.1.4 auf Seite 57 mit $\alpha = \frac{1}{\delta - \frac{1}{4}}$:

$$\alpha^{1-i} \leq \frac{\|\widehat{b}_i\|^2}{\lambda_i(L)^2} \leq \frac{\|b_i\|^2}{\lambda_i(L)^2} \leq \alpha^{n-1}$$

Beweis (zu Satz 7.1.2). Wir zeigen zunächst die obere Schranke $\frac{\|b_i\|^2}{\lambda_i(L)^2} \leq \frac{i+3}{4}$. Für das Gitter $L_i = \pi_i(L)$ gilt:

$$\|\widehat{b}_i\| = \lambda_1(L_i) \leq \lambda_i(L) \tag{7.1}$$

Die Gleichheit folgt aus der HKZ-Eigenschaft. Die Abschätzung gilt, denn es gibt i linear unabhängige Gittervektoren $a_1, a_2, \ldots, a_i \in L$ mit

$$\|a_1\| \leq \|a_2\| \leq \cdots \|a_i\| \leq \lambda_i(L),$$

und es existiert ein j mit $j \leq i$ und $\pi_i(a_j) \neq 0$, also $\pi_i(a_j) \in L_i \setminus \{0\}$. Wir erhalten aus den Eigenschaften einer HKZ-reduzierten Basis und (7.1):

$$\begin{aligned}
\|b_i\|^2 &= \|\widehat{b}_i\|^2 + \sum_{j=1}^{i-1} (\mu_{i,j})^2 \cdot \|\widehat{b}_j\|^2 \\
&\leq \lambda_i(L)^2 + \tfrac{1}{4} \sum_{j=1}^{i-1} \lambda_j(L)^2 \qquad \text{(wegen (7.1) und Basis längenreduziert)} \\
&\leq \tfrac{i+3}{4} \cdot \lambda_i(L)^2
\end{aligned}$$

Wir zeigen die untere Schranke $\frac{4}{i+1} \leq \frac{\|b_i\|^2}{\lambda_i(L)^2}$. Aus der Definition einer HKZ-reduzierten Basis erhalten wir für $j \leq i$:

$$\|\widehat{b}_j\|^2 = \lambda_1(L_j)^2 \leq \|\pi_j(b_i)\|^2 \leq \|b_i\|^2$$

Es gilt:

$$\|b_j\|^2 = \|\widehat{b}_j\|^2 + \sum_{t=1}^{j-1} (\mu_{j,t})^2 \, \|\widehat{b}_t\|^2 \leq \tfrac{j+3}{4} \cdot \|b_i\|^2$$

Wir erhalten die Behauptung:

$$\lambda_i(L)^2 \leq \max_{j=1,2,\ldots,i} \|b_j\|^2 \leq \max_{j=1,2,\ldots,i} \left\{ \frac{j+3}{4} \cdot \|b_i\|^2 \right\} \leq \frac{i+3}{4} \, \|b_i\|^2$$

■

Weitere Sätze über HKZ-reduzierte Gitterbasen finden sich in der Originalarbeit [LLS90] von J.C. Lagarias, H.W. Lenstra und C.P. Schnorr.

7.2 β-reduzierte Gitterbasen

C.P. Schnorr [Schnorr87, Schnorr94a] hat die Begriffe der HKZ- und der LLL-reduzierten Basen zu β-reduzierten Basen verallgemeinert. Während alle bekannten Reduktionsalgorithmen zu HKZ-Basen expontielle Laufzeit haben, gibt es für kleine Blockgrößen praktikable Reduktionsalgorithmen mit Laufzeiten wie bei der schwächeren LLL-Reduktion. Allerdings ist offen, ob diese Algorithmen polynomielle Laufzeit haben.

Definition 7.2.1 (β-reduzierte Basis)
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine geordnete Basis und $\beta \in \{2, 3, \ldots, n\}$ gegeben. Die geordnete Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ heißt β-reduziert (blockreduziert mit Blockgröße β), wenn:

a) $|\mu_{i,j}| \leq \frac{1}{2}$ *für* $1 \leq j < i \leq n$

b) $\pi_i(b_i), \pi_i(b_{i+1}), \ldots, \pi_i(b_{i+\beta-1})$ *ist HKZ-reduzierte Basis für* $i = 1, 2, \ldots, n - \beta + 1$.

Jede $(\beta+1)$-reduzierte Basis ist auch β-reduziert und eine n-reduzierte Basis $b_1, b_2, \ldots, b_n$ ist eine HKZ-reduzierte Basis. Die zweite Eigenschaft bedeutet:

$$\|\widehat{b}_i\| \leq \lambda_1\Big(\pi_i\big(L(b_1, b_2, \ldots, b_{\min(i+\beta-1,n)})\big)\Big)$$

Wir werden in Defintion 7.4.1 im Abschnitt 7.4 den Begriff der β-reduzierten Basis zur (β, δ)-reduzierten Basis verallgemeinern. Die LLL-Reduktion ist ein Spezialfall der Blockreduktion mit $\beta = 2$:

Satz 7.2.2
Die 2-reduzierten Basen entsprechen genau den LLL-reduzierten Basen mit $\delta = 1$.

Beweis. Sei $b_1, b_2, \ldots, b_n$ eine 2-reduzierte Basis. Dann gilt für $i = 1, 2, \ldots, n-1$:

$$\lambda_1\big(L(\pi_i(b_i), \pi_i(b_{i+1}))\big)^2 = \|\widehat{b}_i\|^2 \leq \|\pi_i(b_{i+1})\|^2 = \|\widehat{b}_{i+1}\|^2 + \mu_{i+1,i}^2 \|\widehat{b}_i\|^2$$

Mit $\delta = 1$ ist dies das zweite Kriterium von LLL-reduziert:

$$\delta \cdot \|\widehat{b}_i\|^2 \leq \|\widehat{b}_{i+1}\|^2 + \mu_{i+1,i}^2 \|\widehat{b}_i\|^2 \qquad \text{für } i = 1, 2, \ldots, n-1$$

Weil jede β-reduzierte Basis längenreduziert ist, erfüllt $b_1, b_2, \ldots, b_n$ die LLL-Eigenschaften.

Umgekehrt zeigen wir, jede mit $\delta = 1$ LLL-reduzierte Basis $b_1, b_2, \ldots, b_n$ ist auch 2-reduziert, also jeder Vektor in $L(\pi_i(b_i), \pi_i(b_{i+1}))$ ungleich dem Nullvektor ist nicht kürzer als $\pi_i(b_i) = \widehat{b}_i$. Wegen

$$\|u \cdot \pi_i(b_i) + v \cdot \pi_i(b_{i+1})\|^2 = (u + v \cdot \mu_{i+1,i})^2 \cdot \|\widehat{b}_i\|^2 + v^2 \|\widehat{b}_{i+1}\|^2$$

ist zu zeigen, daß für alle $(u, v) \in \mathbb{Z}^2 \setminus \{(0,0)\}$ gilt:

$$(u + v\mu_{i+1,i})^2 \cdot \|\widehat{b}_i\|^2 + v^2 \cdot \|\widehat{b}_{i+1}\|^2 \geq \|\widehat{b}_i\|^2 \tag{7.2}$$

Die Ungleichung (7.2) gilt für $v = 0$, da dann der Koeffizient $u \in \mathbb{Z}$ ungleich Null ist. Da die Basis mit Parameter $\delta = 1$ LLL-reduziert ist, gilt nach dem zweiten LLL-Kriterium

$$\underbrace{\delta}_{=1} \cdot \|\widehat{b}_i\|^2 \leq \|\widehat{b}_{i+1}\|^2 + \mu_{i+1,i}^2 \cdot \|\widehat{b}_i\|^2 \tag{7.3}$$

und wir erhalten unmittelbar Ungleichung (7.2) den Fall $v = 1$. Da die Basis längenreduziert ist, gilt $\mu_{i+1,i}^2 \leq \frac{1}{4}$ und nach (7.3) folgt $\frac{3}{4} \cdot \|\widehat{b}_i\|^2 \leq \|\widehat{b}_{i+1}\|^2$. Für den Fall $|v| \geq 2$ erhalten wir aus

$$(u + v\mu_{i+1,i})^2 \cdot \|\widehat{b}_i\|^2 + v^2 \cdot \|\widehat{b}_{i+1}\|^2 \geq v^2 \cdot \|\widehat{b}_{i+1}\|^2 \geq 4 \cdot \|\widehat{b}_{i+1}\|^2 \geq 3 \cdot \|\widehat{b}_i\|^2$$

die Ungleichung (7.2). ■

Wir untersuchen die Eigenschaften β-reduzierter Basen. Wir geben zwei Sätze über β-reduzierte Basen an, die wir später in diesem Abschnitt beweisen werden:

Satz 7.2.3
Für jede β-reduzierte Basis $b_1, b_2, \ldots, b_n$ des Gitters L gilt:

a) $\dfrac{\|\widehat{b}_i\|^2}{\lambda_i(L)^2} \leq (\gamma_\beta)^{2 \cdot \frac{n-i}{\beta-1}}$ *für* $i = 1, 2, \ldots, n$

b) $\dfrac{\|b_i\|^2}{\lambda_i(L)^2} \leq (\gamma_\beta)^{2 \cdot \frac{n-i}{\beta-1}} \cdot \dfrac{i+3}{4}$ *für* $i = 1, 2, \ldots, n$

Dabei ist γ_β die Hermite-Konstante der Dimension β.

Wir betrachten eine untere Schranke für $\frac{\|b_i\|^2}{\lambda_i(L)^2}$, also eine obere Schranke für $\frac{\lambda_i(L)^2}{\|b_i\|^2}$.

Satz 7.2.4
Für jede β-reduzierte Basis $b_1, b_2, \ldots, b_n$ des Gitters L gilt:

$$\frac{\lambda_i(L)^2}{\|b_i\|^2} \leq (\gamma_\beta)^{2 \cdot \frac{i-1}{\beta-1}} \cdot \frac{i+3}{4} \qquad \text{für } i = 1, 2, \ldots, n.$$

Für $i \leq \beta$ können wir die Schranken der Sätze 7.2.3 und 7.2.4 durch die stärkeren Schranken aus 7.1.2 ersetzen, da die β-reduzierten Basen für $i \leq \beta$ HKZ-reduzierte Basen sind. Die Werte $(\gamma_\beta)^{\frac{2}{\beta-1}}$ aus den Sätzen 7.2.3 und 7.2.4 sind bekannt für $\beta = 2, 3, \ldots, 8$:

β	2	3	4	5	6	7	8
$(\gamma_\beta)^{\frac{2}{\beta-1}}$	$\frac{4}{3}$	$2^{1/3}$	$2^{1/3}$	$2^{3/10}$	$2^{2/5} \cdot 3^{-1/15}$	$2^{2/7}$	$2^{2/7}$
$\approx$	1,333	1,260	1,260	1,231	1,226	1,219	1,219

Es ist ein offenes Problem, minimale Konstanten $C_{\beta,n}$ zu finden, so daß

$$\frac{\|b_1\|^2}{\lambda_1(L)^2} \leq C_{\beta,n}$$

für alle β-reduzierten Basen $b_1, b_2, \ldots, b_n$ vom Rang n gilt. Die Schranke aus Satz 7.2.3 ist für $n \geq 3$ nicht scharf. Bisher weiß man:

- $C_{2,n} = \left(\frac{4}{3}\right)^{n-1}$ aus [BaKa84]
- $C_{3,n} = \left(\sqrt{\frac{3}{2}}\right)^{n-3}$ für ungerade $n \geq 3$ aus [Schnorr94a]

Es gilt:

Lemma 7.2.5
Für jede β-reduzierte Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ gilt

$$\|b_1\| \leq (\gamma_\beta)^{\frac{n-1}{\beta-1}} \cdot M$$

mit $M := \max\left\{\|\widehat{b}_{n-\beta+2}\|, \|\widehat{b}_{n-\beta+3}\|, \ldots, \|\widehat{b}_n\|\right\}$.

Beweis. Wir erweitern die Basis $b_1, b_2, \ldots, b_n$ durch $\beta - 2$ linear unabhängige Vektoren zu

$$b_{-\beta+3}, b_{-\beta+4}, \ldots, b_{-1}, b_0, b_1, \ldots, b_n \tag{7.4}$$

so, daß gilt:

$$\|b_i\| = \|b_1\| \qquad \text{für } i \leq 0 \tag{7.5}$$

$$\langle b_i, b_j \rangle = 0 \qquad \text{für } i \leq 0,\ i \leq j \text{ und } j = -\beta+3, -\beta+4, \ldots, n \tag{7.6}$$

Dazu betten wir die Basis in den $\mathbb{R}^{m+\beta-2}$ ein: Wir wählen zum Beispiel $b_{-\beta+3}, b_{-\beta+4}, \ldots, b_{-1}, b_0$ als $\|b_1\|$-Vielfaches der kanonischen Einheitsvektoren in die zusätzlichen $\beta - 2$ Richtungen. Die Gitterbasis (7.4) ist β-reduziert und besteht aus mindestens $2(\beta - 1)$ Vektoren Für jedes i mit $-\beta + 3 \leq i \leq n - \beta + 1$ bilden die Vektoren

$$\pi_i(b_i), \pi_i(b_{i+1}), \ldots, \pi_i(b_{i+\beta-1})$$

eine β-reduzierte und damit eine HKZ-reduzierte Basis des Gitters $L\left(\widehat{b}_i, \widehat{b}_{i+1}, \ldots, \widehat{b}_{i+\beta-1}\right)$ (Beachte Eigenschaft (7.6)), so daß gilt:

$$\|\widehat{b}_i\|^2 \leq \lambda_1\left(L\left(\widehat{b}_i, \widehat{b}_{i+1}, \ldots, \widehat{b}_{i+\beta-1}\right)\right)^2$$

Aus der Definition der Hermite-Konstanten

$$\gamma_\beta = \sup\left\{\frac{\lambda_1(L)^2}{(\det L)^{2/\beta}} \;\middle|\; \mathrm{Rang}\,(L) = \beta\right\}$$

folgt für $i = -\beta + 3, -\beta + 4, \ldots, n - \beta + 1$:

$$\|\widehat{b}_i\|^\beta \leq (\gamma_\beta)^{\frac{\beta}{2}} \prod_{s=0}^{\beta-1} \|\widehat{b}_{i+s}\|$$

Durch Multiplikation der $n - 1$ Ungleichungen erhalten wir:

$$\begin{aligned}\prod_{i=-\beta+3}^{n-\beta+1} \|\widehat{b}_i\|^\beta \leq (\gamma_\beta)^{\frac{\beta(n-1)}{2}} &\|\widehat{b}_{-\beta+3}\|^1 \cdot \|\widehat{b}_{-\beta+4}\|^2 \ldots \|\widehat{b}_1\|^{\beta-1} \\ &\cdot \|\widehat{b}_2\|^\beta \cdot \|\widehat{b}_3\|^\beta \ldots \|\widehat{b}_{n-\beta+1}\|^\beta \\ &\cdot \|\widehat{b}_{n-\beta+2}\|^{\beta-1} \ldots \|\widehat{b}_{n-1}\|^2 \cdot \|\widehat{b}_n\|^1\end{aligned}$$

Dies impliziert:

$$\begin{aligned}&\|\widehat{b}_{-\beta+3}\|^{\beta-1} \cdot \ldots \|\widehat{b}_0\|^2 \cdot \|\widehat{b}_1\|^1 \\ &\leq (\gamma_\beta)^{\frac{\beta(n-1)}{2}} \|\widehat{b}_{n-\beta+2}\|^{\beta-1} \cdot \|\widehat{b}_{n-\beta+2}\|^{\beta-1} \ldots \|\widehat{b}_{n-1}\|^2 \cdot \|\widehat{b}_n\|^1\end{aligned} \tag{7.7}$$

Wegen Eigenschaft (7.5) ist $\|\widehat{b}_i\| = \|b_1\|$ für $i \leq 0$, so daß aus Abschätzung (7.7) folgt

$$\|b_1\|^{\binom{\beta}{2}} \leq (\gamma_\beta)^{\frac{\beta(n-1)}{2}} \cdot M^{\binom{\beta}{2}}$$

mit $M := \max\left\{\|\widehat{b}_{n-\beta+2}\|, \|\widehat{b}_{n-\beta+3}\|, \ldots, \|\widehat{b}_n\|\right\}$. Wir erhalten die Behauptung:

$$\|b_1\| \leq (\gamma_\beta)^{\frac{n-1}{\beta-1}} \cdot M$$

■

Wie gut approximiert der erste Vektor einer β-reduzierten Basis einen kürzesten, nicht-trivialen Vektor?

Korollar 7.2.6
Für jede β-reduzierte Basis $b_1, b_2, \ldots, b_n$ des Gitters L gilt:

$$\|b_1\| \leq (\gamma_\beta)^{\frac{n-1}{\beta-1}} \cdot \lambda_1(L)$$

Beweis. Induktion über n:

- Für $n = \beta$ ist $b_1, b_2, \ldots, b_n$ eine HKZ-reduzierte Basis und es gilt $\|b_1\| = \lambda_1(L)$. Weil für die Hermite-Konstante $\gamma_n \geq 1$ gilt, erhalten wir die Behauptung.
- Sei $n > \beta$ und v einer der kürzesten Gittervektoren ungleich dem Nullvektor. O.B.d.A. sei $v \notin L(b_1, b_2, \ldots, b_{n-1})$, denn sonst betrachte die Basis $b_1, b_2, \ldots, b_{n-1}$ und die Behauptung folgt aus der Induktionsannahme. Wegen $v \notin L(b_1, b_2, \ldots, b_{n-1})$. gilt $\pi_i(v) \neq 0$ für $i = n-\beta+1, n-\beta+2, \ldots, n-1$. Wir erhalten mit $L_i = \pi_i(L)$ für $i = n-\beta+1, n-\beta+2, \ldots, n$:

  $$\lambda_1(L) = \|v\| \geq \lambda_1(L_i) = \|\widehat{b}_i\|$$

  Die Gleichheit $\lambda_1(L_i) = \|\widehat{b}_i\|$ gilt, da $\pi_{n-\beta+1}(b_i)$, $i = n-\beta+1, n-\beta+2, \ldots, n$ eine HKZ-reduzierte Basis ist. Wegen

  $$\begin{aligned}\lambda_1(L) &\geq \max\left\{\|\widehat{b}_i\| \;:\; i = n-\beta+1, n-\beta+2, \ldots, n\right\} \\ &\geq \max\left\{\|\widehat{b}_i\| \;:\; i = n-\beta+2, n-\beta+3, \ldots, n\right\} = M\end{aligned}$$

  folgt die Behauptung unmittelbar aus Lemma 7.2.5:

  $$\|b_1\| \leq (\gamma_\beta)^{\frac{n-1}{\beta-1}} \cdot M \leq (\gamma_\beta)^{\frac{n-1}{\beta-1}} \cdot \lambda_1(L)$$

■

Beweis (zu Satz 7.2.3). Korollar 7.2.6 liefert für die Gitter $L_i = \pi_i(L)$ mit $i = 1, 2, \ldots, n$:

$$\|\widehat{b}_i\| \leq (\gamma_\beta)^{\frac{n-i}{\beta-1}} \lambda_1(L_i) \tag{7.8}$$

Ferner ist $\lambda_1(L_i) \leq \lambda_i(L)$, denn es gibt i linear unabhängige Gittervektoren v, deren Länge höchstens $\lambda_i(L)$ ist und von denen ein Vektor $\pi_i(v) \neq 0$ erfüllt. Also:

$$\lambda_1(L_i) \leq \pi_i(v) \leq \lambda_i(L)$$

Wir erhalten die erste Behauptung, daß für $i = 1, 2, \ldots, n$ gilt:

$$\frac{\|\widehat{b}_i\|^2}{\lambda_i(L)^2} \leq (\gamma_\beta)^{2 \cdot \frac{n-i}{\beta-1}}$$

Aus (7.8), $\mu_{i,j}^2 \leq \frac{1}{4}$ (die Basis ist längenreduziert) und $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_j$ folgt:

$$\begin{aligned}\|b_i\|^2 &= \|\widehat{b}_i\|^2 + \sum_{j=1}^{i-1} (\mu_{i,j})^2 \, \|\widehat{b}_j\|^2 \\ &\leq (\gamma_\beta)^{2\cdot\frac{n-i}{\beta-1}} \cdot \lambda_i(L)^2 + \tfrac{1}{4} \sum_{j=1}^{i-1} (\gamma_\beta)^{2\cdot\frac{n-j}{\beta-1}} \cdot \lambda_j(L)^2 \\ &\leq (\gamma_\beta)^{2\cdot\frac{n}{\beta-1}} \cdot \left((\gamma_\beta)^{2\cdot\frac{-i}{\beta-1}} + \tfrac{1}{4} \sum_{j=1}^{i-1} (\gamma_\beta)^{2\cdot\frac{-j}{\beta-1}} \right) \cdot \lambda_i(L)^2\end{aligned}$$

Wir schätzen die Summanden durch $(\gamma_\beta)^{2\cdot\frac{-1}{\beta-1}}$ nach oben ab und erhalten:

$$\begin{aligned}\|b_i\|^2 &\leq (\gamma_\beta)^{2\cdot\frac{n}{\beta-1}} \cdot (\gamma_\beta)^{2\cdot\frac{-1}{\beta-1}} \cdot \left(1 + \frac{i-1}{4}\right) \cdot \lambda_i(L)^2 \\ &\leq (\gamma_\beta)^{2\cdot\frac{n-1}{\beta-1}} \cdot \frac{i+3}{4} \cdot \lambda_i(L)^2\end{aligned}$$

Damit haben wir die zweite Behauptung auch gezeigt. ■

Beweis (zu Satz 7.2.4). Nach Definition der sukzessiven Minima gilt:

$$\lambda_i^2 \leq \max_{j=1,2,\ldots,i} \|b_j\|^2$$

Aus

$$\|b_i\|^2 = \|\widehat{b}_i\|^2 + \sum_{j=1}^{i-1} (\mu_{i,j})^2 \, \|\widehat{b}_j\|^2$$

und $\mu_{i,j}^2 \leq \frac{1}{4}$ folgt:

$$\|b_i\|^2 \leq \left(1 + (i-1)\cdot\tfrac{1}{4}\right) \cdot \max_{j=1,2,\ldots,i} \|\widehat{b}_j\|^2$$

Wir erhalten:

$$\lambda_i^2 \leq \tfrac{i+3}{4} \cdot \max_{j=1,2,\ldots,i} \|\widehat{b}_j\|^2 \tag{7.9}$$

Lemma 7.2.5 angewandt auf die β-reduzierte Basis $\pi_j(b_j), \pi_j(b_{j+1}), \ldots, \pi_j(b_i)$ liefert für $1 \leq j \leq i - \beta + 1$:

$$\|\widehat{b}_j\| \leq (\gamma_\beta)^{\frac{i-j}{\beta-1}} \max\left\{ \|\widehat{b}_{i-\beta+2}\|, \|\widehat{b}_{i-\beta+3}\|, \ldots, \|\widehat{b}_i\| \right\} \tag{7.10}$$

Andererseits gilt für $i - \beta + 2 \leq j \leq i$:

$$\|\widehat{b}_j\| \leq \|\pi_j(b_j)\| \leq \|b_i\| \tag{7.11}$$

Aus (7.10) und (7.11) erhalten wir für $1 \leq j \leq i$:

$$\|\widehat{b}_j\| \leq (\gamma_\beta)^{\frac{i-j}{\beta-1}} \, \|b_i\|$$

Diese Ungleichung liefert mit (7.9) die Behauptung, daß für $i = 1, 2, \ldots, n$ gilt:

$$\frac{\lambda_i(L)^2}{\|b_i\|^2} \leq (\gamma_\beta)^{2\cdot\frac{i-1}{\beta-1}} \cdot \frac{i+3}{4}$$

■

7.3 Kritische β-reduzierte Basen für $\beta = 2, 3$

In diesem Abschnitt konstruieren wir für $\beta = 2, 3$ kritische β-reduzierte Basen.

Definition 7.3.1 (kritische β-reduzierte Basis)
Eine β-reduzierte Basis $b_1, b_2, \ldots, b_n$ des Gitters L heißt kritisch für n und β, falls $\frac{\|b_1\|}{\lambda_1(L)}$ maximal für alle β-reduzierten Basen vom Rang n ist.

Für $\beta = 2$ konstruieren wir die Basismatrix $A_n := [b_1, b_2, \ldots, b_n] \in M_{n,n}(\mathbb{R})$ wie folgt. Sei $\rho := \sqrt{\frac{3}{4}}$:

$$A_n := \begin{bmatrix} 1 & \frac{1}{2} & 0 & \cdots & \cdots & 0 \\ 0 & \rho & \frac{1}{2}\rho & \ddots & 0 & \vdots \\ \vdots & 0 & \rho^2 & \frac{1}{2}\rho^2 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \rho^{n-2} & \frac{1}{2}\rho^{n-2} \\ 0 & \cdots & \cdots & \cdots & 0 & \rho^{n-1} \end{bmatrix} \tag{7.12}$$

Es gilt

$$A_2 := \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \sqrt{\frac{3}{4}} \end{bmatrix}$$

und für $n \geq 2$ die Rekursion:

$$A_n := \begin{bmatrix} 1 & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & & & & \\ \vdots & & & \rho \cdot A_{n-1} & \\ 0 & & & & \end{bmatrix}$$

Satz 7.3.2
Seien $b_1, b_2, \ldots, b_n$ die Spaltenvektoren der Matrix A_n aus (7.12). Für $\rho = \sqrt{\frac{3}{4}}$ und das Gitter $L = L(b_1, b_2, \ldots, b_n)$ gilt:

a) $b_1, b_2, \ldots, b_n$ ist eine kritische, 2-reduzierte Basis.

b) $\dfrac{\|b_1\|}{\lambda_1(L)} = \dfrac{1}{\rho^{n-2}} = \rho^{-n+2}$

c) $\lambda_1(L) = \rho^{2(n+2)} \left(\frac{1}{4} + \rho^2\right) = \rho^{2(-n+2)}$

Beweis. Siehe [Schnorr94a, Theorem 9]. ■

Für $\beta = 3$ definieren wir die Basismatrix $B_n := [b_1, b_2, \ldots, b_n] \in M_{n,n}(\mathbb{R})$ wie folgt (zur Konstruktion siehe [Schnorr94a]):

$$B_4 := \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{-1}{2\sqrt{3}} & \frac{1}{2\sqrt{3}} \\ 0 & 0 & \sqrt{\frac{2}{3}} & \frac{1}{2}\sqrt{\frac{2}{3}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{7.13}$$

Die Matrizen B_2, B_3 seien die 2×2- bzw. 3×3-Matrizen in der linken, oberen Ecke von B_4. Für $n \geq 4$ definieren wir die Basismatrix B_n rekursiv:

$$B_n := \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{-1}{2\sqrt{3}} & \frac{1}{2\sqrt{3}} & 0 & \cdots & 0 \\ 0 & 0 & & & & & \\ \vdots & \vdots & & \sqrt{\frac{2}{3}} \cdot B_{n-2} & & & \\ 0 & 0 & & & & & \end{bmatrix} \tag{7.14}$$

Es gilt:

Satz 7.3.3
Seien $b_1, b_2, \ldots, b_{2k+1}$ die Spaltenvektoren der Basismatrix B_{2k+1} aus (7.14) *bzw.* (7.13)*. Dann ist $b_1, b_2, \ldots, b_{2k+1}$ für $k = 1, 2, \ldots$ eine kritische, 3-reduzierte Basis.*

Beweis. Siehe [Schnorr94a, Theorem 14]. ■

7.4 Praktisches Verfahren zur β-Reduktion

Wir modifizieren die Bedingung für β-reduziert durch Einführung eines Parameters δ für die Praxis (siehe [SchnEu91, Ritter97]):

Definition 7.4.1 ((β, δ)-reduzierte Basis)
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine geordnete Basis, $\beta \in \{2, 3, \ldots, n\}$ und δ mit $\frac{1}{4} < \delta < 1$ gegeben. Die geordnete Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ heißt (β, δ)-reduziert, wenn:

a) $|\mu_{i,j}| \leq \frac{1}{2}$ *für* $1 \leq j < i \leq n$

b) $\delta \cdot \|\widehat{b}_j\|^2 \leq \lambda_1 \left(\pi_j(L(b_j, b_{j+1}, \ldots, b_k)\right)^2$ *für* $j = 1, 2, \ldots, n$

Der Algorithmus 7.4.1 aus [SchnEu91] transformiert eine gegebene Basis in eine β-reduzierte Basis des gleichen Gitters. Das Unterprogramm L^3FP zur LLL-Reduktion für Gleitkommazahlen haben wir auf Seite 71 angegeben. Das Unterprogramm ENUM (Algorithmus 8.1.1) bzw. GAUSS-ENUM (zur geschnittenen Aufzählung) stellen wir im nächsten Kapitel vor.

Die Variable j wird zyklisch durch die Zahlen $1, 2, \ldots, n-1$ geschoben. Die Variable z zählt die Zahl der Positionen j, welche die Ungleichung

$$\delta \cdot \|\widehat{b}_j\|^2 \leq \lambda_1 \Big(\pi_j\big(L(b_j, b_{j+1}, \ldots, b_k)\big)\Big)^2 \tag{7.15}$$

Algorithmus 7.4.1 Block-Korkine-Zolotareff-Reduktion (kurz BKZ)

EINGABE: ▷ Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{Z}^n$,
▷ $\delta \in \left]\frac{1}{2}; 1\right[$
▷ $\beta \in \{3, 4, \ldots, n-1\}$

1. $L^3FP(b_1, b_2, \ldots, b_n, \delta)$, $z := 0$, $j := 0$

2. WHILE $(z < n-1)$ DO

2.1. $j := j + 1$

$k := \min(j + \beta - 1, n)$

IF $j = n$ THEN $j := 1$, $k := \beta$

/* $b_j, b_{j+1}, \ldots, b_k$ ist LLL-reduziert mit δ */

2.2. ENUM(j, k)

/* Finde Minimalstelle $(u_j, u_{j+1}, \ldots, u_k) \in \mathbb{Z}^{k-j+1} \setminus \{0\}$ zu:

$$c_j(u_j, u_{j+1}, \ldots, u_k) := \sum_{s=j}^{k} \sum_{s=i}^{k} (u_i \mu_{i,s})^2 c_s = \left\| \pi_j \left(\sum_{i=j}^{k} u_i b_i \right) \right\|^2$$

und $b_j^{\text{neu}} := \sum_{s=j}^{k} u_s b_s$. Sei $\overline{c}_j$ der Minimalwert. */

2.3. $h := \min(k+1, n)$

2.4. IF $\delta c_j > \overline{c}_j$ THEN

2.4.1. Ergänze $b_1, b_2, \ldots, b_{j-1}, b_j^{\text{neu}}$ zur Basis von $L(b_1, b_2, \ldots, b_h)$

2.4.2. $L^3FP\left(b_1, b_2, \ldots, b_{j-1}, b_j^{\text{neu}}, b_{j+1}^{\text{neu}}, \ldots, b_h^{\text{neu}}, \delta\right)$ /* Stufe j mit F :=true */

2.4.3. $z := 0$

ELSE

2.4.1. $z := z + 1$

2.4.2. $L^3FP(b_1, b_2, \ldots, b_h, \delta)$ /* auf Stufe $h-1$ */

END if

END while

AUSGABE: (β, δ)-reduzierte Basis $b_1, b_2, \ldots, b_n$

erfüllen. Falls diese Ungleichung nicht für j gilt, fügen wir den Vektor b_j^{neu} in die Basis ein, rufen den LLL-Algorithmus auf und setzen $z = 0$. Den Fall $j = n$ überspringen wir, da (7.15) dann trivialerweise gilt.

Offenbar ist eine Basis $b_1, b_2, \ldots, b_n$ (β, δ)-reduziert, falls sie längenreduziert ist und $z = n - 1$ gilt. Da vor Terminierung der LLL-Algorithmus aufgerufen wird, ist die ausgegebene Basis längenreduziert. Einen Beweis, daß der Algorithmus in polynomieller Zeit arbeitet, gibt es bisher nicht. In der Praxis [SchnEu91, LARIFARI] hat sich der Algorithmus jedoch bewährt.

Kapitel 8

Konstruktion eines kürzesten Gittervektors

Im Kapitel 7.4 haben wir einen Algorithmus zur Block-Reduktion vorgestellt. Als Unterprogramm mußte der kürzeste, nicht-triviale Gittervektor berechnet werden. Wir lernen in diesem Kapitel einen solchen Algorithmus kennen, der durch vollständige Aufzählung einen kürzesten Gittervektor findet. Anschließend werden wir durch die Volumen-Heuristik versuchen, die Aufzählung zu beschränken. Polynomialzeit-Verfahren sind nicht bekannt.

8.1 Algorithmus mit vollständiger Aufzählung

Wir möchten zu einer gegebenen Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ bezüglich der Euklidischen Norm einen kürzesten Gittervektor konstruieren. Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ die Basis mit zugehörigem Orthogonalsystem $\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n$ und Gram-Schmidt-Koeffizienten $\mu_{i,j}$, also:

$$b_i = \sum_{j=1}^{i} \mu_{i,j} \widehat{b}_j \qquad i = 1, 2, \ldots, n,$$

Zur orthogonalen Projektion $\pi_i : \mathbb{R}^m \to \operatorname{span}(b_1, b_2, \ldots, b_i)^{\perp}$ bezeichne:

$$c_t(u_t, u_{t+1}, \ldots, u_n) := \left\| \pi_t \left(\sum_{i=t}^{n} u_i b_i \right) \right\|^2 = \left\| \sum_{i=t}^{n} \sum_{s=t}^{n} u_i \mu_{i,s} \widehat{b}_s \right\|^2 = \sum_{s=t}^{n} \left(\sum_{i=t}^{n} u_i \mu_{i,s} \right)^2 \cdot \|\widehat{b}_s\|^2$$

C.P. Schnorr und M. Euchner stellen in [SchnEu91] den ENUM-Algorithmus (Algorithmus 8.1.1) vor. Die Funktion $a' := \operatorname{next}(a, r)$ liefert zu $a \in \mathbb{Z}$ und $r \in \mathbb{R}$ die in der Reihenfolge nach a betragsmäßig nächste, ganze Zahl zur reellen Zahl r (siehe Grafik 8.1.1). Es gilt:

- $|a - r| \leq |a' - r| \leq |a - r| + 1$
- $\operatorname{sign}(a' - r) \neq \operatorname{sign}(a - r)$

Falls es zu r zwei ganze Zahlen mit Abstand $\frac{1}{2}$ gibt, fordern wir zusätzlich, daß zunächst der kleinere Wert gewählt wird, also aus $|a - r| = |a' - r|$ folgt $a < r < a'$.

Die Korrektheit des ENUM-Algorithmus' 8.1.1 folgt aus den folgenden Beobachtungen:

- Stets gilt: $\tilde{c}_t = c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$. Beweis durch Induktion über Anzahl der Iterationen. Durch die Zuweisungen im ersten Schritt gelten die Behauptungen vor der ersten Iteration

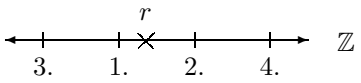

Abbildung 8.1.1: Reihenfolge der Approximationen bei next(· , r)

Algorithmus 8.1.1 ENUM: kürzester Gittervektor (vollständige Aufzählung)

EINGABE: $\|\widehat{b}_i\|^2, \mu_{i,t}$ für $1 \le t \le i \le n$

1. FOR $i = 1, 2, \ldots, n$ DO $\tilde{c}_i := u_i := \tilde{u}_i := y_i := 0$

2. $\tilde{u}_1 := u_1 := 1$; $t := 1$;

3. $c_1^{\min} := \tilde{c}_1 := \|\widehat{b}_1\|^2$

/∗ stets gilt: $\tilde{c}_t = c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$ und $c_1^{\min}$ ist aktuelles Minimum der Funktion c_1 ∗/

4. WHILE $t \le n$ DO

4.1. $\tilde{c}_t := \tilde{c}_{t+1} + (y_t + \tilde{u}_t)^2 \cdot \|\widehat{b}_t\|^2$

4.2. IF $\tilde{c}_t < c_1^{\min}$ THEN

IF $t > 1$ THEN

$t := t - 1$

$y_t := \sum_{i=t+1}^{n} \tilde{u}_i \mu_{i,t}$

$\tilde{u}_t := \lfloor -y_t \rceil$

ELSE

$c_1^{\min} = \tilde{c}_1$

FOR $i = 1, 2, \ldots, n$ DO $u_i := \tilde{u}_i$

END if

ELSE

$t := t + 1$

/∗ $t_{\max}$ bezeichne den bisherigen maximalen Wert von t vor Erhöhung ∗/

$$\tilde{u}_t := \begin{cases} \tilde{u}_t + 1 & \text{falls } t = t_{\max} \\ \text{next}(\tilde{u}_t, -y_t) & \text{sonst} \end{cases}$$

END if

END while

AUSGABE: Minimalstelle $(u_1, u_2, \ldots, u_n) \in \mathbb{Z}^n \setminus \{0\}$ und Minimalwert $c_1^{\min}$ für Funktion c_1

(Induktionsverankerung). Induktionsschluß:

$$\begin{aligned} c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n) &= \underbrace{c_{t+1}(\tilde{u}_{t+1}, \tilde{u}_{t+2}, \ldots, \tilde{u}_n)}_{\text{nach Ind.Annahme: } = \tilde{c}_{t+1}} + \left(\sum_{i=t}^{n} \tilde{u}_i \mu_{i,t} \right)^2 \cdot \|\widehat{b}_t\|^2 \\ &= \tilde{c}_{t+1} + (y_t + \tilde{u}_t)^2 \cdot \|\widehat{b}_t\|^2 \end{aligned}$$

Bei der letzten Umformung nutzen wir, daß $y_t = \sum_{i=t+1}^{n} \tilde{u}_i \mu_{i,t}$ und $\tilde{u}_t = \tilde{u}_t \cdot 1 = \tilde{u}_t \cdot \mu_{t,t}$.

- Der Algorithmus zählt (engl. enumerate) in Depth-First-Order alle Vektoren

$$(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n) \in \mathbb{Z}^{n-t+1} \setminus \{0\}$$

für $t = 1, 2, \ldots, n$ auf, für die gilt ($c_1^{\min}$ ist das aktuelle Minimum der Funktion c_1):

$$c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n) < c_1^{\min}$$

Alle Vektoren erfüllen $\tilde{u}_i > 0$ für das größte i mit $\tilde{u}_i \neq 0$.

- Für feste $\tilde{u}_{t+1}, \tilde{u}_{t+2}, \ldots, \tilde{u}_n$ gilt für die Folge der $\tilde{u}_t$-Werte, erzeugt durch next($\cdot, -y_t$), daß $c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$ monoton wachsend ist. Falls die Abfrage $\tilde{c}_t < \overline{c}$ negativ für den aktuellen Vektor $(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$ ist, kann die Aufzählung der Vektoren $(u, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$, wobei u für die weiteren Werte der next($\cdot, -y_t$)-Funktion steht, entfallen. Denn nach den vorherigen Überlegungen führen diese Vektoren nicht zum Minimum der c_1-Funktion.

Am Ende des ENUM-Algorithmus gilt $c_1^{\min} = \lambda_1^2$.

8.2 Algorithmus mit geschnittener Aufzählung

Um die Laufzeit des Aufzählungsverfahrens zu verkürzen, führen wir eine Heuristik ein, um die Aufzählung abzubrechen, wenn wir in diesem Teil der Aufzählung mit hoher Wahrscheinlichkeit keinen kürzeren Vektoren finden werden (vergleiche [SchnHö95, Ritter97]).

8.2.1 Volumen-Heuristik und Gauß-ENUM

Folgende Heuristik geht auf C.F. Gauß zurück:

Lemma 8.2.1 (Volumen-Heuristik)
Sei $S \subseteq \mathrm{span}(L)$ Jordan-meßbar, $z \in \mathrm{span}(L)$ zufällig, dann gilt:

$$\mathrm{E}_z[\,|(z+S) \cap L|\,] = \frac{\mathrm{vol}(S)}{\det L}$$

Beweis. Denn:

$$\frac{1}{\det L} = \frac{\text{Anzahl Gitterpunkte}}{\text{Volumen Grundmasche}} = \frac{\text{Anzahl Gitterpunkte}}{\text{Volumeneinheit}}$$

■

Angenommen, wir haben $(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n) \in \mathbb{Z}^{n-t+1} \setminus \{0\}$ fest und suchen

$$\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1} \in \mathbb{Z}$$

mit $c_1(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n) < c_1^{\min}$. Setze:

$$\overline{L} := L(b_1, b_2, \ldots, b_{t-1})$$

Wir möchten zu gegebenem Gittervektor $b = \sum_{i=t}^{n} \tilde{u}_i b_i$ einen Vektor

$$\overline{b} = \sum_{i=1}^{t-1} \tilde{u}_i b_i \in \overline{L}$$

addieren, so daß $\left\|b+\overline{b}\right\|^2 < c_1^{\min}$. Wir zerlegen b in orthogonale Anteile:

$$b = \underbrace{\sum_{j=1}^{t-1}\sum_{i=1}^{n} \tilde{u}_i \mu_{i,j} \widehat{b}_j}_{=:-z \in \mathrm{span}(\,\overline{L}\,)} + \underbrace{\sum_{j=t}^{n}\sum_{i=t}^{n} \tilde{u}_i \mu_{i,j} \widehat{b}_j}_{=:y \in \mathrm{span}(\,\overline{L}\,)^{\perp}} \tag{8.1}$$

Das bedeutet, wir suchen nach einem Gitterpunkt $\overline{b} \in \overline{L}$ in

$$\left(b+\overline{L}\,\right) \cap S_{t-1}\left(\sqrt{c_1^{\min} - \tilde{c}_t}, y\right) = \overline{L} \cap S_{t-1}\left(\sqrt{c_1^{\min} - \tilde{c}_t}, y\right),$$

wobei $S_d(r,c)$ eine d-dimensionale Sphäre mit Radius r und Zentrum c ist. Grafik 8.2.1 verdeutlicht die Aufgabe. Wir wenden die Volumen-Heuristik auf das Gitter $\overline{L}$ und die Sphäre

$$S_{t-1}\left(\sqrt{c_1^{\min} - \tilde{c}_t}, y\right) \subseteq \mathrm{span}\left(\,\overline{L}\,\right)$$

an und erhalten:

$$\mathrm{E}_z\left[\,\left|\left\{(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) \in \mathbb{Z}^{t-1} \;:\; c_1(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) < c_1^{\min}\right\}\right|\right] = \frac{\mathrm{vol}\left(S_{t-1}\left(\sqrt{c_1^{\min} - \tilde{c}_t}, y\right)\right)}{\det \overline{L}}$$

Wir werden die Anwendung der Volumen-Heuristik anschließend rechtfertigen.

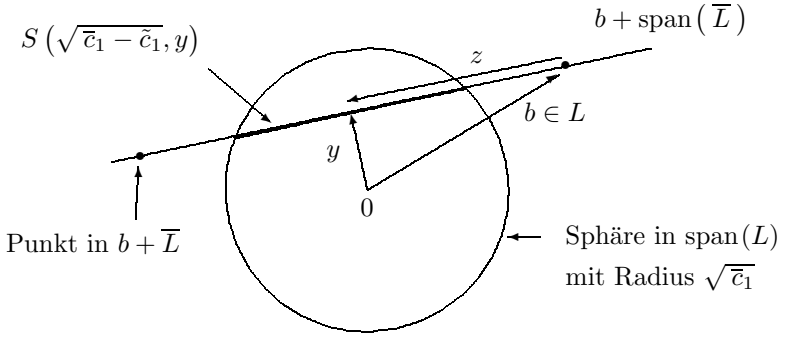

Abbildung 8.2.1: Volumenheuristik bei Gauß-ENUM

Wir beenden die weitere Aufzählung (engl. pruning), falls der Quotient kleiner als 2^{-p}, p fest vorgegeben, ist. Je größer p, desto umfangreicher die Aufzählung. Für $p = \infty$ erhalten wir die vollständige Aufzählung. Wähle $\eta_{t,p}$ als

$$2^{-p} = \frac{\mathrm{vol}\left(S_{t-1}\left(\sqrt{\eta_{t,p}}\right)\right)}{\det \overline{L}}.$$

Aus der Stirling'schen Approximation (siehe (**??**) auf Seite **??**) erhalten wir:

$$2^{-p} = \frac{\mathrm{vol}\left(S_{t-1}\left(\sqrt{\eta_{t,p}}\right)\right)}{\det \overline{L}} = \frac{\left(\pi \cdot \eta_{t,p}\right)^{\frac{t-1}{2}}}{\Gamma\left(1+\frac{t-1}{2}\right)} \cdot \frac{1}{\prod_{i=1}^{t-1} \|\widehat{b}_i\|} \approx \frac{\left(\frac{2e\pi}{t-1} \cdot \eta_{t,p}\right)^{\frac{t-1}{2}}}{\sqrt{\pi(t-1)} \cdot \prod_{i=1}^{t-1} \|\widehat{b}_i\|}$$

Es folgt:

$$\eta_{t,p} = \frac{1}{\pi} \cdot \Gamma\left(1+\frac{t-1}{2}\right)^{\frac{2}{t-1}} \cdot \left(2^{-p} \cdot \prod_{i=1}^{t-1} \|\widehat{b}_i\|\right)^{\frac{2}{t-1}} \approx \frac{t-1}{2e\pi} \left(\sqrt{\pi(t-1)} \cdot 2^{-p} \cdot \prod_{i=1}^{t-1} \|\widehat{b}_i\|\right)^{\frac{2}{t-1}}$$

Im Algorithmus 8.1.1 (Seite 92) ersetzen wir Schritt 4.2 durch

$$\text{IF } \tilde{c}_t < c_1^{\min} - \eta_{t,p}$$

Den erhaltenen Algorithmus 8.2.1 nennen wir Gauß-ENUM. Eine Analyse, mit welcher Wahrscheinlichkeit der kürzeste Gittervektor gefunden wird, erweist sich als schwierig.

Algorithmus 8.2.1 Gauß-ENUM: kürzester Gittervektor (geschnittene Aufzählung)

EINGABE: $\|\widehat{b}_i\|^2, \mu_{i,t}$ für $1 \le t \le i \le n$

1. FOR $i = 1, 2, \ldots, n$ DO $\quad \tilde{c}_i := u_i := \tilde{u}_i := y_i := 0$

2. $\tilde{u}_1 := u_1 := 1$; $t := 1$;

3. $c_1^{\min} := \tilde{c}_1 := \|\widehat{b}_1\|^2$

/* stets gilt: $\tilde{c}_t = c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$ und $c_1^{\min}$ ist aktuelles Minimum der Funktion c_1 */

4. WHILE $t \le n$ DO

4.1. $\tilde{c}_t := \tilde{c}_{t+1} + (y_t + \tilde{u}_t)^2 \cdot \|\widehat{b}_t\|^2$

4.2. IF $\tilde{c}_t < c_1^{\min} - \eta_{t,p}$ THEN

IF $t > 1$ THEN

$t := t - 1$

$$y_t := \sum_{i=t+1}^{n} \tilde{u}_i \mu_{i,t}$$

$\tilde{u}_t := \lfloor -y_t \rceil$

ELSE

$c_1^{\min} = \tilde{c}_1$

FOR $i = 1, 2, \ldots, n$ DO $u_i := \tilde{u}_i$

END if

ELSE

$t := t + 1$

/* $t_{\max}$ bezeichne den bisherigen maximalen Wert von t vor der Erhöhung */

$$\tilde{u}_t := \begin{cases} \tilde{u}_t + 1 & \text{falls } t = t_{\max} \\ \text{next}(\tilde{u}_t, -y_t) & \text{sonst} \end{cases}$$

END if

END while

AUSGABE: Wahrscheinliche Minimalstelle $(u_1, u_2, \ldots, u_n) \in \mathbb{Z}^n \setminus \{0\}$ und Minimalwert $c_1^{\min}$ für die Funktion c_1

Wir müssen noch die Anwendung der Volumen-Heuristik rechtfertigen.

Lemma 8.2.2
Sei $\overline{L} := L(b_1, b_2, \ldots, b_{t-1})$, $(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n) \in \mathbb{Z}^{n-t+1} \setminus \{0\}$ *fest und der Punkt* z *aus* (8.1).

Dann gilt, sofern $z \in_{\mathrm{R}} \operatorname{span}\left(\overline{L}\right)$*:*

$$\mathrm{E}_z\left[\left|\left\{(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) \in \mathbb{Z}^{t-1} \;:\; c_1(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n) \le c_1^{\min}\right\}\right|\right] = \frac{\operatorname{vol}\left(S_{t-1}\left(\sqrt{c_1^{\min} - \tilde{c}_t}, z\right)\right)}{\det \overline{L}}$$

Weiter:

$$\mathrm{E}_z\left[\left|\left\{(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) \in \mathbb{Z}^{t-1} \;:\; c_1(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n) \le c_1^{\min}\right\}\right|\right] \ge 2^{-p} \qquad \Longleftrightarrow \qquad \tilde{c}_t \le c_1^{\min} - \eta_{t,p}$$

Beweis. Wir setzen:

$$S := S_{t-1}\left(\sqrt{c_1^{\min} - \eta_{t,p}}\right)$$

und $L := \overline{L}$. Es gilt:

$$|(z+S) \cap L| = \left|\left\{(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) \in \mathbb{Z}^{t-1} \;:\; c_1(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n) \le c_1^{\min}\right\}\right|$$

Aus Lemma 8.2.1 folgt die Behauptung, da nach Voraussetzung $z \in_{\mathrm{R}} \operatorname{span}\left(\overline{L}\right)$. ■

Zu $r \in \mathbb{R}$ bezeichne $\{r\} \in [0,1[$ die Nachkommastellen.

Definition 8.2.3 (gleichverteilt mod L (kurz: u.d. mod L))
Für ein Gitter L mit der Basis $b_1, b_2, \ldots, b_n$ heißt eine Wahrscheinlichkeitsverteilung der Punkte $\sum_{i=1}^{n} t_i b_i$ in $\operatorname{span}(L)$ gleichverteilt (uniformly distributed) modulo L (kurz: u.d. mod L), falls der Vektor $(\{t_1\}, \{t_2\}, \ldots, \{t_n\})$ gleichverteilt auf $[0,1[^n$ ist.

Bemerkung 8.2.4
Diese Eigenschaft bleibt bei Basiswechsel und Übergang zum Orthogonalsystem erhalten. In den Lemmata 8.2.1 und 8.2.2 genügt es, daß z gleichverteilt mod L ist, denn:

$$z \equiv \overline{z} \pmod{L} \quad \Rightarrow \quad |z + S \cap L| = |\overline{z} + S \cap L|$$

Wir wenden Lemma 8.2.2 und Bemerkung 8.2.4 auf die Situation in ENUM an mit:

- festen $\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n$,
- $\tilde{c}_t = c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$,
- $c_1^{\min} > \tilde{c}_t$,
- wir suchen einen Gitterpunkt aus $\overline{L}$ mit der Sphäre $\left(\sqrt{c_1^{\min} - \tilde{c}_t}, z\right)$ und Zentrum:

$$z = -\sum_{j=1}^{t-1} \sum_{i=t}^{n} \tilde{u}_i \mu_{i,j} \widehat{b}_j$$

Satz 8.2.5
Sei $(\{\mu_{i,j}\} \mid 1 \le j < i \le n)$ gleichverteilt in $[0,1[^{\binom{n}{2}}$. Dann folgt in obiger Situation für festes $(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n) \in \mathbb{Z}^{n-t-1}$:

- *z ist uniformly distributed modulo $\overline{L}$.*
- $\mathrm{E}_z\left[\left|\left\{(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) \in \mathbb{Z}^{t-1} \;:\; c_1(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n \le c_1^{\min}\right\}\right|\right]$

Beweis. Wir nehmen an, daß $\tilde{u}_n \neq 0$, denn sonst können wir n vermindern. Die Eigenschaft $\{\mu\} \in_{\mathrm{R}} [0,1[$ bleibt bei Multiplikation mit einer ganzen Zahl $z \in \mathbb{Z} \setminus \{0\}$ erhalten, d.h. $\{z \cdot \mu\} \in_{\mathrm{R}} [0,1[$. Die Vektoren

$$\begin{gathered}(\{\tilde{u}_n \mu_{n,j}\} \mid j = 1,2,\ldots,t-1) \\ \left(\left\{\sum\nolimits_{i=t}^{n} \tilde{u}_n \mu_{i,j}\right\} \mid j = 1,2,\ldots,t-1\right)\end{gathered}$$

sind unabhängig und gleichverteilt in $[0,1[^{t-1}$. Daraus folgt, daß

$$\left(\left\{\tilde{u}_n \mu_{n,j} - \sum\nolimits_{i=t}^{n} \tilde{u}_n \mu_{i,j}\right\} \mid j = 1,2,\ldots,t-1\right)$$

gleichverteilt in $[0,1[^{t-1}$ ist. Die Behauptung folgt aus Lemma 8.2.2. ■

8.3 Bemerkung zur LLL-reduzierten Basis

Der folgende Satz zeigt, daß es Verteilungen von Gitterbasen gibt, für die die LLL-Reduktion nur sehr schwache Approximationen des kürzesten Gittervektors liefert (vergleiche Satz 5.1.4 auf Seite 57):

Satz 8.3.1
Sei $b_1, b_2, \ldots, b_n$ zufällige Gitterbasis, so daß:

$$(\mu_{i,j} \mid 1 \leq i < j \leq n) \in_{\mathrm{R}} [0,1[^{\binom{n-1}{2}}$$

Es gibt eine Verteilung D auf $[0,1[^{\binom{n-1}{2}}$, so daß die Gitterbasis $b_1, b_2, \ldots, b_n$ mit

$$(\mu_{i,j} \mid 1 \leq i < j \leq n) \in_{D} [0,1[^{\binom{n-1}{2}}$$

stets LLL-reduziert ist und für den Erwartungswert gilt:

$$\mathrm{E}_D\left[\frac{\lambda_1^2}{\|b_1\|^2}\right] \leq \gamma_n \cdot \left(\frac{11}{12}\right)^{\frac{n-1}{2}},$$

wobei γ_n die Hermite-Konstante der Dimension n ist.

Beweis. Sei:

$$(\mu_{i,j} \mid 1 \leq i < j \leq n) \in_{\mathrm{R}} \left[-\tfrac{1}{2}; +\tfrac{1}{2}\right[^{\binom{n-1}{2}}$$

Definiere Basis $b_1, b_2, \ldots, b_n$ des Gitters L durch das zugehörige Orthogonalsystem: Wähle $\|\widehat{b}_i\|^2$, $i = 1,2,\ldots,n$, mit:

$$\begin{aligned}\|\widehat{b}_1\| &= 1 \\ \|\widehat{b}_{i+1}\|^2 &= \|\widehat{b}_i\|^2 \cdot (1 - \mu_{i+1,i}^2) \qquad i = 1,2,\ldots,n-1\end{aligned}$$

Die Basis ist LLL-reduziert mit $\delta = 1$, da nach Konstruktion:

$$\|\pi_i(b_{i+1})\|^2 = \|\widehat{b}_{i+1}\|^2 + \mu_{i+1,i}^2 \|\widehat{b}_i\| = \|\widehat{b}_i\|^2$$

Aus der Definition der Hermite-Konstanten $\gamma_n := \sup\left\{ \frac{\lambda_1(L)^2}{(\det L)^{2/n}} \;\middle|\; \mathrm{Rang}(L) = n \right\}$ folgt:

$$\lambda_1^2 \leq \gamma_n \cdot (\det L)^{\frac{2}{n}} = \gamma_n \cdot \prod_{i=1}^{n} \|\widehat{b}_i\|^{\frac{2}{n}} = \gamma_n \cdot \|\widehat{b}_1\|^{\frac{2}{n}} \cdot \prod_{i=1}^{n-1} (1-\mu_{i+1,i}^2)^{\frac{n-i}{n}}$$

Wir erhalten für x gleichverteilt in $\left[-\frac{1}{2}; +\frac{1}{2}\right[$:

$$\begin{aligned}
\mathrm{E}_x\left[\frac{\lambda_1^2}{\|b_1\|^2}\right] &\leq \gamma_n \cdot \mathrm{E}\left[\prod_{i=1}^{n-1}(1-\mu_{i+1,i}^2)^{\frac{n-i}{n}}\right] && \text{(für } i \neq j \text{ sind } \mu_{i+1,i} \text{ und } \mu_{j+1,j} \text{ unabhängig)} \\
&= \gamma_n \cdot \prod_{i=1}^{n-1} \mathrm{E}_x\left[1-x^2\right]^{\frac{n-i}{n}} \\
&\leq \gamma_n \cdot \prod_{i=1}^{n-1} \mathrm{E}_x\left[1-x^2\right]^{\frac{n-i}{n}} && \left(\text{es gilt: } \mathrm{E}_x\left[1-x^2\right] = 1-\tfrac{1}{12} = \tfrac{11}{12}\right) \\
&\leq \gamma_n \cdot \left(\frac{11}{12}\right)^{\frac{1}{n}\sum_{i=1}^{n-1} i} \\
&\leq \gamma_n \cdot \left(\frac{11}{12}\right)^{\frac{1}{n}\binom{n-1}{2}} \\
&\leq \gamma_n \cdot \left(\frac{11}{12}\right)^{\frac{n-3}{2}}
\end{aligned}$$

Im letzten Schritt nutzen wir die folgende Abschätzung (da $\frac{11}{12} < 1$, müssen wir für eine obere Schranke den Exponenten nach unten abschätzen):

$$\frac{1}{n} \cdot \binom{n-1}{2} = \frac{(n-1)(n-2)}{2n} = \frac{n^2-3n+2}{2n} = \frac{n}{2} - \frac{3}{2} + \frac{1}{n} > \frac{n-3}{2}$$

■

Kapitel 9

Gitterreduktion in beliebiger Norm

Bisher haben wir Gitter bezüglich der Euklidischen Norm reduziert. In diesem Kapitel betrachten wir allgemeine Normen. Besonders die sup-Norm ist von Interesse (siehe Kapitel 10). Bis auf den Gauß-Reduktionsalgorithmus aus Kapitel 4 für aus zwei Vektoren bestehende Basen ist die Reduktion in beliebiger Norm in der Praxis „schwierig“.

9.1 Grundbegriffe

Sei $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ eine beliebige Norm, d.h. es gilt für alle $u, v \in \mathbb{R}^m$ und $\mu \in \mathbb{R}$:

$$\begin{aligned} \|\mu v\| &= |\mu| \cdot \|v\| && \text{(positive Homogenität)} \\ \|u+v\| &\le \|u\| + \|v\| && \text{(Dreiecksungleichung)} \\ \|u\| &\ge 0 \qquad \text{für } u \neq 0 && \text{(positive Definitheit)} \end{aligned}$$

Wir definieren zu einer gegebenen fest geordneten Gitterbasis Abstandsfunktionen:

Definition 9.1.1 (Abstandsfunktion F_i)
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine fest geordnete Gitterbasis. Die i-te Abstandsfunktion (auch Höhen- oder Distanzfunktion) für $1 \le i \le n$

$$F_i : \operatorname{span}(b_1, b_2, \ldots, b_n) \to \mathbb{R}$$

ist bezüglich der gegebenen Norm $\|\cdot\|$ definiert als:

$$\begin{aligned} F_1(x) &:= \|x\| \\ F_i(x) &:= \min_{t_1,t_2,\ldots,t_{i-1}\in\mathbb{R}} \left\| x - \sum\nolimits_{j=1}^{i-1} t_j b_j \right\| = \min_{t\in\mathbb{R}} F_{i-1}\left(x - t b_{i-1}\right) \qquad i = 2, 3, \ldots, n \end{aligned}$$

Die Höhe F_i eines Vektors ist sein Abstand zu dem von $b_1, b_2, \ldots, b_{i-1}$ erzeugten Unterraum. Es gilt $F_i(x) = 0$ genau dann, wenn $x \in \operatorname{span}(b_1, b_2, \ldots, b_{i-1})$. Man rechnet leicht nach, daß jede Abstandsfunktion F_i eine Norm auf $\operatorname{span}(b_1, b_2, \ldots, b_{i-1})^\perp$ ist. Im Fall der Euklidischen Norm ist $F_i(b_i) = \|\widehat{b}_i\|_2$. Die Determinante des Gitters $L = L(b_1, b_2, \ldots, b_n)$ ist:

$$\det L = \prod_{i=1}^{n} \|\widehat{b}_i\|_2$$

Wie sieht die Gleichung $\det L = \prod_i \|\widehat{b}_i\|$ bezüglich $F_1, F_2, \ldots, F_n$ aus? Zu gegebener Norm $\|\cdot\|$ definieren wir:

$$S_{\|\cdot\|}(1) := \{x \in \mathbb{R}^m \ : \ \|x\| \leq 1\}$$

Diese Menge ist konvex, nullsymmetrisch und abgeschlossen. Zu gegebener Norm $\|\cdot\|$ und Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ definiere

$$V_i := \mathrm{vol} \underbrace{\{x \in \mathrm{span}\,(b_1, b_2, \ldots, b_i) \ : \ \|x\| \leq 1\}}_{=\mathrm{span}(b_1,b_2,\ldots,b_i) \cap S_{\|\cdot\|}(1)} \qquad i = 1, 2, \ldots, n \tag{9.1}$$

Man beachte, daß sich Volumen stets auf die Euklidische Metrik bezieht.

Lemma 9.1.2
Für jede Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ gilt:

$$\frac{2^n}{n!} \cdot \prod_{i=1}^{n} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)} \leq V_n \leq 2^n \cdot \prod_{i=1}^{n} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)}$$

Bevor wir das Lemma beweisen, eine Folgerung: Da V_n unabhängig von der Basis ist, gilt:

Korollar 9.1.3
Seien $b_1, b_2, \ldots, b_n$ und $b'_1, b'_2, \ldots, b'_n$ Basen des Gitters L, dann gilt:

$$\prod_{i=1}^{m} F_i(b_i) \leq n! \prod_{i=1}^{m} F'_i(b'_i) \qquad \textit{oder} \qquad \prod_{i=1}^{m} F'_i(b'_i) \leq n! \prod_{i=1}^{m} F_i(b_i)$$

Beweis (zu Lemma 9.1.2). Wir zeigen durch Induktion über n:

$$\frac{2^n}{n!} \cdot \prod_{i=1}^{n} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)} \leq V_n \leq 2^n \cdot \prod_{i=1}^{n} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)}$$

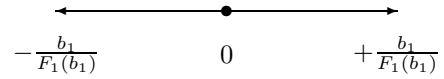

Abbildung 9.1.1: Induktionsverankerung im Beweis zu Lemma 9.1.2

- Induktionsverankerung $n = 1$: Es gilt (vergleiche Abbildung 9.1.1):

$$V_1 = 2 \cdot \frac{\|b_1\|_2}{\|b_1\|} = 2 \cdot \frac{\|\widehat{b}_1\|_2}{F_1(b_1)}$$

- Induktionsschluß von $n-1$ auf n: Wir wählen einen Punkt $z = b_n - \sum_{i=1}^{n-1} t_i b_i \in \mathbb{R}^n$ mit $\|z\| = F_n(b_n)$ (siehe Abbildung 9.1.2). Man erhält eine obere Schranke für V_n durch:

$$V_n \leq 2 \cdot V_{n-1} \cdot \frac{\|\widehat{b}_n\|_2}{F_n(b_n)}$$

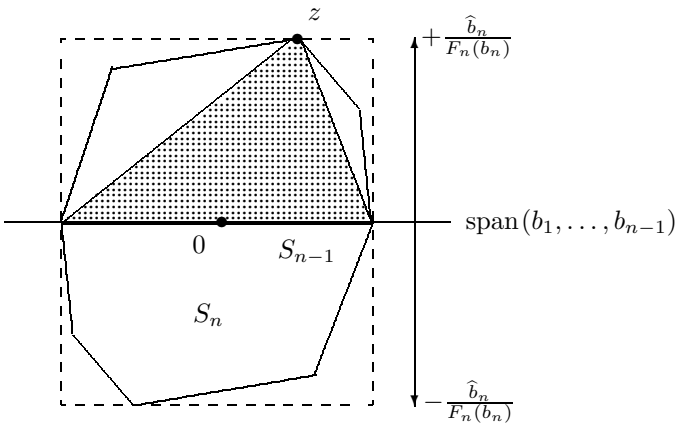

Abbildung 9.1.2: Induktionsschluß im Beweis zu Lemma 9.1.2

Die konvexe Hülle von S_{n-1} und z (gepunktetes Gebiet in Abbildung 9.1.2) ist in S_n enthalten. Da es sich um eine Pyramide mit Grundfläche S_{n-1} und Höhe $\frac{\|\widehat{b}_n\|_2}{F_n(b_n)}$ handelt, gilt:

$$\frac{V_{n-1}}{n} \cdot \frac{\|\widehat{b}_n\|_2}{F_n(b_n)} = \mathrm{vol}_n\,(\text{konvexe Hülle von } S_{n-1} \text{ und } z) \leq \frac{V_n}{2}$$

Es folgt wegen der Symmetrie:

$$2 \cdot \frac{V_{n-1}}{n} \cdot \frac{\|\widehat{b}_n\|_2}{F_n(b_n)} \leq V_n$$

Aus der Induktionsannahme erhalten wir:

$$V_n \geq 2 \cdot \frac{V_{n-1}}{n} \cdot \frac{\|\widehat{b}_n\|}{F_n(b_n)} \geq \left[\frac{2}{n} \cdot \frac{\|\widehat{b}_n\|}{F_n(b_n)}\right] \cdot \left[\frac{2^{n-1}}{(n-1)!} \cdot \prod_{i=1}^{n-1} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)}\right] = \frac{2^n}{n!} \cdot \prod_{i=1}^{n} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)}$$

Für die obere Schranke betrachten wir den Quader mit Grundfläche S_{n-1} und Höhe $\frac{\|\widehat{b}_n\|}{F_n(b_n)}$. Da S_n in zwei dieser Quader enthalten ist (siehe Abbildung 9.1.2), gilt:

$$V_n \leq 2 \cdot V_{n-1} \cdot \frac{\|\widehat{b}_n\|_2}{F_n(b_n)}$$

Aus der Induktionsannahme erhalten wir:

$$V_n \leq 2 \cdot V_{n-1} \cdot \frac{\|\widehat{b}_n\|_2}{F_n(b_n)} \leq 2 \cdot \frac{\|\widehat{b}_n\|_2}{F_n(b_n)} \cdot 2^{n-1} \cdot \prod_{i=1}^{n-1} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)} = 2^n \cdot \prod_{i=1}^{n} \frac{\|\widehat{b}_i\|_2}{F_i(b_i)}$$

■

Es gilt für das erste sukzessive Minimum:

Satz 9.1.4 (Kaib 1994)
Für jede Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ gilt:

$$\min_{i=1,2,\ldots,n} F_i(b_i) \leq \lambda_{1,\|\cdot\|} \leq \left(n! \cdot \prod_{i=1}^{n} F_i(b_i)\right)^{\frac{1}{n}}$$

Zum Vergleich für die ℓ_2-Norm: Wir wissen aus der Minkowski'schen Ungleichung 3.2.11 (Seite 45), daß wegen $\lambda_{1,\|\cdot\|} \leq \lambda_{2,\|\cdot\|} \leq \cdots \leq \lambda_{n,\|\cdot\|}$ für das Gitter $L = L(b_1, b_2, \ldots, b_n)$ gilt:

$$\min_{i=1,2,\ldots,n} \|\widehat{b}_i\| \leq \lambda_{i,\ell_2} \leq (\gamma_n)^{\frac{1}{2}} \cdot (\det L)^{\frac{1}{n}}$$

Beweis (zu Satz 9.1.4). Betrachten wir beide Abschätzungen:

- Wir zeigen:

  $$\min_{i=1,2,\ldots,n} F_i(b_i) \leq \lambda_{i,\|\cdot\|}$$

  Sei $b = \sum_{i=1}^{n} t_i b_i \in L$ mit $\|b\| = \lambda_{1,\|\cdot\|}$. Setze $s := \max\{i \mid t_i \neq 0\}$. Wegen $t_s \in \mathbb{Z} \setminus \{0\}$ gilt:

  $$\lambda_{1,\|\cdot\|} = \|b\| \geq F_s(b) = F_s(t_s b_s) = \underbrace{|t_s|}_{\geq 1} \cdot F_s(b_s) \geq F_s(b_s)$$

  Die Behauptung folgt aus:

  $$\min_{i=1,2,\ldots,n} F_i(b_i) \leq F_s(b_s) \leq \lambda_{i,\|\cdot\|}$$

- Sei $L = L(b_1, b_2, \ldots, b_n)$. Aus dem zweiten Satz von Minkowski 3.3.9 (Seite 49) folgt wegen $\lambda_{1,\|\cdot\|} \leq \lambda_{2,\|\cdot\|} \leq \cdots \leq \lambda_{n,\|\cdot\|}$:

  $$V_n \cdot \lambda_{1,\|\cdot\|}^n \leq 2^n \cdot \det L \tag{9.2}$$

  Aus Lemma 9.1.2 wissen wir:

  $$\frac{n!}{2^n} \cdot \prod_{i=1}^{n} \frac{F_i(b_i)}{\|\widehat{b}_i\|_2} \geq \frac{1}{V_n} \tag{9.3}$$

  Aus $\det L = \prod_{i=1}^{n} \|\widehat{b}_i\|_2$ erhalten wir:

  $$\begin{aligned}
  \lambda_{1,\|\cdot\|}^n &\leq 2^n \cdot V_n^{-1} \cdot \prod_{i=1}^{n} \|\widehat{b}_i\|_2 && \text{(wegen (9.2))} \\
  &\leq 2^n \cdot \left(\frac{n!}{2^n} \cdot \prod_{i=1}^{n} \frac{F_i(b_i)}{\|\widehat{b}_i\|_2} \right) \cdot \left(\prod_{i=1}^{n} \|\widehat{b}_i\|_2 \right) && \text{(wegen (9.3))} \\
  &= n! \cdot \prod_{i=1}^{n} F_i(b_i)
  \end{aligned}$$

■

Es gilt für das Produkt der sukzessiven Minima:

Satz 9.1.5
Für jede Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ gilt:

$$\frac{1}{n!} \cdot \prod_{i=1}^{n} F_i(b_i) \leq \prod_{i=1}^{n} \lambda_{i,\|\cdot\|} \leq n! \cdot \prod_{i=1}^{n} F_i(b_i)$$

Zum Vergleich: Die Minkowski'sche Ungleichung für die ℓ_2-Norm, Satz 3.2.11 auf Seite 45, besagt:

$$\prod_{i=1}^{n} \lambda_{i,\ell_2} \leq (\gamma_n)^{\frac{n}{2}} \cdot \det L$$

Beweis (zu Satz 9.1.5). Die Behauptung folgt aus dem Beweis zu Satz 9.1.4 durch Anwenden des zweiten Satzes von Minkowski 3.3.9 auf Seite 49:

$$\frac{\det L}{n!} \leq \frac{V_n}{2^n} \cdot \prod_{i=1}^{n} \lambda_{i,\|\cdot\|} \leq \det L$$

■

9.2 Reduzierte Basen zur Norm $\|\cdot\|$

Analog zur Euklidischen Norm führen wir Reduktionsbegriffe ein und versuchen, Eigenschaften reduzierter Basen zu beweisen.

9.2.1 Definitionen

Wir übertragen die Reduktionsbegriffe auf den Fall einer beliebig vorgegebenen Norm:

Definition 9.2.1 (HKZ-reduzierte Basis zu $\|\cdot\|$)
Eine geordnete Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ ist eine HKZ-reduzierte Basis zur Norm $\|\cdot\|$, wenn:

a) $F_j(b_i) \leq F_j(b_i \pm b_j)$ *für* $1 \leq j < i \leq n$ *(längenreduziert)*

b) $F_i(b_i) = \min\{F_i(b) \mid b \in L(b_i, b_{i+1}, \ldots, b_n) \setminus \{0\}\}$ *für* $i = 1, 2, \ldots, n$

Beim zweiten Kriterium kann b auch aus $L(b_1, b_2, \ldots, b_n) \setminus \{0\}$ gewählt werden. Für die Eigenschaft „längenreduziert“ gilt mit $1 \leq j < i \leq n$:

$$F_j(b_i) \leq F_j(b_i \pm b_j) \qquad \Longleftrightarrow \qquad F_j(b_i) = \min_{t \in \mathbb{Z}} F_j\left(b_i + t \cdot b_j\right)$$

Diese Äquivalenz nutzt die Konvexität der Norm F_j. Die „$\Leftarrow$“-Richtung folgt unmittelbar und für die „$\Rightarrow$“-Richtung beachtet man, daß gilt:

$$F_j(b_i) \leq F_j(b_i - b_j) \qquad \text{und} \qquad F_j(b_i) \leq F_j(b_i + b_j)$$

Definition 9.2.2 (β-reduzierte Basis zu $\|\cdot\|$)
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine geordnete Basis und $\beta \in \{2, 3, \ldots, n\}$ gegeben. $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ heißt β-reduziert (blockreduziert mit Blockgröße β) zur Norm $\|\cdot\|$, wenn:

a) $F_j(b_i) \leq F_j(b_i \pm b_j)$ *für* $1 \leq j < i \leq n$

b) $F_i(b_i) = \min\left\{F_i(b) \;\middle|\; b \in L(b_i, b_{i+1}, \ldots, b_{\min(i+\beta-1,n)}) \setminus \{0\}\right\}$ *für* $i = 1, 2, \ldots, n-1$

Wir betrachten den Spezialfall einer 2-reduzierten Basis zu $\|\cdot\|$: Die geordnete Basis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ ist 2-reduziert zur Norm $\|\cdot\|$, wenn:

a) $F_j(b_i) \leq F_j(b_i \pm b_j)$ für $1 \leq j < i \leq n$

b) $F_i(b_i) = \min\left\{F_i(sb_i + tb_{i+1}) \;\middle|\; (s,t) \in \mathbb{Z}^2 \setminus \{(0,0)\}\right\}$ für $i = 1, 2, \ldots, n-1$

Eine 2-reduzierte Basis zur ℓ_2-Norm ist eine LLL-reduzierte Basis.

9.2.2 Eigenschaften 2-reduzierter Gitterbasen

Wir untersuchen die Eigenschaften 2-reduzierter Basen und vergleichen die Resultate, die wir im Spezialfall der ℓ_2-Norm (LLL-reduziert) in Kapitel 5.1 (Seite 55 und folgende) bewiesen haben.

Satz 9.2.3
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine 2-reduzierte Basis zur Norm $\|\cdot\|$. Dann gilt für $i = 1, 2, \ldots, n-1$:

$$F_{i+1}(b_{i+1}) \geq \tfrac{1}{2} \cdot F_i(b_i)$$

Zum Vergleich: Für die ℓ_2-Norm ist mit $\delta = \frac{3}{4}$ nach Lemma 5.1.2 auf Seite 56 $\|\widehat{b}_i\|_2^2 \leq 2 \cdot \|\widehat{b}_{i+1}\|_2^2$, also:

$$\|\widehat{b}_{i+1}\|_2 \geq \sqrt{\tfrac{1}{2}} \cdot \|\widehat{b}_i\|_2$$

Beweis (zu Satz 9.2.3). Nach Definition gilt

a) $F_j(b_i) \leq F_j(b_i \pm b_j)$ für $1 \leq j < i \leq n$ und

b) $F_i(b_i) = \min\left\{F_i(sb_i + tb_{i+1}) \;\middle|\; (s,t) \in \mathbb{Z}^2 \setminus \{(0,0)\}\right\}$ für $i = 1, 2, \ldots, n-1$.

Die Behauptung $F_i(b_i) \leq \frac{1}{2} \cdot F_i(b_{i+1})$ erhalten wir aus:

$$\begin{aligned} F_i(b_i) &\leq F_i(b_{i+1}) && \text{(wegen Eigenschaft b))} \\ &\leq \min\{F_i(b_{i+1} - tb_i) \mid t \in \mathbb{Z}\} && \text{(wegen Eigenschaft a))} \\ &\leq F_i(b_{i+1}) + \tfrac{1}{2} \cdot F_i(b_i) \end{aligned}$$

Wir nutzen, daß die Abstandsfunktionen F_i jeweils Normen auf $\operatorname{span}(b_1, b_2, \ldots, b_{i-1})^\perp$ sind:

$$\begin{aligned} F_{i+1}(b_{i+1}) &= \min\{F_i(b_{i+1} - sb_i) \mid s \in \mathbb{R}\} && \text{(Definition)} \\ &= \min\left\{F_i\big(b_{i+1} - (r+t)b_i\big) \;\middle|\; t \in \mathbb{Z}, r \in \left[-\tfrac{1}{2}, +\tfrac{1}{2}\right]\right\} \\ &\leq \min\{F_i(b_{i+1} - tb_i) \mid t \in \mathbb{Z}\} + F_i\left(\tfrac{1}{2} \cdot b_i\right) && \text{(Dreiecksungleichung)} \\ &\leq \min\{F_i(b_{i+1} - tb_i) \mid t \in \mathbb{Z}\} + \tfrac{1}{2} \cdot F_i(b_i) && \text{(Linearität)} \end{aligned}$$

■

Im folgenden Satz untersuchen wir, wie gut im allgemeinen Fall der erste Vektor der 2-reduzierten Basis das erste sukzessive Minimum approximiert.

Satz 9.2.4
Sei $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$ eine 2-reduzierte Basis zur Norm $\|\cdot\|$. Dann gilt:

$$\|b_1\| \leq 2^{n-1} \cdot \lambda_{1,\|\cdot\|}$$

Zum Vergleich: Für die ℓ_2-Norm wissen wir mit $\delta = \frac{3}{4}$ aus Satz 5.1.4 auf Seite 57:

$$\|b_1\|_2 \leq \left(\tfrac{4}{3}\right)^{\frac{n-1}{2}} \cdot \lambda_{1,\ell_2}$$

Beweis. Sei $b = \sum_{i=1}^{n} t_i b_i$ $\|\cdot\|$-minimaler Vektor in $L = (b_1, b_2, \ldots, b_n) \setminus \{0\}$. O.B.d.A. sei $t_n \in \mathbb{Z} \setminus \{0\}$. Es gilt:

$$\begin{aligned}
\|b\| &\geq F_n(b) \\
&= \min_{t_1,t_2,\ldots,t_{n-1}\in\mathbb{R}} \left\| b - \sum\nolimits_{j=1}^{n-1} t_j b_j \right\| && \text{(Definition)} \\
&= F_n(t_n b_n) \\
&= |t_n| \cdot F_n(b_n) && \text{(Linearität der Norm } F_n\text{)} \\
&\geq F_n(b_n) && \text{(wegen } t_n \in \mathbb{Z} \setminus \{0\}\text{)} \\
&\geq 2^{-n+1} \cdot F_1(b_1) && \text{(induktiv aus Satz 9.2.3)} \\
&= \|b_1\| \cdot 2^{-n+1} && \text{(wegen } F_1(b_i) = \|b_i\| \text{ und } \widehat{b}_1 = b_1\text{)}
\end{aligned}$$

Wegen $\lambda_{1,\|\cdot\|} = \|b\|$ folgt die Behauptung. ■

9.2.3 Eigenschaften HKZ-reduzierter Basen

Wir untersuchen die Eigenschaften von HKZ-Basen und vergleichen die Resultate, die wir im Spezialfall der ℓ_2-Norm in Kapitel 7.1 (Seite 81 und folgende) bewiesen haben. Es gilt für HKZ-reduzierte Basen zur Norm $\|\cdot\|$:

Satz 9.2.5 (Lovász, Scarf 1992)
Sei $b_1, b_2, \ldots, b_n$ eine HKZ-reduzierte Basis zu $\|\cdot\|$ des Gitters L. Es gilt für $i = 1, 2, \ldots, n$:

$$\frac{2}{i+1} \cdot \|b_i\| \leq \lambda_{i,\|\cdot\|} \leq \frac{i+1}{2} \cdot F_i(b_i) \leq \frac{i+1}{2} \cdot \|b_i\|$$

Zum Vergleich: Für die ℓ_2-Norm wissen wir aus Satz 7.1.2 auf Seite 82, daß für $i = 1, 2, \ldots, n$ gilt:

$$\frac{i+3}{4} \cdot \|b_i\| \leq \lambda_{i,\ell_2} \leq \frac{i+1}{4} \|b_i\|$$

Beweis (zu Satz 9.2.5). Wir zeigen die untere und obere Schranke:

- Wir zeigen $\frac{2}{i+1} \cdot \|b_i\| \leq \lambda_{i,\|\cdot\|}$ für $i = 1, 2, \ldots, n$. Angenommen, $h_1, h_2, \ldots, h_n \in L$ realisieren die sukzessiven Minima $\lambda_1, \lambda_2, \ldots, \lambda_n$, d.h. es ist $\|h_i\| = \lambda_{i,\|\cdot\|}$ für $i = 1, 2, \ldots, n$, und die Vektoren $h_1, h_2, \ldots, h_n$ sind linear unabhängig. Es gilt

  $$\max_{j \leq i} F_i(h_j) \geq F_i(b_i), \tag{9.4}$$

  weil:

  - wegen $\dim(\operatorname{span}(h_1, h_2, \ldots, h_i)) = i$ ist $\max_{j \leq i} F_i(h_j) \neq 0$ und
  - $b_1, b_2, \ldots, b_n$ eine HKZ-reduzierte Basis ist, also

    $$F_i(b_i) = \min\{F_i(b) \mid b \in L(b_i, \ldots, b_n) \setminus \{0\}\}$$

    gilt.

  Wir erhalten aus (9.4) und $\lambda_{1,\|\cdot\|} \leq \lambda_{2,\|\cdot\|} \leq \cdots \leq \lambda_{n,\|\cdot\|}$:

  $$\lambda_{i,\|\cdot\|} = \|h_i\| = \max_{j \leq i} \|h_i\| \geq F_i(b_i) \tag{9.5}$$

Wir wenden aus Beweis zu Satz 9.2.3 die Ungleichung

$$\min_{\mu\in\mathbb{Z}} F_j(x+\mu\cdot b_j) \le F_{j+1}(x) + \tfrac{1}{2}\cdot F_j(b_j)$$

rekursiv beginnend mit $x := b_i$ und $j = i-1$ an:

$$\begin{aligned} F_{i-1}(b_i) &\le F_{i-1}(b_i + \mu_{i,i-1}\cdot b_{i-1}) && \text{für alle } \mu_{i,i-1}\in\mathbb{Z} \\ &\le F_i(b_i) + \tfrac{1}{2}\cdot F_{i-1}(b_{i-1}) && \text{für Minimalstelle } \mu_{i,i-1}\in\mathbb{Z} \end{aligned}$$

Im nächsten Schritt sei $x := b_i + \mu_{i,i-1}\cdot b_{i-1}$ mit Minimalstelle $\mu_{i,i-1}\in\mathbb{Z}$ und $j := i-1$ usw. Nach $i-1$ Schritten erhalten wir mit Abschätzung (9.5) die Behauptung:

$$\|b_i\| = F_1(b_i) \le F_1\left(b_i + \sum_{j=1}^{i-1}\mu_{i,j}\cdot b_j\right) \le \frac{i+1}{2}\cdot\lambda_{i,\|\cdot\|} \tag{9.6}$$

- Wir zeigen für $i = 1, 2, \ldots, n$:

$$\lambda_{i,\|\cdot\|} \le \frac{i+1}{2}\cdot F_i(b_i) \le \frac{i+1}{2}\cdot\|b_i\|$$

Es gilt:

$$\begin{aligned} \lambda_{i,\|\cdot\|} &\le \max_{j\le i} F_1(b_j) && (\text{wegen } F_1(b) = \|b\|) \\ &\le \max_{j\le i}\left\{F_j(b_j) + \tfrac{1}{2}\cdot\sum_{t=1}^{j-1} F_t(b_t)\right\} && (\text{wegen (9.6)}) \\ &\le \max_{j\le i}\left\{\tfrac{j+1}{2}\cdot F_j(b_j)\right\} && (\text{wegen (9.6)}) \\ &\le \tfrac{i+1}{2}\cdot F_1(b_i) && (\text{wegen } F_t(b_t)\le F_1(b_j) \text{ für } t<j) \\ &\le \tfrac{i+1}{2}\cdot\|b_i\| && (\text{wegen } F_1(b_i) = \|\widehat{b}_i\| \le \|b\|) \end{aligned}$$

■

9.2.4 Eigenschaften β-reduzierter Gitterbasen

Wir untersuchen die Eigenschaften von HRZ-Basen und vergleichen die Resultate, die wir im Spezialfall der ℓ_2-Norm in Kapitel 7.2 (Seite 83 und folgende) bewiesen haben. Wir definieren:

Definition 9.2.6 (α_β)
Wir setzen:

$$\alpha_\beta := \sup\left\{\frac{\|b_1\|}{F_\beta(b_\beta)} \;\middle|\; \begin{array}{l} b_1, b_2, \ldots, b_n \textit{ HKZ-reduzierte} \\ \textit{Basis und } \|\cdot\| \textit{ Norm} \end{array}\right\}$$

Satz 9.2.7
Für jede β-reduzierte Basis $b_1, b_2, \ldots, b_n$ zu $\|\cdot\|$ gilt:

$$\|b_1\| \le \alpha_\beta^{\lceil\frac{n-1}{\beta-1}\rceil}\cdot\lambda_{1,\|\cdot\|}$$

Beweis. Sei $h_i := F_i(b_i)$. Bestimme Index μ mit minimalem h_μ. Nach Satz 9.1.4 gilt $h_\mu \le \lambda_{1,\|\cdot\|}$. Für $j < \beta$ sind die Basen $b_i, b_{i+1} \ldots, b_{i+j}$ HKZ-reduzierte Basen zur Norm F_i. Nach Definition von α_β und wegen $\alpha_k \le \alpha_{k+1}$ gilt:

$$h_i \le \alpha_\beta \cdot h_{i+j} \tag{9.7}$$

Wir erhalten durch wiederholtes Anwenden von (9.7):

$$h_1 \le \alpha_\beta \cdot h_{1+1(\beta-1)} \le \alpha_\beta^2 \cdot h_{1+2(\beta-1)} \le \alpha_\beta^3 \cdot h_{1+3(\beta-1)} \le \ldots \le \alpha_\beta^{\lfloor \frac{\mu-1}{\beta-1} \rfloor} \cdot h_{1+\lfloor \frac{\mu-1}{\beta-1} \rfloor(\beta-1)}$$

Insgesamt erhalten wir:

$$h_1 \le \alpha_\beta^{\lceil \frac{\mu-1}{\beta-1} \rceil} \cdot h_\mu \le \alpha_\beta^{\lceil \frac{n-1}{\beta-1} \rceil} \cdot \lambda_{1,\|\cdot\|}$$

■

Für die ℓ_2-Norm zeigt C.P. Schnorr [Schnorr87, Korollar 2.5], daß für

$$\alpha_{\beta,\ell_2} := \sup \left\{ \frac{\|b_1\|_2}{\|\widehat{b}_\beta\|} \ : \ b_1, b_2, \ldots, b_n \text{ HKZ-reduzierte Basis} \right\} \tag{9.8}$$

gilt, wobei α_{β,ℓ_2} als Quadrat von (9.8) definiert und als Korkine-Zolotareff-Konstante bezeichnet wird):

$$\alpha_{\beta,\ell_2} \le k^{\frac{1+\ln k}{2}}$$

Analog zeigt man für beliebige Norm:

$$\alpha_k \le k(k-1)^{\ln(k-1)}$$

Satz 9.2.8
Jede β-reduzierte Basis $b_1, b_2, \ldots, b_n$ erfüllt für $i = 1, 2, \ldots, n$

$$\frac{2}{i+1} \cdot {\gamma'_\beta}^{-\frac{i-\beta/2}{\beta-1}} \le \frac{\|b_i\|}{\lambda_{i,\|\cdot\|}} \le \frac{i+1}{2} \cdot {\gamma'_\beta}^{\frac{n-\beta/2}{\beta-1}}$$

mit $\gamma'_\beta = (\beta!)^{\frac{2}{\beta}} \approx \left(\frac{\beta}{e}\right)^2$.

Zum Vergleich: In der ℓ_2-Norm gilt für die Hermite-Konstante γ_β nach Satz 7.2.3 auf Seite 84:

$$\sqrt{\frac{4}{i+3}} \cdot \gamma_\beta^{-\frac{i-1}{\beta-1}} \le \frac{\|b_i\|}{\lambda_{i,\ell_2}} \le \sqrt{\frac{i+3}{4}} \cdot \gamma_\beta^{\frac{n-1}{\beta-1}}$$

Der Beweis zu Satz 9.2.8 ist im wesentlichen analog zum Beweis zur ℓ_2-Norm. Wichtiger „Baustein“ ist folgendes Analogon zu Lemma 7.2.5 auf Seite 85: Für jede β-reduzierte Basis $b_1, b_2, \ldots, b_n$ gilt:

$$\|b_1\| \le {\gamma'_\beta}^{\frac{m-\beta/2}{\beta-1}} \cdot M \qquad \text{mit} \quad M := \max_{n-\beta+2 \le i \le n} F_i(b_i)$$

9.3 Konstruktion einer HKZ-reduzierten Gitterbasis

Gegeben sei ein Gitter L vom Rang n. Wir konstruieren eine HKZ-reduzierte Basis in zwei Schritten, wobei die Konstruktion allerdings nicht effizient ist.

1. Wir wählen für $i = 1, 2, \ldots, n$ ein $b_i \in L$ mit:

   $$F_i(b_i) = \min \{ F_i(b) \mid b \in L, F_i(b) \neq 0 \}$$

   Beachte, $F_i(b)$ ist definiert, da wir im i-ten Schritt bereits $b_1, b_2, \ldots, b_{i-1}$ festgelegt haben. Es gilt genau dann $F_i(b) \neq 0$, wenn $b \notin \operatorname{span}(b_1, b_2, \ldots, b_{i-1})$. Die Vektoren $b_1, b_2, \ldots, b_n$ bilden eine Basis von L: Falls dies nicht der Fall ist, existiert ein minimales i, so daß $b_1, b_2, \ldots, b_i$ kein primitives System ist:

   $$L \cap \operatorname{span}(b_1, b_2, \ldots, b_{i-1}) = L(b_1, b_2, \ldots, b_{i-1}) \tag{9.9}$$

   $$L \cap \operatorname{span}(b_1, b_2, \ldots, b_i) \supsetneq L(b_1, b_2, \ldots, b_i) \tag{9.10}$$

   Wegen 9.10 existiert ein $b \in L \cap \operatorname{span}(b_1, b_2, \ldots, b_i) \setminus L(b_1, b_2, \ldots, b_i)$ mit:

   $$b = \sum_{j=1}^{i-1} t_j b_j + t_i b_i \qquad \text{mit } t_1, t_2, \ldots, t_{i-1} \in \mathbb{Z} \text{ und } t_i \notin \mathbb{Z}$$

   Sei $k > 1$ der Index der additiven Untergruppe $L(b_1, b_2, \ldots, b_i)$ in $\operatorname{span}(b_1, b_2, \ldots, b_i) \cap L$. Wegen (9.9) gilt:

   $$t_i \in \tfrac{1}{k} + \mathbb{Z}$$

   Wähle $t' = t_i \mod \mathbb{Z}$, d.h. $t' = \frac{1}{k} \in]0, 1[$. Es folgt der Widerspruch zur Minimalität:

   $$F_i(t_1 b_1 + t_2 b_2 + \cdots + t' b_i) = F_i(t' b_i) = |t'| \cdot F_i(b_i) < F_i(b_i).$$

2. Längenreduktion: Für $i = 1, 2, \ldots, n$, für $j = i-1, i-2, \ldots, 1$ wähle $\mu_{i,j} \in \mathbb{Z}$, so daß:

   $$F_j(b_i + \mu_{i,i-1} b_{i-1} + \mu_{i,i-2} b_{i-2} + \cdots + \mu_{i,j} b_j)$$

   minimal ist. Setze:

   $$b_i := b_i + \sum_{j=1}^{i-1} \mu_{i,j} b_j$$

   Die Längenreduktion sichert $F_j(b_j) \leq F_j(b_i \pm b_j)$ für $j < i$.

Lemma 9.3.1
Obige Konstruktion liefert eine HKZ-reduzierte Basis $b_1, b_2, \ldots, b_n$ des Gitters L.

Beweis. Nachrechnen! ■

9.4 Alternative zur Reduktion in $\|\cdot\|$

Alternativ zur Reduktion in $\|\cdot\|$ kann man $S_{\|\cdot\|}(1)$ durch $S_{\|\cdot\|_{\mathrm{E}}}(1)$ mit Ellipsoid-Norm $\|\cdot\|_{\mathrm{E}}$ approximieren und die Reduktion in der Ellipsoid-Norm durchführen.

$$\|x\|_{\mathrm{E}}^2 := x^{\mathsf{T}} B^{\mathsf{T}} B x$$

Die Sätze für die ℓ_2-Norm übertragen sich. Nach [John48] gilt: Zu jeder Norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ gibt es eine Ellipsoid-Norm $\|\cdot\|_{\mathrm{E}}$ mit

$$\|x\|_{\mathrm{E}} \leq \|x\| \leq \sqrt{n} \cdot \|x\|_{\mathrm{E}}$$

Dann folgt für die $\|\cdot\|_{\mathrm{E}}$ β-reduzierte Basis nach Satz 7.2.3:

$$\frac{1}{n} \cdot \sqrt{\frac{4}{i+3}} \cdot \gamma_\beta^{-\frac{i-1}{\beta-1}} \leq \frac{\|b_i\|}{\lambda_{i,\|\cdot\|}} \leq n \cdot \sqrt{\frac{i+3}{4}} \cdot \gamma_\beta^{\frac{n-1}{\beta-1}} \tag{9.11}$$

Dabei geht ein Faktor $\sqrt{n}$ verloren bei der Approximation von $\|\cdot\|$ durch $\|\cdot\|_{\mathrm{E}}$. Ein weiterer Faktor $\sqrt{n}$ geht verloren durch die Approximation von $\lambda_{i,\|\cdot\|}$ durch λ_{i,ℓ_2}.

Für kleine Blockweiten β ist die Aussage (9.11) schärfer als Satz 9.2.8, weil $\gamma'_\beta = \Theta\left(\gamma_\beta^2\right)$. Für große Blockweiten $\beta \approx n$ ist Satz 9.2.8 schärfer. Für $\beta = n$ sind die Schranken für HKZ-reduzierte Basen zu $\|\cdot\|$ um den Faktor $\sqrt{n}$ besser als die Schranken (9.11).

9.5 Konstruktion eines $\|\cdot\|$-kürzesten Gittervektors

Wir übertragen unseren Algorithmus zur Bestimmung eines kürzesten Gittervektors für die Euklidische Norm aus Kapitel 8.1 auf beliebige Normen. H. Ritters [Ritter96, Ritter97] gibt eine Übersicht über die Aufzählung kürzester Gittervektoren in der sup-Norm.

9.5.1 ENUM-Algorithmus für beliebige Norm

Wir verallgemeinern Algorithmus 8.1.1 von Seite 92. Es bezeichne:

$$c_t(u_t, u_{t+1}, \ldots, u_n) := F_t\left(\sum_{i=t}^{n} u_i b_i\right)$$

Wir bezeichnen zu $\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n$ mit $\mathrm{next}_{F_t}(u)$ die erste, ganzzahlige Minimalstelle u' von

$$\left| F_t\left(u \cdot b_t + \sum_{i=t}^{n} \tilde{u}_i b_i\right) - F_t\left(u' \cdot b_t + \sum_{i=t}^{n} \tilde{u}_i b_i\right) \right| \tag{9.12}$$

und mit $\mathrm{next}_{F_t}(\tilde{u}_t, u)$ die nächste, ganzzahlige Nullstelle von (9.12) nach $\tilde{u}_t$. Falls $S_{\|\cdot\|}$ ein Polytop ist, z.B. für die 1- und sup-Norm, kann F_t durch lineare Optimierung bestimmt werden.

9.5.2 Gauß-ENUM-Algorithmus für beliebige Norm

Betrachten wir Schritt 2 des Algorithmus' 9.5.1. Gegeben sind $b_1, b_2, \ldots, b_n$ sowie $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$ und $c_1^{\min}$. Sei $\overline{L} := L(b_1, b_2, \ldots, b_{t-1})$ und setze:

$$z := -\sum_{i=t}^{n} \tilde{u}_i b_i$$

Dann gilt:

$$\begin{aligned} \left|\left\{(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) \in \mathbb{Z}^{t-1} \,:\, c_1(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_{t-1}) \leq c_1^{\min}\right\}\right| &= \left|\left(\overline{L} + w\right) \cap S_{\|\cdot\|}(c_1^{\min})\right| \\ &= \left|\overline{L} \cap \left(S_{\|\cdot\|}(c_1^{\min}) + z\right)\right| \end{aligned}$$

Algorithmus 9.5.1 $\|\cdot\|$-ENUM: kürzester Gittervektor (vollständige Aufzählung)

EINGABE: Gitterbasis $b_1, b_2, \ldots, b_n \in \mathbb{R}^m$

1. FOR $i = 1, 2, \ldots, n$ DO $\tilde{c}_i := u_i := \tilde{u}_i := y_i := 0$

2. $\tilde{u}_1 := u_1 := 1$; $t := 1$;

3. $c_1^{\min} := \tilde{c}_1 := \|b_1\|^2$

/* stets gilt: $\tilde{c}_t = c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n)$ und $c_1^{\min}$ ist aktuelles Minimum der Funktion c_1 */

4. WHILE $t \leq n$ DO

4.1. $\tilde{c}_t := c_t(\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_n) = F_t\left(\sum_{i=t}^{n} \tilde{u}_i b_i\right)$

4.2. IF $\tilde{c}_t < c_1^{\min}$ THEN

IF $t > 1$ THEN

$t := t - 1$

u reelle Minimalstelle von $F_t\left(u b_t + \sum_{i=t+1}^{n} \tilde{u}_i b_i\right)$

$\tilde{u}_t := \mathrm{next}_{F_t}(u)$

ELSE

$c_1^{\min} := \tilde{c}_1$

FOR $i = 1, 2, \ldots, n$ DO $u_i := \tilde{u}_i$

END if

ELSE

$t := t + 1$

/* $t_{\max}$ bezeichne den bisherigen maximalen Wert von t vor der Erhöhung */

$$\tilde{u}_t := \begin{cases} \tilde{u}_t + 1 & \text{falls } t = t_{\max} \\ \mathrm{next}_{F_t}(\tilde{u}_t, u) & \text{sonst} \end{cases}$$

END if

END while

AUSGABE: Minimalstelle $(u_1, u_2, \ldots, u_n) \in \mathbb{Z} \setminus \{0\}$ und Minimalwert $c_1^{\min}$ für die Funktion c_1

Nach der Volumenheuristik ist:

$$\begin{aligned} \left|\overline{L} \cap \left[S_{\|\cdot\|}(\, c_1^{\min}) + z\right]\right| &\approx \frac{\mathrm{vol}_{t-1}\left(\mathrm{span}\left(\,\overline{L}\,\right) \cap \left[S_{\|\cdot\|}(\, c_1^{\min}) + z\right]\right)}{\det \overline{L}} \\ &= \frac{\mathrm{vol}_{t-1}\left(\,\left[w + \mathrm{span}\left(\,\overline{L}\,\right)\right] \cap S_{\|\cdot\|}(\, c_1^{\min})\right)}{\det \overline{L}} \end{aligned}$$

Wann gilt die Volumen-Heuristik streng? Hinreichende Voraussetzung: Bei festem y ist z uniformly distributed modulo $\overline{L}$ (vergleiche Definition 8.2.3 auf Seite 96):

$$b = \underbrace{\sum_{j=1}^{t-1}\sum_{i=1}^{n} \tilde{u}_i \mu_{i,j} \widehat{b}_j}_{=-z\in \operatorname{span}(\overline{L})} + \underbrace{\sum_{j=t}^{n}\sum_{i=t}^{n} \tilde{u}_i \mu_{i,j} \widehat{b}_j}_{:=y\in \operatorname{span}(\overline{L})^{\perp}}$$

Die Menge $\left(b + \operatorname{span}\left(\overline{L}\right)\right) \cap S_{\|\cdot\|}(c_1^{\min})$ hängt nur von z, aber nicht von y ab. Es folgt aus der Volumenheuristik Lemma 8.2.1 (Seite 93):

Lemma 9.5.1
Angenommen, z ist uniformly distributed modulo $\overline{L}$ und unabhängig von y. Dann gilt

$$\mathrm{E}\left[\left|\left[w + \overline{L}\right] \cap S_{\|\cdot\|}(c_1^{\min})\right|\right] = \frac{\mathrm{vol}_{t-1}\left(\left[y + \operatorname{span}\left(\overline{L}\right)\right] \cap S_{\|\cdot\|}(c_1^{\min})\right)}{\det \overline{L}},$$

wobei $y + \operatorname{span}\left(\overline{L}\right) = b + \operatorname{span}\left(\overline{L}\right)$.

Wir erhalten analog zu Satz 8.2.5:

Satz 9.5.2
Angenommen, $(\{\mu_{i,j}\} : 1 \le j < i \le n)$ ist gleichverteilt in $[0,1[^{\binom{n}{2}}$. Dann gilt in Algorithmus 9.5.1 $\|\cdot\|$-ENUM stets:

- *z ist uniformly distributed modulo $\overline{L}$ und unabhängig von y.*
- $\mathrm{E}\left[\left|\left[w + \overline{L}\right] \cap S_{\|\cdot\|}(c_1^{\min})\right|\right] = \dfrac{\mathrm{vol}_{t-1}\left(\left[y + \operatorname{span}\left(\overline{L}\right)\right] \cap S_{\|\cdot\|}(c_1^{\min})\right)}{\det \overline{L}}$

Wir erhalten Gauß-$\|\cdot\|$-ENUM aus Algorithmus' 9.5.1, indem wir Schritt 2 ersetzen durch:

$$\text{IF } \frac{\mathrm{vol}_{t-1}\left(\left[y + \operatorname{span}\left(\overline{L}\right)\right] \cap S_{\|\cdot\|}(c_1^{\min})\right)}{\det \overline{L}} \ge 2^{-p}$$

Kapitel 10

Anwendungen der Gitterreduktion

In Kapitel 6 (Seite 73 und folgende) haben wir versucht, Subsetsum-Aufgaben durch Gitterreduktion zu lösen. In diesem Kapitel werden wir weitere Anwendungen der Gitterreduktion kennenlernen: Lösen des 3-SAT-Problems, Angriff auf Dåmgards Hashfunktion (finde zwei verschiedene Vektoren, denen der gleiche Werte zugewiesen wird) und Faktorisieren ganzer Zahlen. Für weitere Anwendungen der Gitterreduktion in der Kryptographie verweisen wir auf die Arbeit [JoSt94] von A. Joux und J. Stern.

Für eine effiziente Aufzählung kürzester Gittervektor in der sup-Norm verweisen wir auf H. Ritters Arbeit [Ritter96], in der er mit Hilfe der Gitterreduktion G. Ortons [Orton1994] Kryptosystem basierend auf dem Subsetsum-Problem mit Dichte größer als 1 bricht. Die Methoden können auf beliebige Normen ℓ_p übertragen werden.

10.1 Gitterbasis zu 3-SAT

Wir beschreiben zunächst die konjunktive Normalform. Seien $x_1, x_2, \ldots, x_n$ Boole'sche Variablen. Wir schreiben $x_i^{-1} := \neg\, x_i$, $x_i^1 := x_i$ und $x_i^0 := 0$. Die Klauseln der konjunktiven Normalform (KNF) schreiben wir als:

$$C_j = x_1^{a_{j1}} \vee x_2^{a_{j2}} \vee \cdots \vee x_n^{a_{jn}}$$

mit $(a_{j1}, a_{j2}, \ldots, a_{jn}) \in \{0, \pm 1\}^n$. Falls eine Variable x_i nicht in der Klausel C_j auftritt, setze $a_{ji} := 0$. Die KNF γ hat folgenden Aufbau

$$\gamma(x_1, x_2, \ldots, x_n) := \bigwedge_{j=1}^{m} C_j(x_1, x_2, \ldots, x_n)$$

Wir betrachten nur konjunktive Normalformen, deren Klauseln aus maximal drei Literalen bestehen, also $\sum_{i=1}^{n} |a_{ji}| \leq 3$ für $j = 1, 2, \ldots, m$. Beim 3-SAT-Problem ist zu entscheiden, ob eine erfüllende Belegung für die konjunktive Normalform existiert:

Definition 10.1.1 (3-SAT)
Das 3-SAT-Problem lautet:

- *Gegeben: KNF* $\gamma(x_1, x_2, \ldots, x_n) := \bigwedge_{j=1}^{m} C_j(x_1, x_2, \ldots, x_n)$ *mit max. 3 Literalen pro Klausel*
- *Finde* $(y_1, y_2, \ldots, y_n) \in \{0, 1\}^n$ *mit* $\gamma(y_1, y_2, \ldots, y_n) = 1$ *oder zeige, daß keine erfüllende Belegung existiert.*

Das 3-SAT-Problem ist $\mathcal{NP}$-vollständig [GaJo79]. Wir ordnen dem 3-SAT-Problem eine Gitterbasis zu und versuchen, durch Gitterreduktion in der sup-Norm eine erfüllende Belegung der konjunktiven Normalform zu bestimmen.

Wir reduzieren zunächst 3-SAT auf $\{0,1\}$-Integer-Programming, indem wir ein äquivalentes Ungleichungssystem bilden:

$$c_j := 2 - |\{i \,:\, a_{ji} = -1\}| \leq 1 \qquad \text{für } j = 1, 2, \ldots, m$$

Betrachte das folgende Ungleichungssystem in den Unbekannten $y_1, y_2, \ldots, y_n \in \{0,1\}$:

$$\left| \sum_{i=1}^{n} a_{ji} y_i - c_j \right| \leq 1 \qquad \text{für } j = 1, 2, \ldots, m \tag{10.1}$$

Beispiel:

$$\begin{aligned} x_1 \vee x_2 \vee x_3 \quad &\leftrightarrow \quad |y_1 + y_2 + y_3 - 2| \leq 1 \\ \neg\, x_1 \vee x_2 \vee x_3 \quad &\leftrightarrow \quad |-y_1 + y_2 + y_3 - 1| \leq 1 \end{aligned}$$

Wir können jede Restriktion in zwei $\leq$-Relationen aufspalten. Durch Fallunterscheidung über die Anzahl negierter/nicht-negierter Literale in der Klausel folgt:

Lemma 10.1.2
Die $\{0,1\}$-IP-Aufgabe (10.1) hat genau dann eine Lösung $y \in \{0,1\}^n$, wenn $\gamma(y) = 1$.

Die Gitterbasis zum 3-SAT-Problem besteht aus den folgenden $n+1$ ganzzahligen Zeilenvektoren $b_1, b_2, \ldots, b_{n+1} \in \mathbb{Z}^{n+m+1}$:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ b_{n+1} \end{bmatrix} := \begin{bmatrix} 2 & 0 & \cdots & 0 & a_{11} & a_{21} & \cdots & a_{m1} & 0 \\ 0 & 2 & & 0 & a_{12} & a_{22} & \cdots & a_{m2} & 0 \\ \vdots & & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & 2 & a_{1n} & a_{2n} & \cdots & a_{mn} & 0 \\ -1 & -1 & \cdots & -1 & -c_1 & -c_2 & \cdots & -c_m & +1 \end{bmatrix} \tag{10.2}$$

Sei $y = (y_1, y_2, \ldots, y_n)$ eine erfüllende Belegung der KNF. Der zugehörige Lösungsvektor ist:

$$b(y) = \sum_{i=1}^{n} y_i b_i + b_{n+1}$$

Dieser Vektor liegt wegen (10.1) in $\{\pm 1\}^n \times \{\pm 1, 0\}^m \times \{+1\}$.

Satz 10.1.3
Sei L das von den Zeilenvektoren $b_1, b_2, \ldots, b_{n+1}$ aus (10.2) erzeugte Gitter. Dann gilt für alle Gittervektoren $z \in L$:

$$\begin{array}{l} \textit{Es existiert eine erfüllende Belegung} \\ y \in \{0,1\}^n \textit{ zu } \gamma(y) \textit{ mit } z = \pm b(y) \end{array} \quad \Longleftrightarrow \quad \|z\|_\infty = 1$$

Beweis. Wir zeigen beide Richtungen:

„⇒“ Wegen $b(y) \in \{\pm 1, 0\}^{n+m+1}$ gilt $\|z\|_\infty = 1$.

„⇐“ Gegeben ist ein Vektor $z \in L$ mit $\|z\|_\infty = 1$. Der Vektor habe die Darstellung

$$z = \sum_{i=1}^{n+1} y'_i b_i \tag{10.3}$$

mit $y'_1, y'_2, \ldots, y'_{n+1} \in \mathbb{Z}$. Wegen $\|z\|_\infty = 1$ folgt aus der letzten Komponente der Basisvektoren, daß $y'_{n+1} = \pm 1$ ist. Aus den ersten n Einträgen erhalten wir nach Fallunterscheidung:

1. Aus $y_{n+1} = +1$ folgt $(y'_1, y'_2, \ldots, y'_n) \in \{0, +1\}^n$.
2. Aus $y_{n+1} = -1$ folgt $(y'_1, y'_2, \ldots, y'_n) \in \{0, -1\}^n$.

Setze:

$$y := y'_{n+1} \cdot (y'_1, y'_2, \ldots, y'_n)$$

Es ist $y \in \{0, +1\}^n$ und $(y, 1) = y'_{n+1} \cdot y'$. Wegen $\|z\|_\infty = 1$ und $y'_{n+1} \in \{\pm 1\}$ gilt nach (10.3)

$$\left| \sum_{i=1}^{n} a_{ji} y_i - c_j \right| = \left| y'_{n+1} \right| \cdot \left| \sum_{i=1}^{n} a_{ji} y_i - c_j \right| \leq \|z\|_\infty \leq 1$$

für $j = 1, 2, \ldots, m$. Nach Lemma 10.1.2 ist y eine erfüllende Belegung.

■

Wir versuchen durch Gitterreduktionen, einen in der sup-Norm kürzesten, nicht-trivalen Gittervektor zu finden, um eine erfüllende Belegung der konjunktive Normalformen mit höchstens drei Literalen pro Klausel zu bestimmen. Unter der Cook'schen Hypothese $\mathcal{P} \neq \mathcal{NP}$ ist dies in einigen Fällen schwierig, denn das 3-SAT-Problem ist $\mathcal{NP}$-vollständig.

Wir haben mit Satz 10.1.3 einen alternativen Beweis zu Korollar 1.2.9 von Seite 11 kennengelernt: Das Problem $\|\cdot\|_\infty$-kürzester Gittervektor ist $\mathcal{NP}$-vollständig.

10.2 Angriff auf Dåmgards Hashfunktion

I.B. Dåmgard [Dåmgard89] hat der EuroCrypt-Konferenz 1989 die folgende kryptographische Hashfunktion basierend auf dem Subsetsum-Problem vorgestellt. Wähle zufällig und unabhängig

$$a = (a_1, a_2, \ldots, a_n) \in_{\mathrm{R}} [1, 2^m - 1]^n$$

und definiere zu a die Hashfunktion:

$$\begin{array}{rcccl} h_a & : & \{0,1\}^n & \to & \mathbb{N} \\ & & (x_1, x_2, \ldots, x_n) & \mapsto & \sum_{i=1}^{n} a_i x_i \end{array}$$

Eine *Kollision* nennen wir $x, x' \in \{0,1\}^n$, $x \neq x'$, mit $h_a(x) = h_a(x')$. Als *Pseudo-Kollision* bezeichnen wir $x, x' \in \{0,1\}^n$, $x \neq x'$, mit $h_a(x) = h_a(x') \pmod{2^m}$. Wir werden versuchen, Pseudo-Kollisionen zu finden.

Weshalb suchen wir nach Kollisionen? Um eine lange Nachricht M durch eine kurze, digitale Unterschrift zu versehen, wendet man in der Kryptographie die Hashfunktion h auf die Nachricht an und erhält einen im Vergleich zur Nachricht kleinen Wert. Nur $h(M)$ wird digital unterschrieben. Der Teilnehmer veröffentlicht M und seine digitale Unterschrift von $h(M)$. Falls wir eine andere Nachricht M' mit $h(M) = h(M')$ finden, können wir die digitale Unterschrift für M einfach übernehmen. Wir haben eine Nachricht mit digitaler Unterschrift eines fremden Teilnehmers.

Jedem Vektor $\overline{x} = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n) \in \{\pm 1, 0\}^n \setminus \{0\}$ mit

$$\sum_{i=1}^{n} a_i \cdot \overline{x}_i = 0 \pmod{2^m}$$

entspricht eine Pseudo-Kollision x, x' gemäß

$$x_i := \begin{cases} 1 & \text{falls } \overline{x}_i = 1 \\ 0 & \text{sonst} \end{cases} \qquad x'_i := \begin{cases} 1 & \text{falls } \overline{x}_i = -1 \\ 0 & \text{sonst} \end{cases}$$

und umgekehrt. Wir wählen als Basis:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ b_{n+1} \end{bmatrix} := \begin{bmatrix} 1 & 0 & \cdots & 0 & a_1 n \\ 0 & 1 & & 0 & a_2 n \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & & 1 & a_n n \\ 0 & 0 & \cdots & 0 & 2^m n \end{bmatrix} \tag{10.4}$$

Wir bezeichnen:

$$\text{P-Kollision} := \left\{ (x_1, x_2, \ldots, x_{n+1}) \in \{\pm 1, 0\}^{n+1} \;\middle|\; \begin{array}{l} x_{n+1} = 0 \text{ und } (x_1, x_2, \ldots, x_n) \\ \text{entspricht Pseudo-Kollision} \end{array} \right\}$$

Es gilt offenbar: P-Kollision $\subseteq L(b_1, b_2, \ldots, b_{n+1})$. Wir versuchen durch Gitterreduktion, einen kurzen Gittervektor in der Euklidischen Norm zu finden. Was ist das Minimum der Menge

$$\{ \|x\|_2 \,:\, x \in \text{P-Kollision} \},$$

also die Länge des kürzesten Gittervektors, der einer Pseudo-Kollision entspricht? Wir führen eine probabilistische Analyse zu $a \in_{\mathrm{R}} [1, 2^m - 1]^n$ und festem x durch. Zu $\alpha \in \left[0, \frac{1}{2}\right]$ mit $\alpha n \in \mathbb{N}$ sei:

$$\mathcal{N}_\alpha := \left\{ (x_1, x_2, \ldots, x_n) \in \{\pm 1, 0\}^n \;\middle|\; \sum_{i=1}^{n} |x_i| = \alpha n \right\}$$

Für $x \in \mathcal{N}_\alpha$ ist $\|x\|_2 = \sqrt{\alpha n}$, da $x \in \{\pm 1, 0\}^n$. Es gilt:

$$N_\alpha := |\mathcal{N}_\alpha| = \binom{n}{\alpha n} \cdot 2^{\alpha n} \tag{10.5}$$

Denn wir können die αn Einträge ungleich 0 beliebig auf die n Positionen verteilen und als Wert jeweils $+1$ oder -1 setzen. Es gilt

$$N_\alpha \approx 2^{(H(\alpha, 1-\alpha)+\alpha) \cdot n}, \tag{10.6}$$

wobei H die Shannon'sche Entropie-Funktion ist:

$$H(\alpha, 1-\alpha) = -\alpha \cdot \log_2 \alpha - (1-\alpha) \cdot \log_2 (1-\alpha)$$

Wir möchten bezüglich $a \in_{\mathrm{R}} [1, 2^m - 1]^n$ die Wahrscheinlichkeit berechnen, mit der in $\mathcal{N}_\alpha$ ein Vektor aus P-Kollision liegt. Dazu definieren wir zu a und festem $x \in \mathcal{N}_\alpha$ die Zufallsvariable:

$$\xi_x := \begin{cases} 1 & \text{falls } \sum\limits_{i=1}^{n} a_i x_i = 0 \pmod{2^m} \\ 0 & \text{sonst} \end{cases}$$

Offenbar ist

$$\mathrm{E}_a [\, \xi_x] = 2^{-m} \tag{10.7}$$

Wir definieren die Zufallsvariable $\overline{\xi}_x := \xi_x - 2^{-m}$, so daß:

$$\mathrm{E}_a\left[\overline{\xi}_x\right] = 0 \tag{10.8}$$

$$\mathrm{E}_a\left[\overline{\xi}_x^2\right] = \mathrm{E}_a\left[\xi_x^2\right] - 2 \cdot 2^{-m} \cdot \mathrm{E}_a[\xi_x] + 2^{-2m} \leq 2 \cdot 2^{-m} \tag{10.9}$$

Beachte, daß für die Indikatorvariable ξ_i gilt $\mathrm{Ws}_a[\xi_i = 1] = \mathrm{Ws}_a\left[\xi_i^2 = 1\right]$. Wir verwenden aus der Stochastik (siehe u.a. [Feller68]) für eine Zufallsvariable X:

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 && \text{(Definition Varianz)}\\
\mathrm{Var}[cX] &= c^2 \cdot \mathrm{Var}[X] && (c > 0 \text{ konstant})\\
\mathrm{Ws}[|X - \mathrm{E}[X]| \geq \epsilon] &\leq \frac{1}{\epsilon^2} \cdot \mathrm{Var}[X] && \text{(Tschebycheff-Ungleichung)}
\end{aligned}$$

Wir wenden die Tschebycheff-Ungleichung auf $\frac{1}{N_\alpha} \cdot \sum_{x \in \mathcal{N}_\alpha} \xi_x$ an und erhalten wegen der Erwartungswerte (10.8) und (10.9):

$$\begin{aligned}
\mathrm{Ws}_a\left[\left|\frac{1}{N_\alpha} \cdot \sum_{x \in \mathcal{N}_\alpha} \xi_x - 2^{-m}\right| \geq \epsilon\right] &\leq \frac{1}{\epsilon^2} \cdot \mathrm{Var}\left[\frac{1}{N_\alpha} \cdot \sum_{x \in \mathcal{N}_\alpha} \xi_x\right]\\
&= \frac{1}{\epsilon^2 \cdot N_\alpha^2} \cdot \sum_{x \in \mathcal{N}_\alpha} \sum_{y \in \mathcal{N}_\alpha} \mathrm{E}_a\left[\overline{\xi}_x \cdot \overline{\xi}_y\right]
\end{aligned}$$

Wegen des Erwartungswerts (10.9) ist:

$$\begin{aligned}
\sum_{x \in \mathcal{N}_\alpha} \sum_{y \in \mathcal{N}_\alpha} \mathrm{E}_a\left[\overline{\xi}_x \cdot \overline{\xi}_y\right] &= \sum_{x \in \mathcal{N}_\alpha} \mathrm{E}_a\left[\overline{\xi}_x^2\right] + \sum_{\substack{x,y \in \mathcal{N}_\alpha\\ x \neq y}} \mathrm{E}_a\left[\overline{\xi}_x\right] \cdot \mathrm{E}_a\left[\overline{\xi}_y\right]\\
&= \sum_{x \in \mathcal{N}_\alpha} \mathrm{E}_a\left[\overline{\xi}_x^2\right] \qquad (10.10)\\
&= N_\alpha \cdot \mathrm{E}_a\left[\overline{\xi}_x^2\right]
\end{aligned}$$

Aus Abschätzung (10.10) und dem Erwartungswert (10.8) folgt:

$$\mathrm{Ws}_a\left[\left|\frac{1}{N_\alpha} \cdot \sum_{x \in \mathcal{N}_\alpha} \xi_x - 2^{-m}\right| \geq \epsilon\right] \leq \frac{1}{\epsilon^2 \cdot N_\alpha} \cdot \mathrm{E}_a\left[\overline{\xi}_x^2\right] \leq \frac{2}{\epsilon^2 \cdot N_\alpha \cdot 2^m}$$

Für $\epsilon = 2^{-m}$ erhalten wir

$$\mathrm{Ws}_a\left[\sum_{x \in \mathcal{N}_\alpha} \xi_x = 0\right] \leq \frac{2^{m+1}}{N_\alpha} \tag{10.11}$$

und für $\epsilon = 2^{-m-l}$:

$$\mathrm{Ws}_a\left[\sum_{x \in \mathcal{N}_\alpha} \xi_x \leq N_\alpha \cdot \left(2^{-m} - 2^{-m-l}\right)\right] \leq \frac{2^{m+1+2l}}{N_\alpha} \tag{10.12}$$

Aus (10.11) folgt unmittelbar:

Satz 10.2.1
Es gilt:

a) Für $m \leq \log_2 N_\alpha - 2 \approx (H(\alpha, 1-\alpha) + \alpha) \cdot n$ gibt es bezüglich $a \in_{\mathrm{R}} [1, 2^m - 1]^n$ mit Wahrscheinlichkeit mindestens $\frac{1}{2}$ Pseudo-Kollisionen in $\mathcal{N}_\alpha$.

b) Für $m \leq \log_2 N_\alpha - 4 \approx (H(\alpha, 1-\alpha) + \alpha) \cdot n$ gibt es bezüglich $a \in_{\mathrm{R}} [1, 2^m - 1]^n$ mit Wahrscheinlichkeit mindestens $\frac{1}{2}$ mindestens $N_\alpha \cdot 2^{-m-1}$ Pseudo-Kollisionen in $\mathcal{N}_\alpha$.

Beweis. Die Aussage a) folgt aus (10.11), die Aussage b) folgt aus (10.12) mit $l = 1$. ■

Im nächsten Schritt möchten wir $N_\alpha = |\mathcal{N}_\alpha|$ maximieren: Aus dem Ansatz

$$\frac{\partial N_\alpha}{\partial \alpha} = 0$$

mit (10.6) erhalten wir:

$$-\log_2 2 \approx \frac{\partial \left(-\alpha \cdot \log_2 \alpha - (1-\alpha) \cdot \log_2 (1-\alpha)\right)}{\partial \alpha}$$

Dazu äquivalent:

$$\log_2 \alpha + \log_2 (1-\alpha) \approx -\log_2 2$$

Wegen $-\log_2 2 = -1$ und $-1 + \log_2 \alpha = \log_2 \frac{1}{\alpha}$ erhalten wir den Anzatz $\alpha = \frac{k-1}{k}$

$$-\log_2 \tfrac{k-1}{k} + \log_2 \tfrac{1}{k} = \log_2 (k-1) \approx -\log_2 2$$

und somit $k - 1 = 2$ bzw. $k = 3$, also $\alpha = \frac{2}{3}$ als ungefähre Maximalsstelle von N_α.

Satz 10.2.2
Bezüglich $a \in_{\mathrm{R}} [1, 2^m - 1]^n$ gibt es mit Wahrscheinlichkeit mindestens $\frac{1}{2}$ Pseudo-Kollisionen, wenn $N_{2/3} \geq 2^{m-1}$ oder äquivalent $n \geq \frac{m-1}{\log_2 3}$ ist.

Beweis. Aus (10.7) wissen wir, daß:

$$\mathrm{E}_a [\,\text{Anzahl Pseudo-Kollisionen in } \mathcal{N}_\alpha] = N_\alpha \cdot 2^{-m}$$

Wegen $N_{2/3} \geq 2^{m-1}$ folgt:

$$\mathrm{Ws}_a \left[\,\text{Anzahl Pseudo-Kollisionen in } \mathcal{N}_{2/3}\right] \geq \tfrac{1}{2}$$

Damit $N_{2/3} \geq 2^{m-1}$ ist, muß wegen

$$\begin{aligned}
\log_2 N_{2/3} &= n \cdot \left[-\tfrac{2}{3} \cdot \log_2 \tfrac{2}{3} - \tfrac{1}{3} \cdot \log_2 \tfrac{1}{3} + \tfrac{2}{3} \cdot \log_2 2\right] \\
&= n \cdot \left[-\tfrac{2}{3} \cdot (-\log_2 3 + \log_2) - \tfrac{1}{3} \cdot (-\log_2 3) + \tfrac{2}{3}\right] \\
&= n \cdot \left[+\tfrac{2}{3} \cdot \log_2 3 - \tfrac{2}{3} + \tfrac{1}{3} \cdot \log_2 3 + \tfrac{2}{3}\right] \\
&= n \cdot \log_2 3
\end{aligned}$$

gelten $n \cdot \log_2 3 \geq m - 1$ oder äquivalent $n \geq \frac{m-1}{\log_2 3}$. ■

Betrachten wir die Situation bei den von I.B. Dåmgard vorgeschlagenen Parametern:

- Für $m = 120$ gibt es Pseudo-Kollisionen, falls $n \geq 77$.
- Für $m = 120$ und $n = 100$ gibt es im Mittel $N_{2/3} \cdot 2^{-m} \approx 3,8 \cdot 10^{11}$ Pseudo-Kollisionen.

Betrachten wir die Anzahl der kurzen Gittervektoren:

$$\left|\left\{z \in L(b_1, b_2, \ldots, b_{n+1}) \ : \ \|z\|_2^2 \leq \alpha n\right\}\right| \approx \frac{N(0, n, \alpha)}{2^m}$$

J.E.Mazo und A.M.Odlyzko haben in [MaOd90] die Funktion

$$N(z, n, \alpha) := \left|\left\{x \in \mathbb{Z}^n \ : \ \|x - z\|^2 \leq \alpha \cdot n\right\}\right|$$

untersucht. Als Vektoren z kommen nur Vektoren in Frage, deren letzter Eintrag 0 ist. Der Anteil der Vektoren $(x_1, x_2, \ldots, x_n)$ mit $\sum_{i=1}^n a_i x_i = 0 \pmod{2^m}$ ist 2^{-m}. Es gilt:

$$\left|\left\{z \in L(b_1, b_2, \ldots, b_{n+1}) \ : \ \|z\|_2^2 \leq \alpha n\right\}\right| \approx \frac{(2e\pi\alpha)^{\frac{n}{2}}}{2^m}$$

Wir wählen m und α derart, daß in etwa gilt

$$\mathrm{E}_a\left[\text{Anzahl Pseudo-Kollisionen in } \mathcal{N}_\alpha\right] \approx 1,$$

Also wegen (10.5) und (10.6):

$$m = n \cdot [H(\alpha, 1-\alpha) + \alpha] \qquad \text{bzw.} \qquad N_\alpha = \binom{n}{\alpha n} \cdot 2^{\alpha n} = 2^m$$

Gibt es zur Länge $\sqrt{\alpha n}$ kürzere, nicht-triviale Gittervektoren, die keiner Pseudo-Kollision entsprechen? Wir würden bei der Gitter-Reduktion unter Umständen diese kürzeren Vektoren anstatt der gewünschten erhalten. Für $m = 120$ und $n = 100$ ist $\frac{(2e\pi\alpha)^{\frac{n}{2}}}{2^m} \approx 2^{0{,}0039n}$, d.h. die sog. parasitären, kurzen Gittervektoren sind leicht in der Überzahl.

10.3 Faktorisieren ganzer Zahlen mit Hilfe Diophantischer Approximationen

Sei $N \in \mathbb{Z}$ das Produkt mindestens zweier verschiedener Primzahlen und $p_1, p_2, \ldots$ die Folge der Primzahlen. C.P. Schnorr [Schnorr91b, Schnorr93] hat das Problem der Faktorisierung von N auf das Finden von $t+2$ Lösungen $(e_1, e_2, \ldots, e_t) \in \mathbb{Z}^t$ der Ungleichungen ($c > 1$ fest)

$$\begin{aligned} &\left|\sum_{i=1}^t e_i \ln p_i - \ln N\right| \leq N^{-c} \cdot p_t^{o(1)} \\ &\sum_{i=1}^t |e_i \ln p_i| \leq (2c-1) \cdot \log_2 N + 2 \cdot \ln p_t \end{aligned} \tag{10.14}$$

reduziert. Wir möchten in diesem Kapitel die Idee und den Algorithmus vorstellen. Für eine ausführliche Betrachtung verweisen wir auf die Originalarbeit.

Wir assoziieren mit den Primzahlen $p_1, p_2, \ldots, p_t$ ein Gitter im $\mathbb{R}^{t+1}$ und mit N einen Punkt $\vec{N}$ im $\mathbb{R}^{t+1}$. Das Problem der Diophantischen Approximation (10.14) reduzieren wir auf die Bestimmung eines in der ℓ_1-Norm hinreichend nahen Gitterpunktes.

Sei $p_1, p_2, \ldots, p_t$ die Folge der ersten Primzahlen, also $2, 3, 5, \ldots$. Wir definieren zu N und festem $c > 1$ das Gitter $L_{\alpha,c} \subseteq \mathbb{R}^{t+1}$, welches von den Zeilenvektoren $b_1, b_2, \ldots, b_t, \vec{N}$ aufgespannt wird:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_t \\ \vec{N} \end{bmatrix} := \begin{bmatrix} \ln p_1 & 0 & \cdots & 0 & N^c \cdot \ln p_1 \\ 0 & \ln p_2 & & 0 & N^c \cdot \ln p_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & & \ln p_n & N^c \cdot \ln p_t \\ 0 & 0 & \cdots & 0 & N^c \cdot \ln N \end{bmatrix} \tag{10.15}$$

Die reellen Vektoren kann man durch rationale Vektoren approximieren. In der Praxis genügen sogar ganzzahlige Vektoren.

Betrachten wir Algorithmus 10.3.1. Da N das Produkt mindestens zweier verschiedener Primzahlen ist, hat x^2 mindestens 4 Wurzeln modulo N, wobei sich zwei nur im Vorzeichen unterscheiden.

$$x^2 - y^2 = (x-y)(x+y) = 0 \pmod{N}$$

Im Fall $x \neq \pm y \pmod{N}$ sind $0 < x-y < N$ und $0 < x+y < N$, also $1 < x \pm y < N$, wobei $x \pm y$ modulo N reduziert sei. Dann liefert $\mathrm{ggT}(x \pm y, N)$ nicht-triviale Teiler von N. Falls x und y sich wie unabhängige Zufallsvariable verhalten, hat Schritt 7 Erfolgswahrscheinlichkeit mindestens $\frac{1}{2}$.

Der Schritt 4 erfordert, daß $|u_i - v_i N|$ über der Basis P_t faktorisiert werden kann. Satz 10.3.2 zeigt, daß für Punkte $z \in L_{\alpha,c}$, die „nahe“ bei $\vec{N}$ liegen, diese Voraussetzung mit hoher Wahrscheinlichkeit erfüllt ist.

Satz 10.3.1
Sei $c > 1$, $\beta, \delta \geq 0$ fest und $p_t < N$. Falls $(e_1, e_2, \ldots, e_t) \in \mathbb{Z}^t$ das Ungleichungssystem

$$\left| \sum_{i=1}^{t} a_i \ln p_i - \ln N \right| \leq N^{-c} \cdot p_t^{o(1)}$$
$$\sum_{i=1}^{t} |a_i \ln p_i| \leq (2c-1) \cdot \log_2 N + 2 \cdot \ln p_t$$

löst, gilt für das in Schritt 3 konstruierte Paar (u_i, v_i) mit

$$u_i := \prod_{a_{i,j}>0} p_j^{a_{ij}} \qquad \text{und} \qquad v_i := \prod_{a_{i,j}<0} p_j^{|a_{i,j}|} \tag{10.16}$$

daß:

$$|u_i - v_i N| \leq p_t^{\beta+\delta+o(1)}$$

Beweis. Siehe [Schnorr93, Theorem 1]. ■

Falls $(e_1, e_2, \ldots, e_t) \in \mathbb{Z}^t$ die Ungleichungen (10.16) erfüllt mit $\beta + \delta \leq 1$ und N hinreichend groß ist, erfüllt das Paar (u_i, v_i) mit hoher Wahrscheinlichkeit die zur Faktorisierung in Schritt 4 nötigen Eigenschaften:

$$u_i = \prod_{a_{i,j}>0} p_j^{a_{ij}} \qquad \text{und} \qquad |u_i - v_i N| \leq p_t$$

Dies ist nur mit vernachlässigbarer Wahrscheinlichkeit nicht der Fall [Schnorr93, Theorem 4].

Satz 10.3.2
Sei $\alpha > 1$, $c > 1$, $\delta \geq 0$ und $p_t = (\ln N)^\alpha < N$. Falls $z \in L_{\alpha,c}$ die Ungleichung

$$\|z - N\|_1 \leq (2c-1) \cdot \ln N + 2\delta \cdot \ln p_t$$

erfüllt, gilt für die in Schritt 3 konstruierten Paare (u_i, v_i):

$$|u_i - v_i N| \leq p_t^{\frac{1}{\alpha}+\beta+o(1)}$$

Beweis. Siehe [Schnorr93, Theorem 2]. ■

Aus den beiden Sätzen 10.3.1 und 10.3.2 folgt, daß es zum Faktorisieren von N genügt, hinreichend viele Gittervektoren $z \in L_{\alpha,c}$ zu finden, die in der ℓ_1-Norm nahe bei $\vec{N}$ liegen. Diese Gitterpunkte findet man in der Praxis durch Anwenden mächtiger Reduktionsalgorithmen auf die Vektoren $b_1, b_2, \ldots, b_n, \vec{N}$ in der ℓ_2-Norm. Die reduzierte Basis enthält im allgemeinen dann Vektoren, die bezüglich der ℓ_1-Norm kurz sind. Man kann erreichen, daß diese Vektoren die Form

$$\sum_{i=1}^{t} e_i b_i - \vec{N}$$

haben ($e_1, e_2, \ldots, e_n \in \mathbb{Z}$) und wählen $z := \sum_{i=1}^{t} e_i b_i$ für Satz 10.3.2.

Algorithmus 10.3.1 Faktorisieren einer ganzen Zahl

EINGABE: ▷ $N \in \mathbb{Z}$ (Produkt mindestens zweier verschiedener Primzahlen)
▷ α und $c \in \mathbb{Q}$ mit $c \geq 1$

1. Bilde Liste $p_1, p_2, \ldots, p_t$ der ersten Primzahlen, $p_t = (\ln N)^\alpha$. Sei $P_t := \{p_1, p_2, \ldots, p_t\}$.

2. Reduziere mit „geeignetem“ Reduktionsalgorithmus das Gitter $L_{\alpha,c} \subseteq \mathbb{R}^{t+1}$ (siehe Basismatrix (10.15)).

3. Bilde für Vektoren $z^{(i)} := \sum_{j=1}^{t} a_{i,j} b_j \in L$ mit kleiner l_1-Norm $\left\| z - \vec{N} \right\|_1$ mit Hilfe von P_t und den ganzzahligen Koeffizienten $(a_{i,1}, a_{i,2}, \ldots, a_{i,t})$ $m \geq t+2$ Tupel $(u_i, v_i) \in \mathbb{N}^2$:

$$u_i := \prod_{a_{i,j}>0} p_j^{a_{ij}}$$
$$v_i := \prod_{a_{i,j}<0} p_j^{|a_{i,j}|}$$

4. FOR $i = 1, 2, \ldots, m$ DO

Faktorisiere $u_i - v_i N$ über P_t und $p_0 := -1$:

$$u_i - v_i N := \prod_{j=0}^{t} p_j^{b_{i,j}}$$

Setze $a_{i,0} := 0$. Bezeichne:

$$a_i := (a_{i,0}, a_{i,1}, \ldots, a_{i,t})$$
$$b_i := (b_{i,0}, b_{i,1}, \ldots, b_{i,t})$$

END for

5. Finde eine $\{0,1\}$-Lösung $(c_1, c_2, \ldots, c_m) \neq 0$ zu:

$$\sum_{i=1}^{m} c_i (a_i + b_i) = 0 \pmod 2 \tag{10.13}$$

6. Bilde (Division durch 2 wegen (10.13) möglich):

$$x := \prod_{j=0}^{t} p_j^{\sum_{i=1}^{m} c_i (a_{i,j} + b_{i,j})/2} \pmod N$$
$$y := \prod_{j=0}^{t} p_j^{\sum_{i=1}^{m} c_i b_{i,j}} \pmod N$$

Es gilt $x^2 = y^2 \pmod N$.

7. IF $x = \pm y \pmod N$ THEN GOTO 5 und berechne neue Lösung

AUSGABE: Zwei nicht-triviale Faktoren $\mathrm{ggT}(x+y, N)$ und $\mathrm{ggT}(x-y, N)$

Kapitel 11

$\mathbb{Z}$-Modul und Hermite-Normalform

In diesem Kapitel beschäftigen wir uns mit der Hermite-Normalform (HNF), die wir bereits kurz in Kapitel 2.2.2 auf Seite 17 vorgestellt haben. Wir entwickeln einen Algorithmus zur Berechnung der Hermite-Normalform und stellen eine Modifikation des Verfahrens vor, welche die Beträge der Koeffizienten während der Berechnung beschränkt. Wir definieren die Smith-Normalform und untersuchen, wann ein ganzzahliges Gitter durch eine lineare Kongruenz beschrieben werden kann.

11.1 $\mathbb{Z}$-Modul

Wir definieren:

Definition 11.1.1 (R-Modul)
Sei R ein kommutativer Ring mit 1, der auf der additiven Gruppe $(M, +)$ wie folgt operiere:

$$R \times M \to M \qquad (a, u) \mapsto a \cdot u$$

Mit den folgenden drei Eigenschaften ($a, b \in R$, $u, v \in M$):

- $a \cdot (u + v) = (a \cdot u) + (a \cdot v)$
- $(a + b) \cdot u = (a \cdot u) + (a \cdot v)$
- $(a \cdot b) \cdot u = a \cdot (b \cdot u)$

Dann ist M ein R-Modul.

Ein Gitter $L \subseteq \mathbb{Z}^n$ ist zum Beispiel ein $\mathbb{Z}$-Modul. $(L, +)$ bildet eine (abelsche) Gruppe und die drei Eigenschaften des $\mathbb{Z}$-Moduls mit $a, b \in \mathbb{Z}$ und $u, v \in L$ gelten offenbar.

Bemerkung 11.1.2
Die $\mathbb{Z}$-Moduln sind die abelschen Gruppen $(M, +)$ vermöge $(z, m) \mapsto \underbrace{m + m + \cdots + m}_{z\text{-}mal}$.

Definition 11.1.3 (Endlich erzeugter R-Modul)
Der R-Modul M ist endlich erzeugt mit Erzeugenden $g_1, g_2, \ldots, g_n$, wenn:

$$M = \left\{ \sum_{i=1}^{n} t_i g_i \;\middle|\; t_1, t_2, \ldots, t_n \in R \right\}$$

Definition 11.1.4 (Frei vom Rang n)
Der R-Modul M heißt frei vom Rang n, falls:

$$M \cong R^n := \underbrace{R \times R \times \cdots \times R}_{n\text{-mal}}$$

Gitter sind endlich erzeugte, freie $\mathbb{Z}$-Moduln zusammen mit einer Einbettung in den $\mathbb{R}^n$.

Satz 11.1.5
Sei V endlich erzeugter $\mathbb{Z}$-Modul und

$$V_{tors} := \{ v \in V \mid \exists m \in \mathbb{Z} \setminus \{0\} : \; m \cdot v = 0 \}$$

Dann gilt:

a) V_{tors} ist endliche Gruppe und es gilt $V \cong V_{tors} \times \mathbb{Z}^n$ mit $n \in \mathbb{N}$ (n heißt der Rang von V).

b) Ist V freier $\mathbb{Z}$-Modul, dann ist jeder Untermodul U von V frei mit $\mathrm{Rang}(U) \leq \mathrm{Rang}(V)$.

c) Ist V endlicher $\mathbb{Z}$-Modul, dann gibt es ein $n \in \mathbb{N}$ und einen freien Untermodul L von $\mathbb{Z}^n$ mit $V \cong \mathbb{Z}^n / L$ (Ein freier $\mathbb{Z}$-Modul von $\mathbb{Z}^n$ ist ein Gitter).

Beweis. Siehe u.a. [Lang93]. ■

Die ersten beiden Aussagen des Satzes zeigen, daß die Untersuchung von endlich erzeugten $\mathbb{Z}$-Moduln in die Betrachtung von endlichen $\mathbb{Z}$-Moduln (also endliche, abelsche Gruppen) und freien $\mathbb{Z}$-Moduln endlichen Ranges (also Gitter) zerfällt.

11.2 Hermite-Normalform

Wir hatten in Kapitel 2.2.2 die Hermite-Normalform in Definition 2.2.6 auf Seite 17 definiert und bereits elementare Eigenschaften nachgewiesen. Eine Matrix $[a_{ij}]_{ij} \in M_{m,n}(\mathbb{R})$ mit $n \leq m$ ist in Hermite-Normalform (kurz HNF), wenn:

a) $a_{ij} = 0$ für $j > i$, d.h. A ist eine untere Dreiecksmatrix

b) $a_{ii} > 0$ für $i = 1, 2, \ldots, m$

c) $0 \leq a_{ij} < a_{ii}$ für $j < i$

Zu jeder Matrix $A \in M_{m,n}(\mathbb{Q})$ mit $\mathrm{Rang}(A) = m \leq n$ gibt es eine Matrix $T \in \mathrm{GL}_n(\mathbb{Z})$, so daß AT in Hermite-Normalform ist, und die Hermite-Normalform AT ist eindeutig bestimmt (Satz 2.2.7 auf Seite 17).

Algorithmus 11.2.1 Hermite-Normalform

EINGABE: $A = [A_1, A_2, \ldots, A_n] = [a_{ij}] \in M_{m,n}(\mathbb{Z})$ mit $\mathrm{Rang}(A) = m < n$

/* $A_1, A_2, \ldots, A_n$ sind die Spaltenvektoren */

1. FOR $i = 1, 2, \ldots, m$ DO

/* Bringe i-te Zeile auf HNF-Form */

1.1. Berechne mit erweitertem Euklidischem Algorithmus $u_i, u_{i+1}, \ldots, u_n \in \mathbb{Z}$, $d > 0$ mit:

$$d = \sum_{j=i}^{n} a_{ij} \cdot u_i = \mathrm{ggT}(a_{ii}, a_{i,i+1}, \ldots, a_{in})$$

1.2. $A_i := \sum_{j=i}^{n} A_j \cdot u_i$ /* $a_{ii} = d$ */

1.3. FOR $j = i+1, i+2, \ldots, n$ DO $A_j := A_j - \frac{a_{ij}}{d} \cdot a_i$ /* $a_{ij} = 0$ */

1.4. FOR $j = 1, 2, \ldots, i-1$ DO $A_j := A_j - \lfloor \frac{a_{ij}}{d} \rfloor \cdot a_i$ /* $0 \le a_{ij} < a_{ii}$ */

END for i

AUSGABE: Hermite-Normalform A der Eingabematrix

11.2.1 Berechnung der Hermite-Normalform

Algorithmus 11.2.1 berechnet in Verbindung mit Algorithmus 11.2.2 zu einer Matrix $A \in M_{m,n}(\mathbb{Z})$ mit $\mathrm{Rang}(A) = m < n$ und Spaltenvektoren $A_1, A_2, \ldots, A_n$ die Hermite-Normalform $\mathrm{HNF}(A)$ mit $\mathrm{HNF}(A) \cdot T = A$. Der Algorithmus transformiert A durch Multiplikation von links mit einem Produkt aus unimodularen Matrizen, die jeweils folgende *Spalten-Elementaroperationen* realisieren:

- Vertauschen zweier Spalten
- Multiplikation einer Spalte mit -1
- Addition eines ganzzahligen Vielfachen einer Spalte zu einer anderen

Die Ausgabematrix des Algorithmus' ist offenbar in Hermite-Normalform, und aus der Schleifeninvariante

$$A^{\mathrm{neu}} = A^{\mathrm{alt}} \cdot T$$

folgt die Korrektheit. Nachteil des Algorithmus' 11.2.1: Die Koeffizienten a_{ij} können extrem anwachsen. Wir werden diese Schwachstelle in Kapitel 11.3 lösen.

11.2.2 Anwendungen

Wir betrachten zwei Beispiele zur Anwendung der Hermite-Normalform. In der ersten Anwendung möchten wir den $\mathbb{Z}$-Kern der Matrix $A \in M_{m,n}(\mathbb{Z})$ mit $\mathrm{Rang}(A) = m < n$ (ein Gitter) bestimmen:

$$\mathbb{Z}\text{-}\,\mathrm{Kern}(A) = \{x \in \mathbb{Z}^n \mid Ax = 0\}$$

Algorithmus 11.2.2 Schritte 1.1, 1.2 und 1.3 des Algorithmus' 11.2.1

/∗ Sei beim ersten Aufruf $T = \mathrm{Id}_n \in \mathrm{GL}_n(\mathbb{Z})$ ∗/

1. FOR $j = n, n-1, \ldots, i+1$ DO

1.1. WHILE $a_{ij} \neq 0$ DO

1.1.1. $q := \left\lfloor \frac{a_{i,j-1}}{a_{ij}} \right\rfloor$

1.1.2. $[A_{j-1}, A_j] := [A_{j-1}, A_j] \cdot \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix}$

1.1.3. $T := \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & 0 & \\ & & 1 & & & & \\ & & & \begin{bmatrix} q & 1 \\ 1 & 0 \end{bmatrix} & & & \\ & & & & 1 & & \\ & 0 & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \cdot T$

/∗ Beachte: $\begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix}^{-1} = \begin{bmatrix} q & 1 \\ 1 & 0 \end{bmatrix}$ ∗/

END while

END for

2. Setze $a_{ii} = |a_{ii}|$

AUSGABE: $d = a_{ii}$, $u_j = t_{i,j}$ für $j = i, i+1, \ldots, n$

Satz 11.2.1
Sei $A \in M_{m,n}(\mathbb{Z})$ mit $\mathrm{Rang}(A) = m < n$ und $B = AT$ die Hermite-Normalform zu A mit $T = [T_1, T_2, \ldots, T_n] \in \mathrm{GL}_n(\mathbb{Z})$. Dann gilt:

$$\mathbb{Z}\text{-}\mathrm{Kern}(A) = \left\{ \sum_{i=m+1}^{n} s_i T_i \;\middle|\; s_{m+1}, s_{m+2}, \ldots, s_n \in \mathbb{Z} \right\}$$

Beweis. Seien $e_1, e_2, \ldots, e_n$ die kanonischen Einheitsvektoren. Für $B = AT$ gilt:

$$\mathbb{Z}\text{-}\mathrm{Kern}(B) = \left\{ \sum_{i=m+1}^{n} s_i e_i \;\middle|\; s_{m+1}, s_{m+2}, \ldots, s_n \in \mathbb{Z} \right\} \tag{11.1}$$

Da die Matrix B eine HNF ist, hat sie folgenden Aufbau:

$$\begin{bmatrix} * & 0 & \cdots & 0 & 0 & \cdots & 0 \\ * & * & \ddots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & 0 & 0 & \cdots & 0 \\ * & * & \cdots & * & 0 & \cdots & 0 \end{bmatrix}$$

Die ersten m Spalten haben eine quadratische, untere Dreiecksform, die letzten $n - m$ Spalten

bestehen nur aus 0. Wegen $Bx = ATx$ für alle $x \in \mathbb{Z}^n$, gilt:

$$\begin{aligned}
\mathbb{Z}\text{-Kern}(A) &= T \cdot \mathbb{Z}\text{-Kern}(B) \\
&= \left\{ T \cdot \sum_{i=m+1}^{n} s_i e_i \;\middle|\; s_{m+1}, s_{m+2}, \ldots, s_n \in \mathbb{Z} \right\} \qquad \text{(wegen (11.1))} \\
&= \left\{ \sum_{i=m+1}^{n} s_i T \cdot e_i \;\middle|\; s_{m+1}, s_{m+2}, \ldots, s_n \in \mathbb{Z} \right\} \\
&= \left\{ \sum_{i=m+1}^{n} s_i T_i \;\middle|\; s_{m+1}, s_{m+2}, \ldots, s_n \in \mathbb{Z} \right\}
\end{aligned}$$

■

Eine andere Anwendung der Hermite-Normalform ist, zu gegebenen Basismatrizen $B, \overline{B} \in M_{m,n}(\mathbb{Z})$ zu entscheiden, ob $L(B) = L\left(\overline{B}\right)$. Wir berechnen mit Algorithmus 11.2.1 Matrizen $T, \overline{T} \in \mathrm{GL}_n(\mathbb{Z})$, so daß $B \cdot T$ und $\overline{B} \cdot \overline{T}$. Wegen der Eindeutigkeit der Hermite-Normalform, und da durch Multiplikation von rechts mit unimodularer Matrix das von den Spaltenvektoren erzeugte Gitter dasselbe ist, gilt:

$$L(B) = L\left(\overline{B}\right) \quad \Longleftrightarrow \quad B \cdot T = \overline{B} \cdot \overline{T}$$

11.2.3 Darstellung rationaler Gitter

Sei $L(B) \subseteq \mathbb{Q}^m$ mit Basismatrix $B \in M_{m,n}(\mathbb{Q})$. Es gibt eine eindeutig bestimmte Hermite-Normalform zu B. $L(B)$ ist ein Gitter. Sei $d := \min\{k \in \mathbb{N} \mid k \cdot L(B) \subseteq \mathbb{Z}^n\}$. Wir können das rationale Gitter $L(B)$ durch das Paar (W, d) ($W \in M_{m,n}(\mathbb{Z})$ in Hermite-Normalform) darstellen mit

$$W = \mathrm{HNF}(d \cdot B) = d \cdot \mathrm{HNF}(B)$$

Wir möchten zu gegebenem (W, d) und $x \in \mathbb{Q}^n$ entscheiden, ob $x \in \frac{1}{d} \cdot L(W)$ ist. Sei $W = [W_1, W_2, \ldots, W_m, 0, \ldots, 0]$. Löse das lineare Gleichungssystem

$$\sum_{i=1}^{n} W_i y_i = d \cdot x$$

in $y_1, y_2, \ldots, y_m \in \mathbb{Q}$. Dann gilt:

$$x \in \tfrac{1}{d} \cdot L(W) \quad \Longleftrightarrow \quad y_1, y_2, \ldots, y_m \in \mathbb{Z}$$

11.3 Modulare Berechnung der Hermite-Normalform

Nachteilig an Algorithmus 11.2.1 zur Berechnung der Hermite-Normalform ist, daß Koeffizienten im Laufe des Verfahrens extrem anwachsen können. J. Hafner [HaMcC91] gibt das Beispiel einer 20×20-Matrix mit ganzzahligen Einträgen aus $\{0, 1, \ldots, 10\}$. Es treten während der Berechnung Werte der Größe 10^{5011} auf, während die Determinante kleiner als 10^{33} ist.

In diesem Abschnitt entwickeln wir eine Modifikation des Algorithmus', der dieses Problem durch modulare Arithmetik löst. Für die Berechnung der Gitterdeterminanten wählen wir das Standard-Skalarprodukt. Das Verfahren kann sogar noch verbessert werden (siehe [HaMcC91]).

Lemma 11.3.1
Sei $A = [a_{ij}] \in M_{m,n}(\mathbb{Z})$ mit $\operatorname{Rang}(A) = m < n$. Dann gilt für $D := \det L(A)$:

$$D \cdot \mathbb{Z}^m \subseteq L(A)$$

Beweis. Sei $B = [B_1, B_2, \ldots, B_n]$ die Hermite-Normalform zu A, also eine untere Dreicksmatrix mit 0-Spalten rechts und $L(A) = L(B)$.

$$B = \begin{bmatrix} b_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ b_{21} & b_{22} & \ddots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & 0 & 0 & \cdots & 0 \\ b_{m1} & b_{m2} & \cdots & b_{mm} & 0 & \cdots & 0 \end{bmatrix}$$

Es ist:

$$D = \prod_{j=1}^{m} b_{jj}$$

Setze $D_i := \prod_{j=i}^{m} b_{jj}$ für $i = 1, 2, \ldots, m$. Seien $e_1, e_2, \ldots, e_m \in \mathbb{Z}^m$ die kanonischen Einheitsvektoren. Durch Induktion über $i = m, m-1, \ldots, 1$ zeigen wir, daß $D_i \cdot e_i \in L(B)$ gilt:

- Verankerung für $i = m$. Es gilt:

$$D_m \cdot e_m = b_{mm} \cdot e_m = B_m \in L(B)$$

- Induktionsschluß von $i+1$ auf i: Da $D_{i+1} \in \mathbb{Z}$ und $b_{ji} \in \mathbb{Z}$ gilt:

$$\underbrace{D_{i+1} \cdot B_i}_{\in L(B)} = D_{i+1} \cdot \left(b_{ii} \cdot e_i + \sum_{j=i+1}^{n} b_{ji} \cdot e_j \right) = D_i \cdot e_i + \underbrace{\sum_{j=i+1}^{n} \overbrace{b_{ij} \left(\prod_{k=i}^{j+1} b_{kk} \right)}^{\in \mathbb{Z}} D_j \cdot e_j}_{\in L(B) \text{ nach Induktionsannahme}}$$

  Da $L(B)$ eine additive Gruppe ist, erhalten wir $D_i \cdot e_i \in L(B)$.

Weil $D_j \mid D$ und $L(A) = L(B)$, folgt die Behauptung $D \cdot \mathbb{Z}^n \subseteq L(A)$. ■

Speziell haben wir im Beweis zu Lemma 11.3.1 gezeigt:

Korollar 11.3.2
Sei $B = [b_{ij}] \in M_{m,n}(\mathbb{Z})$ eine untere Dreiecksmatrix mit $\operatorname{Rang}(B) = m < n$. Definiere für $i = 1, 2, \ldots, m$:

$$D_i := \prod_{j=i}^{m} b_{jj} \tag{11.2}$$

Dann gilt $D_i \cdot e_i \in L(B)$ für $i = 1, 2, \ldots, m$.

Aus Lemma 11.3.1 folgt, daß eine Matrix $A = [a_{ij}] \in M_{m,n}(\mathbb{Z})$ und $A' \in M_{m,n+m}(\mathbb{Z})$ mit

$$A' := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & D_1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & a_{2n} & 0 & D_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \ddots & 0 \\ a_{m1} & a_{m2} & \cdots & a_{mn} & 0 & & \cdots & D_m \end{bmatrix}$$

bis auf 0-Spalten rechts die gleiche Hermite-Normalform haben. Wir verfolgen den Ansatz, da wir anstatt der zusätzlichen Spalten die Einträge der i-ten Zeile bei der Berechnung modulo D_i reduzieren. Jedoch kann auf der Diagonalen der Hermite-Normalform D_i stehen. In diesem Fall dürfen wir nicht modulo D_i reduzieren. Links von der Diagonalen kann bei der Hermite-Normalform kein Wert größer oder gleich D_i stehen, da sonst das Diagonalelemente in der Zeile größer als D_i und damit die Determinante größer als $\det L(A)$ wäre.

Lemma 11.3.3
Sei die Matrix $A = [a_{ij}] \in M_{m,n}(\mathbb{Z})$ mit $\operatorname{Rang}(A) = m < n$ in Hermite-Normalform und $B = [b_{ij}] \in M_{m,n}(\mathbb{Z})$ eine untere Dreiecksmatrix, deren Spaltenvektoren im Gitter $L(A)$ liegen. Falls $a_{ii} = b_{ii}$ für $i = 1, 2, \ldots, m$ ist, gilt $L(B) = L(A)$.

Beweis. Durch Multiplikation mit unimodularen Matrizen erreichen wir, daß die Einträge links der Diagonalen größer oder gleich 0 und kleiner als das Diagonalelement sind. Sei B' die erhaltene Matrix. Offenbar ist B' die Hermite-Normalform zu B, und es gilt $L(B') = L(B)$. Da die Hermite-Normalform nach Satz 2.2.7 (Seite 17) eindeutig bestimmt ist, folgt $B' = A$. Wir erhalten die Behauptung $L(B) = L(B') = L(A)$. ■

Zu einer Matrix $A \in M_{m,n}(\mathbb{Z})$ bezeichnen wir mit $g_i(A)$ für $i = 1, 2, \ldots, m$:

$$g_i(A) := \operatorname{ggT}\left(\begin{array}{l} \text{Determinanten aller } i \times i\text{-Teilmatrizen} \\ \text{der ersten } i \text{ Zeilen der Matrix } A \end{array} \right)$$

Wir bezeichnen mit $A_{i,S}$ die $i \times i$-Matrix, deren Spalten jeweils die ersten i Einträge der Spalten $S \subseteq \{1, 2, \ldots, n\}$ (S sei eine geordnete Menge mit $|S| = i$) von A sind. Es gilt für $i = 1, 2, \ldots, m$:

$$g_i(A) = \operatorname{ggT}(\{\det A_{i,S} \mid S \subseteq \{1, 2, \ldots, n\},\ |S| = i\})$$

Diese Größen bleiben beim HNF-Algorithmus 11.2.1 erhalten:

Lemma 11.3.4
Sei $A \in M_{m,n}(\mathbb{Z})$. Dann bleiben $g_1(A), g_2(A), \ldots, g_m(A)$ bei Multiplikation von rechts mit unimodularen Matrizen erhalten.

Beweis. Die Multiplikation einer Spalte mit -1 ändert den Betrag der Determinanten nicht und der größte gemeinsame Teiler $g_i(A)$ bleibt erhalten. Auch das Vertauschen zweier Spalten ändert den größten gemeinsamen Teiler $g_i(A)$ nicht.

Betrachten wir den Fall, wenn ein ganzzahliges Vielfaches $z \in \mathbb{Z}$ einer Spalte j zu einer anderen Spalte k addieren wird. Die Determinante $\det A_{i,S}$ bleibt erhalten, wenn $j, k \in S$ sind. Im anderen Fall, also $j \in S$ und $k \notin S$, gilt wegen der Linearität der Determinante für $S' := (S \cup \{k\}) \setminus \{j\}$:

$$\det A_{i,S}^{\text{neu}} = \det A_{i,S} + z \cdot \det A_{i,S'}$$

Da der größte gemeinsame Teiler $g_i(A)$ auch über $\det A_{i,S'}$ gebildet wird, ändert sich $g_i(A)$ nicht, denn wenn wir zu einem Argument ein ganzzahliges Vielfaches eines anderen Arguments addieren, bleibt der größte gemeinsame Teiler erhalten:

$$\operatorname{ggT}(\ldots, \det A_{i,S}, A_{i,S'}, \ldots) = \operatorname{ggT}(\ldots, \det A_{i,S} + z \cdot \det A_{i,S'}, \det A_{i,S'}, \ldots)$$

■

Es folgt:

Satz 11.3.5
Sei $A \in M_{m,n}(\mathbb{Z})$ mit $m < n$ und $B := \mathrm{HNF}(A)$ die Hermite-Normalform zu A. Dann gilt:

$$\mathrm{ggT}(D_i, g_i(A)) = \mathrm{ggT}(D_i, g_i(B)) \qquad \text{für } i = 1, 2, \ldots, m.$$

Beweis. Nach Lemma 11.3.4 ist $g_i(A) = g_i(B)$ für $i = 1, 2, \ldots, m$. Die Reduktion modulo D_i bedeutet, daß ein ganzzahliges Vielfaches von D_i subtrahiert wird. Wir erhalten mit $z_1, \ldots, z_k \in \mathbb{Z}$:

$$\begin{aligned}
\mathrm{ggT}(D_i, g_i(A)) &= \mathrm{ggT}(D_i, \det A_{i,S_1} \bmod D_i, \det A_{i,S_2} \bmod D_i, \ldots, \det A_{i,S_k} \bmod D_i) \\
&= \mathrm{ggT}(D_i, \det A_{i,S_1} + z_1 D_i, \det A_{i,S_2} + z_2 D_i, \ldots, \det A_{i,S_k} + z_k D_i) \\
&= \mathrm{ggT}(D_i, \det A_{i,S_1}, \det A_{i,S_2}, \ldots, \det A_{i,S_k}) \\
&= \mathrm{ggT}(D, g_i(B))
\end{aligned}$$

■

Korollar 11.3.6
Sei $A = [a_{ij}] \in M_{m,n}(\mathbb{Z})$ eine untere Dreiecksmatrix mit $\mathrm{Rang}(A) = m < n$ und die Matrix $B = [b_{ij}] := \mathrm{HNF}(A)$ die Hermite-Normalform zu A. Zu $D_1, D_2, \ldots, D_m$ (wie in (11.2) definiert) gilt:

$$b_{ii} = \mathrm{ggT}(D_i, a_{ii}) \qquad \text{für } i = 1, 2, \ldots, m. \tag{11.3}$$

Beweis. Weil A und B untere Dreiecksmatrizen sind, gilt für $k = 1, 2, \ldots, m$:

$$g_k(A) = \prod_{j=1}^{k} a_{jj} \qquad\qquad g_k(B) = \prod_{j=1}^{k} b_{jj}$$

Insbesondere ist $D_1 = g_m(B) = \det L(A)$, da $L(A) = L(B)$. Nach Satz 11.3.5 erhalten wir:

$$\mathrm{ggT}(D_1, g_k(A)) = \mathrm{ggT}(D_1, g_k(B)) \qquad \text{für } k = 1, 2, \ldots, m.$$

Da $g_k(A)$ und $g_k(B)$ Teiler von D_1 sind, folgt:

$$\mathrm{ggT}\left(D_1, \prod_{j=1}^{k} a_{jj}\right) = \prod_{j=1}^{k} b_{jj} \qquad \text{für } k = 1, 2, \ldots, m. \tag{11.4}$$

Wir haben die Behauptung (11.3) für $i = 1$ mit $k = 1$ bewiesen. Sei im weiteren $i > 1$. Wegen Definition (11.2) $D_i := \prod_{j=i}^{m} b_{jj}$ gilt:

$$\frac{D_1}{\prod_{j=1}^{i-1} b_{jj}} = \frac{\prod_{j=1}^{m} b_{jj}}{\prod_{j=1}^{i-1} b_{jj}} = \prod_{j=i}^{m} b_{jj} = D_i \tag{11.5}$$

Aus (11.4) folgt mit $k := i - 1$:

$$1 = \frac{\mathrm{ggT}\left(D_1, \prod_{j=1}^{i-1} a_{jj}\right)}{\prod_{j=1}^{i-1} b_{jj}} = \mathrm{ggT}\left(D_i, \prod_{j=1}^{i-1} \frac{a_{jj}}{b_{jj}}\right) \tag{11.6}$$

Wir erhalten aus (11.4) mit $k = i$ und (11.5):

$$b_{ii} = \frac{\mathrm{ggT}\left(D_1, \prod_{j=1}^{i} a_{jj}\right)}{\prod_{j=1}^{i-1} b_{jj}} = \mathrm{ggT}\left(D_i, a_{ii} \cdot \prod_{j=1}^{i-1} \frac{a_{jj}}{b_{jj}}\right)$$

Wegen Identität (11.6) ist der zweite Faktor des zweiten Arguments teilerfremd zu D_i und es gilt:

$$b_{ii} = \mathrm{ggT}(D_i, a_{ii})$$

■

Algorithmus 11.3.1 berechnet die Hermite-Normalform, wobei der Zahlenbereich für die (mod D_i)-Rechnung $\{0, 1, \ldots, D_i - 1\}$ ist. Da nur auf der Diagonalen der Wert D_i stehen kann (vergleiche HNF-Definition), testen wir in Schritt 2.3, ob $a_{ii} \equiv 0 \pmod{D_i}$ gilt. Wegen $\text{Rang}(A) = m < n$ ist $a_{ii} \neq 0$, und wir setzen a_{ii} auf D_i.

Algorithmus 11.3.1 Hermite-Normalform modulo Determinante D

EINGABE: ▷ $A = [A_1, A_2, \ldots, A_n] = [a_{ij}] \in M_{m,n}(\mathbb{Z})$ mit $\text{Rang}(A) = m < n$
▷ Determinante D von $L(A)$

/* $A_1, A_2, \ldots, A_n$ sind die Spaltenvektoren */

1. $D_1 := D$

2. FOR $i = 1, 2, \ldots, m$ DO

/* Bringe i-te Zeile auf HNF-Form */

2.1. Berechne mit erweitertem Euklidischem Algorithmus $u_i, u_{i+1}, \ldots, u_n, u \in \mathbb{Z}$, $d > 0$ mit:

$$d = \sum_{j=i}^{n} a_{ij} \cdot u_i + u \cdot D_i = \text{ggT}(D_i, a_{ii}, a_{i,i+1}, \ldots, a_{in})$$

/* Wir schreiben bei Vektor-Operationen (mod $M_1, M_2, \ldots, M_m$), wenn der j-te Eintrag modulo M_j reduziert wird ($j = 1, 2, \ldots, m$). */

2.2. $A_i := \sum_{j=i}^{n} A_j \cdot u_i \pmod{D_1, D_2, \ldots, D_i, \ldots, D_i}$

2.3. IF $a_{ii} = 0$ THEN $a_{ii} := D_i$ /* Da $a_{ii} \neq 0$ wegen $\text{Rang}(A) = m$ */

2.4. FOR $j = i + 1, i + 2, \ldots, n$ DO /* Erreiche $a_{ij} = 0$ */

$$A_j := A_j - \tfrac{a_{ij}}{d} \cdot a_i \pmod{D_1, D_2, \ldots, D_i, \ldots, D_i}$$

END for j

2.5. FOR $j = 1, 2, \ldots, i - 1$ DO /* Erreiche $0 \leq a_{ij} < a_{ii}$ */

$$A_j := A_j - \lfloor \tfrac{a_{ij}}{d} \rfloor \cdot a_i \pmod{D_1, D_2, \ldots, D_i, \ldots, D_i}$$

END for j

2.6. $D_{i+1} = \frac{D_i}{d}$

END for i

AUSGABE: Hermite-Normalform A der Eingabematrix

Nach Korollar 11.3.6 ist die Berechnung der Diagonalelemente korrekt. Wegen $D_i \cdot e_i \in L$ ($1 \leq i \leq m$) bleibt bei den Modulo-Operationen das Gitter invariant. Wir erhalten eine untere Dreiecksmatrix, deren Diagonalelemente mit denen der Hermite-Normalform übereinstimmen und deren Spaltenvektoren im Gitter liegen. Aus Lemma 11.3.3 folgt die Korrektheit des Algorithmus'. Für $[b_{ij}] := \text{HNF}(A)$ gilt

$$D_i := \prod_{j=i}^{m} b_{jj} \qquad \text{für } i = 1, 2, \ldots, m$$

und insbesondere $D_1 = \det L(A)$. Da wir bei der i-ten Iteration das i-te Diagonalelement der Hermite-Normalform zu A erhalten, kennen wir zu diesem Zeitpunkt $D_1, D_2, \ldots, D_i$. Für die

Zeilen $i+1, i+2, \ldots, n$ reduziert der Algorithmus modulo D_i, einem Vielfachen von D_j für $j > i$. Wir erhalten kein falsches Ergebnis, da wir die Einträge in der j-ten Zeile später modulo D_j reduzieren.

Alternativ kann als Modul auch der Exponent des Gitters, ein Teiler der Gitterdeterminanten, verwendet werden:

Definition 11.3.7 (Exponent eines Gitters)
Sei $L \subseteq \mathbb{R}^n$ Gitter vom Rang n. Der Exponent E von $\mathbb{Z}^n/L$ (bzw. von L) ist:

$$E := \min\{k \in \mathbb{N} \mid k \cdot \mathbb{Z}^n \subseteq L\} = \max\{\operatorname{ord}(x) \mid x \in \mathbb{Z}^n/L\}$$

Als Modul wählt man im allgemeinen die Determinante, da diese effizient zu berechnen ist.

11.4 Approximation von Gittern durch Gitter mit zyklischer Faktorgruppe

In diesem Abschnitt reduzieren wir algorithmische Probleme zu ganzzahligen Gittern auf den Fall eines Gitters $L \subseteq \mathbb{Z}^n$, das durch eine lineare Kongruenz gegeben ist:

$$L = \{x \in \mathbb{Z}^n \mid \langle x, u\rangle \equiv 0 \pmod{d}\}$$

mit $u \in \mathbb{Z}^n \setminus \{0\}$ und $d \in \mathbb{N}$. Wir werden in Satz 11.4.7 und Korollar 11.4.8 zeigen, daß jedes Gitter $L \subseteq \mathbb{Z}^n$ in dieser Form geschrieben werden kann, falls L den Rang n hat und die abelsche Faktorgruppe $\mathbb{Z}^n/L$ zyklisch ist. In Satz 11.4.7 zeigen wir, daß jedes ganzzahlige Gitter effizient approximiert werden kann durch ein rationales Gitter $L \subseteq \frac{1}{k}\mathbb{Z}^n$ mit zyklischer Faktorgruppe $\mathbb{Z}^n/kL$.

Wir definieren die Smith-Normalform, kurz SNF, einer ganzzahligen, quadratischen Matrix. Die Definition geht zurück auf H.J.S. Smith [Smith1861].

Definition 11.4.1 (Smith-Normalform (SNF))
Eine quadratische Matrix $B = [b_{ij}] \in M_{n,n}(\mathbb{Z})$ ist in Smith-Normalform, wenn:

a) $b_{ij} > 0$ *für* $i = j$ *und* $b_{ij} = 0$ *für* $i \neq j$

b) $b_{i+1,i+1} \mid b_{ii}$ *für* $i = 1, 2, \ldots, n-1$

In Worten: Eine quadratische Matrix ist in Smith-Normalform, wenn alle Einträge bis auf die Diagonale gleich 0 sind, die Werte auf der Diagonalen alle positiv sind und jeweils ein Vielfaches des folgenden Eintrags auf der Diagonalen sind. Auch die Smith-Normalform einer Matrix ist eindeutig bestimmt:

Satz 11.4.2
Sei $A \in M_{n,n}(\mathbb{Z})$ mit $\det A \neq 0$. Dann existiert eine eindeutig bestimmte Smith-Normalform $B = \mathrm{SNF}(A)$ von der Form

$$B = SAT \qquad S, T \in \mathrm{GL}_n(\mathbb{Z})$$

Beweis. Beweisidee und einen effizienten Algorithmus zur Bestimmung der Smith-Normalform finden sich in [Cohen93] und [Lang93, Theorem 7.9, Kapitel 3, §8]. Das Produkt $\prod_{k=i}^{n} b_{kk}$ ist der größte gemeinsame Teiler aller $(n-i+1)$-Minore der Matrix A. ■

Beispiel 11.4.3 (Smith-Normalform)
Sei

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

mit $\det A \neq 0$. Setze $b_2 := \mathrm{ggT}(a, b, c, d)$ und $b_1 := \frac{ad-bc}{b_2}$. Dann ist:

$$\mathrm{SNF}(A) = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}$$

Vergleiche Übungsaufgabe 17 des zweiten Kapitels in [Cohen93]. ◇

Satz 11.4.2 für Matrizen ist äquivalent zu folgendem Satz für $\mathbb{Z}$-Moduln.

Satz 11.4.4 (Elementarteilersatz)
Es seien $L \subseteq L'$ freie $\mathbb{Z}$-Moduln mit $\mathrm{Rang}(L) = \mathrm{Rang}(L') = n$. Dann gibt es $d_1, d_2, \ldots, d_n \in \mathbb{N}$ (sog. Elementarteiler von L in L') mit:

a) $d_{i+1} \mid d_i$ für $i = 1, 2, \ldots, n-1$

b) $L'/L \cong \bigoplus_{i=1}^{n} (\mathbb{Z}/d_i\mathbb{Z})$ (eindeutig bestimmt) und $[L' : L] = d_1 d_2 \ldots d_n$.

c) Es gibt $\mathbb{Z}$-Basis $v_1, v_2, \ldots, v_n$ von L', so daß $d_1 v_1, d_2 v_2, \ldots, d_n v_n$ $\mathbb{Z}$-Basis von L ist.

Beweis. Siehe [Lang93, Theorem 7.8, Kapitel 3, §8]. ■

Wir erhalten einen bekannten Satz von L. Kronecker (1877):

Beispiel 11.4.5 (Kanonische Form endlicher, abelscher Gruppen)
Jede endliche, abelsche Gruppe G mit $|G| > 1$ hat die kanonische Form:

$$G \cong \bigoplus_{i=1}^{n} (\mathbb{Z}/d_i\mathbb{Z})$$

mit $d_{i+1} \mid d_i$ und $d_i > 1$ für $i = 1, 2, \ldots, n-1$. ◇

Beispiel 11.4.6 (Kanonische Form von $\mathbb{Z}^n/L$)
Sei $L \subseteq \mathbb{Z}^n$ ein vollständiges Gitter. Dann gilt:

$$\mathbb{Z}^n/L \cong \bigoplus_{i=1}^{n} (\mathbb{Z}/d_i\mathbb{Z})$$

mit $d_{i+1} \mid d_i$ für $i = 1, 2, \ldots, n-1$ eindeutig bestimmt (bis auf $d_i = 1$). Beachte, daß $(\mathbb{Z}/d_i\mathbb{Z}, +)$ eine zyklische Gruppe der Ordnung d_i ist. Es gilt:

$$\begin{aligned} \mathbb{Z}^n/L \text{ zyklisch} \quad &\Longleftrightarrow \quad d_1, d_2, \ldots, d_n \text{ paarweise teilerfremd} \\ &\Longleftrightarrow \quad d_1 > 1 \text{ und } d_i = 1 \text{ für } i = 2, 3, \ldots, n \end{aligned}$$

◇

Satz 11.4.7
Sei $L \subseteq \mathbb{Z}^n$ ein Gitter vom Rang n. $\mathbb{Z}^n/L$ hat genau dann die kanonische Form

$$\bigoplus_{i=1}^{n} (\mathbb{Z}/d_i\mathbb{Z}) \qquad \text{mit} \quad d_{i+1} \mid d_i,$$

wenn L von folgender Form ist (wobei $\langle \cdot, \cdot \rangle$ das Standard-Skalarprodukt ist):

$$L = \{x \in \mathbb{Z}^n \mid \langle x, u_i \rangle \equiv 0 \pmod{d_i} \text{ für } i = 1, 2, \ldots, n\}$$

mit

- $d_{i+1} \mid d_i$ *für* $i = 1, 2, \ldots, n-1$,
- $\det L = d_1 d_2 \ldots d_n$ *und*
- $[u_1, u_2, \ldots, u_n] \in \mathrm{GL}_n(\mathbb{Z})$.

Beweis. Für $U \in \mathrm{GL}_n(\mathbb{Z})$ sind die Einträge der Spaltenvektoren jeweils teilerfremd, denn wir können die Spalte durch den größten gemeinsamen Teiler $d > 0$ dividieren und erhalten eine ganzzahlige Matrix U' mit $\det U = d \cdot \det U'$. Wegen $\det U = 1$ und $\det U' \in \mathbb{Z}$ folgt $d = 1$.

Wir zeigen beide Richtungen der Behauptung:

„⇒“ Sei A die Basismatrix zu L in Smith-Normalform:

$$U \cdot A \cdot V = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \qquad U, V \in \mathrm{GL}_n(\mathbb{Z})$$

Damit ist $\det L = \det A = d_1 d_2 \cdots d_n$. Das folgende Produkt zweier Matrizen ist eine Basismatrix des Gitters L:

$$U^{-1} \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} = AV \tag{11.7}$$

Das duale Gitter L^* ist definiert als:

$$L^* := \{x \in \mathrm{span}(L) \mid \forall a \in L : \langle x, a \rangle \in \mathbb{Z}\}$$

Die Basismatrix des dualen Gitters entsteht durch Invertieren und Transponieren der Basismatrix. Wir erhalten aus (11.7) daher eine Basismatrix des zu L dualen Gitters L^*:

$$\left((AV)^{-1}\right)^{\mathsf{T}} = U^{\mathsf{T}} \cdot \begin{bmatrix} d_1^{-1} & & & \\ & d_2^{-1} & & \\ & & \ddots & \\ & & & d_n^{-1} \end{bmatrix} \tag{11.8}$$

Seien $u_1, u_2, \ldots, u_n$ die Spaltenvektoren von U^{T}. Aus (11.8) folgt, daß die Vektoren $d_i^{-1} \cdot u_i$ $(i = 1, 2, \ldots, n)$ eine Basis von L^* bilden. Wegen $(L^*)^* = L$ (Satz 2.2.29 auf Seite 29) folgt:

$$\begin{aligned} L &= \left\{x \in \mathbb{Z}^n \;\middle|\; \left\langle x, d_i^{-1} \cdot u_i \right\rangle \in \mathbb{Z} \text{ für } i = 1, 2, \ldots, n\right\} \\ &= \{x \in \mathbb{Z}^n \mid \langle x, u_i \rangle \in d_i\mathbb{Z} \text{ für } i = 1, 2, \ldots, n\} \\ &= \{x \in \mathbb{Z}^n \mid \langle x, u_i \rangle \equiv 0 \pmod{d_i} \text{ für } i = 1, 2, \ldots, n\} \end{aligned}$$

„⇐“ Sei das Gitter L von der Form

$$L = \{x \in \mathbb{Z}^n \mid \langle x, u_i\rangle = 0 \pmod{d_i} \quad \text{für } i = 1, 2, \ldots, n\}$$

mit $d_{i+1} \mid d_i$ und die Matrix $(u_{ij}) := [u_1, u_2, \ldots, u_n] \in \mathrm{GL}_n(\mathbb{Z})$. Setze:

$$L_i := \{x \in \mathbb{Z}^n \mid \langle x, u_i\rangle \equiv 0 \pmod{d_i}\} \quad \text{für } i = 1, 2, \ldots, n$$

Da $\mathrm{ggT}(u_{i1}, u_{i2}, \ldots, u_{in}) = 1$, gilt $\det L_i = d_i$. Ferner ist $\mathbb{Z}^n/L_i \cong \mathbb{Z}/d_i\mathbb{Z}$ zyklisch von der Ordnung d_i. Es ist:

$$L = \bigcap_{i=1}^{n} L_i$$

Wegen $\det L = d_1 d_2 \cdots d_n$ (folgt aus Korollar 2.2.21 auf Seite 24) gilt:

$$\left(\bigcap_{i=1}^{j} L_i\right) \Big/ \left(\bigcap_{i=j+1}^{n} L_i\right) \cong \mathbb{Z}^n/L_{j+1}$$

Es folgt aus dem zweiten Isomorphiesatz für Gruppen:

$$\mathbb{Z}^n/L = \mathbb{Z}^n \Big/ \bigcap_{i=1}^{n} L_i \overset{\text{2. Isomorphiesatz}}{\cong} \bigoplus_{i=1}^{n} \mathbb{Z}/L_i = \bigoplus_{i=1}^{n} (\mathbb{Z}/d_i\mathbb{Z})$$

■

Zu einem gegebenen Gitter L können wir effizient durch den Algorithmus zur Bestimmung der Smith-Normalform die Werte $d_1, d_2, \ldots, d_n$ berechnen.

Korollar 11.4.8
Sei $L \subseteq \mathbb{Z}^n$ ein volldimensionales Gitter. Genau dann kann das Gitter dargestellt werden als

$$L = \{x \in \mathbb{Z}^n \mid \langle x, u\rangle \equiv 0 \pmod{d}\}$$

mit $u \in \mathbb{Z}^n \setminus \{0\}$ und $d \in \mathbb{N}$, wenn die Faktorgruppe $\mathbb{Z}^n/L$ zyklisch ist.

Beweis. Da $\mathbb{Z}^n/L$ genau dann zyklisch ist, wenn $d := d_1 > 1$ und $d_2 = d_3 = \cdots = d_n = 1$ (vergleiche Beispiel 11.4.6), folgt die Behauptung aus Satz 11.4.7. ■

Falls das Gitter $L \subseteq \mathbb{Z}^n$ keine zyklische Faktorgruppe $\mathbb{Z}^n/L$ hat, können wir es durch ein rationales Gitter mit zyklischer Faktorgruppe (beliebig) approximieren. Wir definieren:

Definition 11.4.9 (ϵ-Approximation)
Sei $L \subseteq \mathbb{Z}^n$ Gitter vom Rang n und $\epsilon \in]0, 1[$. Sei $f : L \to \mathbb{Z}^n$ eine lineare Abbildung und $k \in \mathbb{Z}$, so daß:

$$\forall x \in L: \qquad \left\|x - \tfrac{1}{k} \cdot f(x)\right\| \leq \epsilon \cdot \|x\|$$

Dann heißt (f, k) ϵ-Approximation zum Gitter L. Das Paar (f, k) heißt zyklische ϵ-Approximation, wenn $\mathbb{Z}^n/f(L)$ zyklisch ist.

Dabei ist $f(L) \subseteq \mathbb{Z}^n$ ein Gitter mit dem gleichen Rang wie L, denn wegen der Linearität von f ist $f(L)$ diskret, und falls der Rang kleiner als der von L wäre, gäbe es einen nicht-trivialen Vektor $x \in L$ mit $f(x) = 0$, so daß die Eigenschaft der ϵ-Approximation für $\epsilon \in]0, 1[$ nicht erfüllt ist.

Satz 11.4.10
Zu jedem Gitter $L \subseteq \mathbb{Z}^n$ vom Rang n und $\epsilon > 0$ gibt es eine zyklische ϵ-Approximation. Diese kann effizient berechnet werden.

Beweis. Sei $B := [b_{ij}] = [b_1, b_2, \ldots, b_n]$ Basismatrix zu L. Die Matrix B sei in (unterer) Hermite-Normalform nach dem absolut kleinsten:

a) $b_{ij} = 0 \quad$ für $i > j$

b) $|b_{ji}| < \frac{1}{2}|b_{jj}| \quad$ für $i < j$

Seien $b_{11}, b_{22}, \ldots, b_{nn}$ paarweise verschieden (für den allgemeinen Fall siehe [PaSchn87]). Setze:

$$k := \left\lfloor \epsilon^{-1} \right\rfloor \cdot \prod_{i \neq j} |b_{ii} - b_{jj}| \tag{11.9}$$

Es ist $k \in \mathbb{Z}$. Betrachte die lineare Abbildung $f : L \to \mathbb{Z}^n$ mit $f(b_i) := k \cdot \bar{b}_i$, wobei:

$$\bar{b}_{ij} = \begin{cases} b_{ij} & \text{falls } i \neq j \\ b_{ij} - \frac{1}{k} & \text{falls } i = j \end{cases} \tag{11.10}$$

Sei $\bar{B} = \left(\bar{b}_{ij}\right)$. Die Diagonalelemente der Matrix $k \cdot \bar{B}$ sind $k \cdot b_{ii} - 1$ für $i = 1, 2, \ldots, n$. Durch Widerspruch zeigen wir, daß diese paarweise teilerfremd sind. Angenommen, es gäbe i, j, $i \neq j$, und eine Primzahl p mit:

$$k \cdot b_{ii} - 1 \equiv k \cdot b_{jj} - 1 \pmod{p}$$

Es folgt $p \mid k(b_{ii} - b_{jj})$. Da nach Definition (11.9) $(b_{ii} - b_{jj}) \mid k$ gilt, ist in jedem Fall $p \mid k$. Es folgt der Widerspruch:

$$p \nmid (kb_{jj} - 1)$$

Da die Zahlen $k \cdot \bar{b}_{ii}$ (siehe (11.10)) nach Voraussetzung paarweise relativ prim sind, ist der größte gemeinsame Teiler aller k-Minore, $k < n$, der Matrix $k \cdot \bar{B}$ gleich 1. Betrachten wir die Smith-Normalform zu $k \cdot \bar{B}$. Nach den vorherigen Überlegungen kann nur das erste Diagonalelement ungleich 1 sein:

$$\mathrm{SNF}\left(k \cdot \bar{B}\right) = \begin{bmatrix} d & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

Die ϵ-Approximation können wir effizient berechnen, da die Laufzeit durch die Bestimmung der Hermite-Normalform dominiert wird. Dazu verwenden wir Algorithmus 11.3.1. ■

Der folgende Satz zeigt den Zusammenhang der sukzessiven Minima des Gitters $L \subseteq \mathbb{Z}^n$ und der ϵ-Approximation (f, k):

Satz 11.4.11
Sei (f, k) eine ϵ-Approximation zum volldimensionalen Gitter $L \subseteq \mathbb{Z}^n$. Dann gilt:

$$\left|\lambda_i(L) - \tfrac{1}{k} \cdot \lambda_i(f(L))\right| \leq \epsilon \cdot \lambda_i(L) \qquad \textit{für } i = 1, 2, \ldots, n.$$

Beweis. Sei $\bar{L} := \frac{1}{k} \cdot f(L)$. Wir wählen linear unabhängige Gittervektoren $b_1, b_2, \ldots, b_n \in L$ mit $\|b_i\| = \lambda_i(L)$ und $\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_n \in \bar{L}$ mit $\left\|\bar{b}_i\right\| = \lambda_i\left(\bar{L}\right)$.

- Wir zeigen $\lambda_i\left(\bar{L}\right) - \lambda_i(L) \leq \epsilon \cdot \lambda_i(L)$ für $i = 1, 2, \ldots, n$. Sei i fest. Da (f, k) eine ϵ-Approximation ist, gilt für $j = 1, 2, \ldots, i$:

  $$\left\|b_j - \tfrac{1}{k} \cdot f(b_j)\right\| \leq \epsilon \cdot \|b_j\| = \epsilon \cdot \lambda_j(L)$$

  Es gilt nach Dreiecksungleichung für $j = 1, 2, \ldots, i$:

  $$\left\|\tfrac{1}{k} \cdot f(b_j)\right\| \leq \|b_j\| + \left\|b_j - \tfrac{1}{k} \cdot f(b_j)\right\| \leq \lambda_j(L) + \epsilon \cdot \lambda_j(L) \tag{11.11}$$

  Da $f(L) \subseteq \mathbb{Z}^n$ ein volldimensionales Gitter ist, sind die Vektoren $\frac{1}{k} \cdot f(b_j) \in \bar{L}$, $j = 1, 2, \ldots, i$, linear unabhängig. Es gilt nach (11.11)

  $$\lambda_i\left(\bar{L}\right) \leq \lambda_j(L) + \epsilon \cdot \lambda_j(L)$$

  und wir erhalten die Behauptung.

- Wir zeigen $\lambda_i(L) - \lambda_i\left(\bar{L}\right) \leq \epsilon \cdot \lambda_i(L)$ für $i = 1, 2, \ldots, n$. Sei i fest. Es gilt für $j = 1, 2, \ldots, i$:

  $$\left\|\bar{b}_j - k \cdot f^{-1}\left(\bar{b}_j\right)\right\| \leq \epsilon \cdot \left\|k \cdot f^{-1}\left(\bar{b}_j\right)\right\| \leq \epsilon \cdot \left\|k \cdot f^{-1}\left(\bar{b}_j\right) - \bar{b}_j\right\| + \epsilon \cdot \left\|\bar{b}_j\right\|$$

  Wir erhalten für $j = 1, 2, \ldots, i$:

  $$(1 - \epsilon) \cdot \left\|\bar{b}_j - k \cdot f^{-1}\left(\bar{b}_j\right)\right\| \leq \epsilon \cdot \left\|\bar{b}_j\right\| \tag{11.12}$$

  Die Vektoren $k \cdot f^{-1}\left(\bar{b}_j\right) \in L$, $j = 1, 2, \ldots, i$, sind linear unabhängig. Es folgt für m mit $1 \leq m \leq i$, und $\left\|k \cdot f^{-1}\left(\bar{b}_m\right)\right\| = \max_{1 \leq j \leq i} \left\|k \cdot f^{-1}\left(\bar{b}_j\right)\right\|$:

  $$\begin{aligned}
  (1 - \epsilon) \cdot \lambda_i(L) &\leq (1 - \epsilon) \cdot \left\|k \cdot f^{-1}\left(\bar{b}_m\right)\right\| && \text{(wegen Wahl von } m\text{)} \\
  &\leq (1 - \epsilon) \cdot \left\|\bar{b}_m - k \cdot f^{-1}\left(\bar{b}_m\right)\right\| + (1 - \epsilon) \cdot \left\|\bar{b}_m\right\| && \text{(Dreiecksungleichung)} \\
  &\leq \epsilon \cdot \left\|\bar{b}_m\right\| + (1 - \epsilon) \cdot \left\|\bar{b}_m\right\| && \text{(wegen (11.12))} \\
  &\leq \lambda_i\left(\bar{L}\right) && \text{(wegen } m \leq i\text{)}
  \end{aligned}$$

  Wir erhalten die Behauptung $\lambda_i(L) - \lambda_i\left(\bar{L}\right) \leq \epsilon \cdot \lambda_i(L)$ für $i = 1, 2, \ldots, n$.

Insgesamt haben wir $\left|\lambda_i(L) - \frac{1}{k} \cdot \lambda_i(f(L))\right| \leq \epsilon \cdot \lambda_i(L)$ für $i = 1, 2, \ldots, n$ bewiesen. ■

Wenn wir in Satz 11.4.11 $\epsilon \leq (3 \cdot \lambda_n(L) + 1)^{-1}$ wählen, können wir die sukzessiven Minima von L durch die der ϵ-Approximation bestimmen, denn die sukzessiven Minima sind für beide Gitter ganzzahlig. Diese Polynomialzeit-Reduktion kann auch auf Gitter $L \subseteq \mathbb{Z}^n$ mit $\text{Rang}(L) < n$ erweitert werden [PaSchn87].

Kapitel 12

Gröbner-Basen

In diesem und den folgenden Kapiteln sei S ein kommutativer Ring mit Einselement 1 und $R := S[x_1, x_2, \ldots, x_n]$ der Ring der Polynome in den Variablen $x_1, x_2, \ldots, x_n$ mit Koeffizienten in S.

Wir beschäftigen uns mit der Reduktion von Basen zu Idealen im Ring $S[x_1, x_2, \ldots, x_n]$. In diesem Kapitel stellen wir das Konzept der Gröbner-Basen vor. Im Kapitel 13 werden wir Reduktions-Algorithmen für Gröbner-Basen entwickeln und Anwendungen geben.

12.1 Definition und Eigenschaften

Wir definieren zum Ring R:

Definition 12.1.1 (Ideal)
Die Teilmenge $I \subseteq R$ ist ein Ideal, wenn:

a) I ist eine additive Untergruppe von R.

b) $R \cdot I \subseteq I$.

Da das Ideal I eine additive Gruppe ist, gilt $I \neq \emptyset$.

Definition 12.1.2 (Basis oder erzeugendes System eines Ideals)
Zu $a_1, a_2, \ldots, a_k \in R$ ist

$$(a_1, a_2, \ldots, a_k) := \left\{ \sum_{i=1}^{k} r_i a_i \;\middle|\; r_1, r_2, \ldots, r_k \in R \right\}$$

das Ideal mit Basis oder erzeugendem System $a_1, a_2, \ldots, a_k$.

Im Vergleich zu Gittern müssen die Elemente der Idealbasis nicht linear unabhängig sein (Koeffizienten aus dem Ring R), d.h. die Basis ist im allgemeinen nicht minimal.

Definition 12.1.3 (Potenzprodukte,Terme)
Die Menge der Potenzprodukte (Terme) in den Variablen $x_1, x_2, \ldots, x_n$ ist (setze $x_i^0 := 1$):

$$PP(x_1, x_2, \ldots, x_n) := \{ x_1^{e_1} x_2^{e_2} \ldots x_n^{e_n} \mid e_1, e_2, \ldots, e_n \in \mathbb{N}_0 \} \subseteq R$$

Der Grad des Potenzproduktes $x_1^{e_1} x_2^{e_2} \ldots x_n^{e_n}$ ist $\sum_{i=1}^{n} e_i$, d.h. die Summe der Exponenten.

Seien $p := x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}$ und $q := x_1^{e_1} x_2^{e_2} \ldots x_n^{e_n}$ Potenzprodukte. Das Potenzprodukt p ist ein Vielfaches von q, also teilt p q, falls $e_i \leq d_i$ für $i = 1, 2, \ldots, n$. Das kleinste, gemeinsame Vielfache von p und q ist:

$$\mathrm{kgV}(p, q) = x_1^{\max(d_1, e_1)} x_2^{\max(d_2, e_2)} \cdots x_n^{\max(d_n, e_n)}$$

Der größte gemeinsame Teiler von p und q ist:

$$\mathrm{ggT}(p, q) = x_1^{\min(d_1, e_1)} x_2^{\min(d_2, e_2)} \cdots x_n^{\min(d_n, e_n)}$$

Das folgende Lemma von Dickson werden wir in einigen Beweisen in diesem Kapitel benutzen:

Lemma 12.1.4 (Dicksons Lemma)
Jede Menge $X \subseteq PP(x_1, x_2, \ldots, x_n)$ von Potenzprodukten enthält eine endliche Teilmenge $Y \subseteq X$, so daß jedes $p \in X$ ein Vielfaches eines Potenzprodukts aus Y ist.

Beweis. Siehe [Mishra93, Lemma 2.2.1]. ■

Für die Reduktion benötigen wir eine Ordnung auf den Potenzprodukten. Wir betrachten nur admissible (zulässige) Ordnungen:

Definition 12.1.5 (Admissible (zulässige) Ordnung)
Eine totale Ordnung $\leq_A$ auf den Potenzprodukten $PP(x_1, x_2, \ldots, x_n)$ ist admissible (zulässig), wenn für alle $p, p', q \in PP(x_1, x_2, \ldots, x_n)$ gilt:

- $1 \leq_A p$
- $p \leq_A p' \quad \Longrightarrow \quad pq \leq_A p'q$

Wir schreiben $p <_A q$, falls $p \leq_A q$ und $q \neq p$ ist. Falls q ein Vielfaches von p ist, gilt $p \leq_A q$.

Lemma 12.1.6
Jede admissible Ordnung $\leq_A$ auf $PP(x_1, x_2, \ldots, x_n)$ ist eine Wohlordnung, d.h. jede nicht-leere Teilmenge von $PP(x_1, x_2, \ldots, x_n)$ hat bezüglich $\leq_A$ ein kleinstes Element.

Beweis. Anwendung von Dicksons Lemma [Mishra93, Lemma 2.2.3]. ■

Ein Beispiel für eine admissible Ordnung ist die lexikographische Ordnung:

Definition 12.1.7 (Lexikographische Ordnung)
Seien $p := x_1^{d_1} \cdots x_n^{d_n}$ und $q := x_1^{e_1} \ldots x_n^{e_n}$ Potenzprodukte aus $PP(x_1, x_2, \ldots, x_n)$. Die lexikographische Ordnung $\leq_{\mathrm{lex}}$ ist auf $PP(x_1, x_2, \ldots, x_n)$ wie folgt definiert: Es ist $p <_{\mathrm{lex}} q$, falls

a) für ein i gilt $d_i \neq e_i$ und

b) für das minimale i mit $d_i \neq e_i$ gilt $d_i < e_i$.

Es gilt $x_1 >_{\mathrm{lex}} x_2 >_{\mathrm{lex}} \cdots >_{\mathrm{lex}} x_n > 1$.

Definition 12.1.8 (Monom)
Ein Monom $m \in R$ ist ein Element $m = ap$ mit $a \in S \setminus \{0\}$ und $p \in PP(x_1, x_2, \ldots, x_n)$.

Im folgenden sei $\leq_A$ eine feste, admissible Ordnung. Jedes Polynom $f \in R\backslash\{0\}$ hat eine eindeutige Darstellung

$$f = \sum_{i=1}^{k} a_i p_i \qquad \text{mit } a_i \in S \setminus \{0\} \text{ und } p_i \in \mathrm{PP}(x_1, x_2, \ldots, x_n) \text{ für } i = 1, 2, \ldots, k \tag{12.1}$$

mit $p_1 >_A p_2 >_A \cdots >_A p_k$. Wir führen zur Darstellung (12.1) folgende Bezeichnungen ein, die von der zugrundeliegenden admissiblen Ordnung abhängig sind:

$\mathrm{Hmono}(f) := a_1 p_1$	Head-Monom
$\mathrm{Hcoef}(f) := a_1$	Head-Koeffizient
$\mathrm{Hterm}(f) := p_1$	Head-Term
$\mathrm{Tail}(f) := f - \mathrm{Hmono}(f) = \sum_{i=2}^{k} a_i p_i$	Schwanz

Falls der Koeffizientenring S nullteilerfrei ist, gilt $\mathrm{Hmono}(fg) = \mathrm{Hmono}(f) \cdot \mathrm{Hmono}(g)$ und insbesondere $\mathrm{Hterm}(fg) = \mathrm{Hterm}(f) \cdot \mathrm{Hterm}(g)$. Wir definieren mit Hilfe von Head-Monom-Idealen Gröbner-Basen:

Definition 12.1.9 (Head-Monom-Ideal)
Zu $G \subseteq R$ sei $\mathrm{Head}(G)$ *folgendes Ideal:*

$$\mathrm{Head}(G) := \big(\{\mathrm{Hmono}(g) \mid g \in G\} \big) = \left\{ \sum_{\text{endl. Summe}} r_i \cdot \mathrm{Hmono}(g_i) \;\middle|\; g_i \in G,\, r_i \in R \right\}$$

Gröbner-Basen (G-Basen) wurden 1965 von B. Buchenberger in seiner Disseration [Bb65] eingeführt und nach seinem Doktorvater, W. Gröbner, benannt. Das Gebiet wurde aber erst in den siebziger Jahren populär (vergleiche Literaturliste in [BeWe93, Mishra93]).

Definition 12.1.10 (Gröbner-Basis)
Zum Ideal $I \subseteq R$ ist $G \subseteq I$ Gröbner-Basis von I, wenn $\mathrm{Head}(G) = \mathrm{Head}(I)$.

Da $G := I$ die obige Behauptung erfüllt, hat jedes Ideal I eine Gröbner-Basis. Ein Ideal kann mehrere Gröbner-Basen haben. Zum Beispiel ist zu einer Gröbner-Basis G auch G' mit $G \subseteq G' \subseteq I$ ebenfalls eine Gröbner-Basis des Ideals I.

Aus $G \subseteq I$ folgt $\mathrm{Head}(G) \subseteq \mathrm{Head}(I)$. Daher genügt es zum Nachweis der Gröbner-Basen-Eigenschaft von G zum Ideal I zu zeigen, daß $\mathrm{Head}(G) \supseteq \mathrm{Head}(I)$ gilt. Folgender Satz rechtfertigt den Ausdruck „Basis“ für Gröbner-Basen eines Ideals:

Satz 12.1.11
Sei $I \subseteq R$ ein Ideal von R und $G \subseteq I$. Dann gilt:

$$\mathrm{Head}(G) = \mathrm{Head}(I) \quad \Longrightarrow \quad (G) = I$$

Das heißt: Jede Gröbner-Basis eines Ideals erzeugt das Ideal.

Beweis. Wegen $G \subseteq I$ ist $(G) \subseteq I$. Angenommen, es gelte $(G) \subsetneq I$. Wähle $f \in I \setminus (G)$ mit $\mathrm{Hmono}(f)$ minimal bezüglich $\leq_A$. Wegen

$$\mathrm{Hmono}(f) \in \mathrm{Head}(G) = \mathrm{Head}(I)$$

gibt es eine Darstellung als endliche Summen mit $t_i \in R$:

$$\text{Hmono}(f) = \sum_{g_i \in G} t_i \cdot \text{Hmono}(g_i) \tag{12.2}$$

$$\text{Hterm}(f) = \sum_{g_i \in G} t_i \cdot \text{Hterm}(g_i) = \sum_{g_i \in G} \text{Hterm}(t_i \cdot g_i) \tag{12.3}$$

Betrachte:

$$f' := \text{Tail}(f) - \sum_{g_i \in G} t_i \cdot \text{Tail}(g_i) \overset{(12.2)}{=} f - \underbrace{\sum_{g_i \in G} t_i \cdot g_i}_{\in (G)}$$

Es gilt $f' \in I$. Offenbar ist $f' \notin (G)$, da sonst $f \in (G)$ im Widerspruch zu $f \in I \setminus (G)$ wäre. Wegen (12.2) ist:

$$\text{Hmono}(f') <_A \text{Hmono}(f)$$

Dies ist ein Widerspruch zur Wahl von f: $\text{Hmono}(f)$ war minimal bezüglich $\leq_A$. ■

Es gilt:

Satz 12.1.12
Sei $I \subseteq R$ ein Ideal von R und $G \subseteq I$. G ist genau dann Gröbner-Basis von I, wenn jedes $h \in I$ von der Form

$$h = \sum_{g_i \in G} f_i g_i \qquad \textit{mit } f_i \in R$$

ist, so daß $\text{Hterm}(f_i g_i) = \text{Hterm}(f_i) \cdot \text{Hterm}(g_i) \leq_A \text{Hterm}(h)$, d.h. der Grad der f_i ist beschränkt.

Beweis. Wir zeigen beide Richtungen:

„⇒“ Sei $h \in I$. Induktiv angenommen, die Behauptung gilt für alle $h' <_A h$, d.h. $\text{Hterm}(h') <_A \text{Hterm}(h)$. Es gilt, da nach Voraussetzung G eine Gröbner-Basis von I ist:

$$\text{Hmono}(h) \in \text{Head}(I) = \text{Head}(G)$$

Es gibt eine Darstellung mit $a_i \in S$ und $p_i \in \text{PP}(x_1, x_2, \ldots, x_n)$:

$$\text{Hmono}(h) = \sum_{g_i \in G} a_i \cdot p_i \cdot \text{Hmono}(g_i), \tag{12.4}$$

so daß für alle i gilt:

$$p_i \cdot \text{Hterm}(h_i) = \text{Hterm}(g_i) \tag{12.5}$$

Bilde:

$$h' := \text{Tail}(h) - \sum_{g_i \in G} a_i \cdot p_i \cdot \text{Tail}(g_i) \; = h - \sum_{g_i \in G} a_i \cdot p_i \cdot g_i$$

Beachte: Das führende Monom hebt sich wegen (12.4) und (12.5) weg. Es ist $\text{Hterm}(h') <_A \text{Hterm}(h)$. Aus der Induktionsannahme angewandt auf h' folgt die Darstellung

$$h = h' + \sum_{g_i \in G} a_i \cdot p_i \cdot g_i = \sum_{g_i \in G} f_i' \cdot g_i + \sum_{g_i \in G} a_i \cdot p_i \cdot g_i$$

mit:

1. $\text{Hterm}(f_i') \cdot \text{Hterm}(g_i) \leq_A \text{Hterm}(h') <_A \text{Hterm}(h)$
2. $\text{Hterm}(p_i) \cdot \text{Hterm}(g_i) = \text{Hterm}(h)$

„⇐“ Es genügt zu zeigen: $\text{Head}(G) \supseteq \text{Head}(I)$. Sei $h \in I$ von der Form

$$h = \sum_{g_i \in G} a_i \cdot p_i \cdot g_i \qquad \text{mit } a_i \in S,\ p_i \in \text{PP}(x_1, x_2, \ldots, x_n)$$

mit $p_i \cdot \text{Hterm}(g_i) \leq_A \text{Hterm}(h)$. Sei:

$$L := \{g_i \ : \ \text{Hterm}(h) = p_i \cdot \text{Hterm}(g_i)\} \neq \emptyset$$

Wir erhalten:

$$\begin{aligned} \text{Hmono}(h) &= \text{Hcoef}(h) \cdot \text{Hterm}(h) \\ &= \sum_{g_i \in L} a_i \cdot \text{Hcoef}(g_i) \cdot p_i \cdot \text{Hterm}(g_i) \\ &= \sum_{g_i \in L} a_i \cdot p_i \cdot \text{Hmono}(g_i) \end{aligned}$$

Also gilt $\text{Hmono}(h) \in \text{Head}(G)$, und G ist eine Gröbner-Basis des Ideals I.

■

Falls die Koeffizienten der Polynome statt aus einem beliebigem Ring aus einem Körper sind, hat jedes Ideal eine endliche Gröbner-Basis. Wir werden sehen, daß dies auch für Noether'sche Ringe (Definition 12.1.14) gilt.

Satz 12.1.13
Sei K ein Körper. Dann hat jedes Ideal $I \subseteq K[x_1, x_2, \ldots, x_n]$ eine endliche Gröbner-Basis.

Beweis. Sei

$$X := \{\text{Hterm}(f) \ : \ f \in I\} \subseteq \text{PP}(x_1, x_2, \ldots, x_n)$$

Nach Dicksons Lemma 12.1.4 gibt es ein endliches $Y \subseteq X$ mit

$$\forall x \in X: \quad \exists y \in Y: \quad y \text{ teilt } x \tag{12.6}$$

Definiere $\phi : Y \to I$ so, daß $\text{Hterm}\big(\phi(y)\big) = y$ für alle $y \in Y$. Damit ist $G := \phi(Y)$ eine Gröbner-Basis des Ideals I, denn:

$$\begin{aligned} \text{Head}(G) &= \big(\ \{\text{Hmono}(g) \ : \ g \in G\}\ \big) \\ &= \big(\ \{\text{Hterm}(g) \ : \ g \in G\}\ \big) && \text{(da Koeffizienten aus Körper)} \\ &= (Y) && \text{(da } G := \phi(Y) \text{ und } \text{Hterm}\big(\phi(y)\big) = y\text{)} \\ &= (X) && \text{(wegen Eigenschaft (12.6))} \\ &= \text{Head}(I) && \text{(da Koeffizienten aus Körper)} \end{aligned}$$

■

Dieser Satz überträgt sich auf Noether'sche Ringe:

Definition 12.1.14 (Noether'scher Ring)
Ein Ring R heißt Noether'sch, wenn jedes Ideal des Ringes endlich erzeugt ist.

Falls die Koeffizienten aus einem Noether'schen Ring S sind, ist auch $R = S[x_1, x_2, \ldots, x_n]$ ein Noether'scher Ring:

Satz 12.1.15 (Hilbert'scher Basissatz)
Ist S ein Noether'scher Ring, so auch $R = S[x_1, x_2, \ldots, x_n]$.

Beweis. Siehe u.a. Theorem 2.3.7 und Korollar 2.3.8 in [Mishra93]. ■

Bevor wir die Aussage aus Satz 12.1.13 für Noether'sche Ringe beweisen, fassen wir algebraische Eigenschaften Noether'scher Ringe zusammen:

Satz 12.1.16
Sei R ein Ring. Dann sind folgende drei Aussagen äquivalent:

1. *R ist ein Noether'scher Ring.*
2. *Jede aufsteigende Idealkette bricht ab, d.h.:*

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \quad \Longrightarrow \quad (\exists n \in \mathbb{N} : \ \forall i \geq n : \ I_i = I_n)$$

3. *Jede nicht-leere Menge von Idealen hat ein maximales Ideal (bezüglich Inklusion).*

Beweis. Siehe u.a. Proposition 2.3.6 in [Mishra93]. ■

Wir übertragen Satz 12.1.13 auf Noether'sche Ringe:

Satz 12.1.17
Sei S ein Noether'scher Ring. Dann hat jedes Ideal $I \subseteq R := S[x_1, x_2, \ldots, x_n]$ eine endliche Gröbner-Basis.

Beweis. Sei Y endliches, erzeugendes System des Ideals $\mathrm{Head}(I) \subseteq R$. Es gilt:

$$Y \subseteq \{\mathrm{Hmono}(f) \ : \ f \in I\}$$

Definiere $\phi : Y \to I$ so, daß $\mathrm{Hmono}\big(\phi(y)\big) = y$. Für $G := \phi(Y)$ gilt:

$$\mathrm{Head}(G) = (Y) = \mathrm{Head}(I)$$

■

Wir betrachten zum Abschluß ein Beispiel zu Gröbner-Basen:

Beispiel 12.1.18 (Gröbner-Basis)
Sei $R := S[x_1, x_2]$. Wir wählen als zulässige Ordnung die lexikographische Ordnung:

$$1 <_{\mathrm{lex}} x_2 <_{\mathrm{lex}} x_1$$

Seien $g_1 := x_1$ und $g_2 := x_1 x_2 - x_2$. Wir wollen eine Gröbner-Basis G zu dem Ideal

$$I := (g_1, g_2) = (x_1, x_1 x_2 - x_2)$$

bestimmen. Unser erster Ansatz ist $G := \{g_1, g_2\}$. Es gilt:

$$\mathrm{Head}(G) = \big(\, \mathrm{Hmono}(g_1), \mathrm{Hmono}(g_2) \,\big) = (x_1, x_1x_2)$$

Wir zeigen, daß G keine Gröbner-Basis zu I ist. Für eine Gröbner-Basis gilt, daß das Headmonom eines Polynoms $f = f_1g_1 + f_2g_2 \in I$ mit $f_1, f_2 \in R$ in $\mathrm{Head}(G)$ liegt. Dies gilt offenbar, wenn das Headmonom von f entweder das Headmonom des ersten oder zweiten Summanden ist. Jedoch kann sich das Headmonom wegheben:

$$f := x_2 \cdot g_1 + (-1) \cdot g_2 = x_1x_2 - x_1x_2 + x_2 = x_2 \tag{12.7}$$

Wegen $x_1 \nmid x_2$ ist $x_2 \notin (G)$, und G ist keine Gröbner-Basis von I. Dies tritt genau bei Polynomen des Typs

$$s_1 \cdot x_1 \cdot \frac{\mathrm{kgV}\,(x_1, x_1x_2)}{x_1} + s_2 \cdot (x_1x_2 - x_2) \cdot \frac{\mathrm{kgV}\,(x_1, x_1x_2)}{x_1x_2}$$

mit $s_1, s_2 \in S$ sowie $s_1 + s_2 = 0$ und deren Vielfachen auf. Wir werden in Kapitel 12.2 Polynome dieses Typs (auch für mehr als 2 Polynome) als S-Polynome zu G bezeichnen. Wenn wir diese durch $\mathrm{Head}(G)$ darstellen können (Syzygy-Bedingung), ist G eine Gröbner-Basis.

In unserem Beispiel fügen wir f aus (12.7) zu G hinzu. Nach diesem Prinzip arbeiten die Reduktions-Algorithmen, die wir in Kapitel 13 kennenlernen. Sie erweitern die Idealbasis durch Hinzunahme weiterer Polynome, die noch nicht in $\mathrm{Head}(G)$ liegen, zu einer Gröbner-Basis. In unserem Beispiel ist G nach der Hinzunahme von f eine Gröbner-Basis von I, denn wegen $G \subseteq I$ ist allgemein $\mathrm{Head}(G) \subseteq \mathrm{Head}(I)$ und aus

$$\mathrm{Head}(G) = \big(\, \mathrm{Hmono}(g_1), \mathrm{Hmono}(g_2), \mathrm{Hmono}(f) \,\big) = (x_1, x_2, x_1x_2)$$

folgt $\mathrm{Head}(G) \supseteq \mathrm{Head}(I)$, da $I \cap S = \{0\}$. ◇

12.2 S-Polynome und Syzygy-Bedingung

Wir lernen in diesem Abschnitt eine alternative Charakterisierung von Gröbner-Basen durch S-Polynome und Syzygy-Bedingung kennen. Wir fassen die äquivalenten Bedingungen in Satz 12.2.4 zusammen. Anschließend vereinfachen wir die S-Polynome für den Fall, daß der Koeffizientenring ein Körper oder der Ring der ganzen Zahlen ist.

Definition 12.2.1 (S-Polynome SP(G))
Zu $G \subseteq R$ besteht die Menge $\mathrm{SP}(G) \subseteq G$ der S-Polynome zu G aus den endlichen Summen:

$$h := \sum_{g_i \in G} s_i \cdot g_i \cdot \frac{m}{\mathrm{Hterm}(g_i)}$$

mit $s_i \in S \setminus \{0\}$, $\sum_i s_i \cdot \mathrm{Hcoef}(g_i) = 0$ und $m = \mathrm{kgV}\,(\{\mathrm{Hterm}(g_i \ : \ g_i \text{ aus Summe}\})$.

Für die S-Polynome gilt:

1. $\mathrm{Hterm}(g_i)$ teilt m.
2. Aus $G' \subseteq G$ folgt $\mathrm{SP}(G') \subseteq \mathrm{SP}(G)$.
3. Die führenden Terme (Head-Terme) von h „annulieren“ sich, d.h.:

   $$h = \sum_i s_i \cdot \mathrm{Tail}(g_i) \cdot \frac{m}{\mathrm{Hterm}(g_i)}$$

   Also gilt $\mathrm{Hterm}(h) <_A m$.

Wir definieren die Syzygy-Bedingung:

Definition 12.2.2 (Syzygy-Bedingung)
$G \subseteq R$ erfüllt die Syzygy-Bedingung, wenn jedes $h \in \mathrm{SP}(G)$ von folgender Form ist:

$$h = \sum_i f_i \cdot g_i \qquad \text{mit } f_i \in R,\ g_i \in G \text{ und } \mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(h) \tag{12.8}$$

Die Syzygy-Bedingung bildet eine alternative Charakterisierung von Gröbner-Basen:

Satz 12.2.3
Sei $I \subseteq R$ Ideal und $G \subseteq I$ mit $(G) = I$. G ist genau dann Gröbner-Basis von I, wenn G die Syzygy-Bedingung erfüllt.

Beweis. Wir zeigen beide Richtungen:

„⇒“ Sei $h \in \mathrm{SP}(G) \subseteq (G)$. Nach Satz 12.1.12 ist h von der Form (12.8).

„⇐“ Beweis durch Widerspruch. Sei $f \in I = (G)$.

$$f = \sum_i f_i \cdot g_i \qquad \text{mit } f_i \in R \text{ und } g_i \in G \tag{12.9}$$

mit $M := \max_i \{\mathrm{Hterm}(f_i \cdot g_i)\}$ bezüglich Ordnung $\leq_A$ minimal. Nach Satz 12.1.12 ist G Gröbner-Basis, falls $\mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(f)$. Angenommen, es sei $M >_A \mathrm{Hterm}(f)$. Dann annulieren sich die führenden Terme in (12.9):

$$\sum_{\mathrm{Hterm}(f_i \cdot g_i) = M} s_i \cdot \mathrm{Hcoef}(g_i) = 0 \qquad \text{mit } s_i := \mathrm{Hcoef}(f_i)$$

Also:

$$f = \underbrace{\sum_{\mathrm{Hterm}(f_i \cdot g_i) = M} s_i \cdot \frac{M}{\mathrm{Hterm}(g_i)} \cdot g_i}_{=:h} + \sum_i f_i' \cdot g_i \tag{12.10}$$

mit

$$f_i' := \begin{cases} \mathrm{Tail}(f_i) & \text{falls } \mathrm{Hterm}(f_i \cdot g_i) = M \\ f_i & \text{sonst} \end{cases}$$

Beachte: $\mathrm{Hterm}\,(f_i' \cdot g_i) <_A M$. Da nach Annahme $\mathrm{Hterm}(f) <_A M$, folgt $\mathrm{Hterm}(h) <_A M$. Für

$$m := \mathrm{kgV}\,\{\mathrm{Hterm}(g_i)\ :\ \mathrm{Hterm}(f_i \cdot g_i) = M\}$$

gilt $m \mid M$, da $\mathrm{Hterm}(g_i) \mid M$. Folglich:

$$h \in \tfrac{M}{m} \cdot \mathrm{SP}(G)$$

Es gibt ein $\bar{h} \in \mathrm{SP}(G)$ mit $h = \frac{M}{m} \cdot \bar{h}$. Wegen der Syzygy-Bedingung (12.8) ist $\bar{h}$ von der Form:

$$\bar{h} = \sum_{\bar{g}_i \in G} \bar{f}_i \cdot \bar{g}_i \qquad \text{mit } \mathrm{Hterm}\left(\bar{f}_i \cdot \bar{g}_i\right) \leq_A \mathrm{Hterm}\left(\bar{h}\right)$$

Es gilt:

$$\mathrm{Hterm}(\bar{f}_i \cdot \bar{g}_i) \leq_A \mathrm{Hterm}(\bar{h}) \leq_A \mathrm{Hterm}(h) <_A M$$

Aus der Darstellung (12.10) erhalten wir einen Widerspruch: M ist nicht minimal.

Wir haben die Aussage bewiesen. ■

Zusammenfassend erhalten wir als äquivalente Charakterisierung von Gröbner-Basen:

Satz 12.2.4
Sei $I \subseteq R$ Ideal und $G \subseteq I$, dann sind äquivalent:

a) G ist eine Gröbner-Basis des Ideals I.

b) $\mathrm{Head}(G) = \mathrm{Head}(I)$

c) $\forall h \in I : \exists f_i \in R : h = \sum_{g_i \in G} f_i g_i$ *(endliche Summe) mit* $\mathrm{Hterm}(g_i \cdot f_i) \leq_A \mathrm{Hterm}(h)$.

d) Es gilt $(G) = I$, und G erfüllt die Syzygy-Bedingung.

Wir betrachten im weiteren die Spezialfälle, daß der Koeffizientenring ein Körper oder $\mathbb{Z}$ ist. In beiden Fällen können wir die Syzygy-Bedingung durch einfachere S-Polynome definieren:

Definition 12.2.5 (S-Polynome $\mathrm{SP}_K(G)$ und $\mathrm{SP}_{\mathbb{Z}}(G)$)
Sei K ein Körper. Zu $G \subseteq K[x_1, x_2, \ldots, x_n]$ bestehe die Menge $\mathrm{SP}_K(G)$ *aus den Polynomen ($g_j, g_k \in G$):*

$$S(g_j, g_k) := \frac{m}{\mathrm{Hmono}(g_j)} \cdot g_j - \frac{m}{\mathrm{Hmono}(g_k)} \cdot g_k \tag{12.11}$$

mit $m := \mathrm{kgV}(\mathrm{Hmono}(g_i), \mathrm{Hmono}(g_k))$. *Zu $G \subseteq \mathbb{Z}[x_1, x_2, \ldots, x_n]$ bestehe die Menge* $\mathrm{SP}_{\mathbb{Z}}(G)$ *aus den Polynomen* (12.11).

Beachte, daß sich der führende Term in der Differenz in (12.11) weghebt. Zu $G \subseteq K[x_1, x_2, \ldots, x_n]$ bzw. $G \subseteq \mathbb{Z}[x_1, x_2, \ldots, x_n]$ und $g_j, g_k \in G$ bezeichnen wir:

$$\begin{aligned} c_j &:= \mathrm{Hcoef}(g_j) \\ T_j &:= \mathrm{Hterm}(g_j) \\ c_{j,k} &:= \mathrm{kgV}(c_j, c_k) \\ T_{j,k} &:= \mathrm{kgV}(T_j, T_k) \end{aligned}$$

Mit diesen Bezeichnungen schreibt sich (12.11) als:

$$S(g_j, g_k) := \frac{c_{j,k}}{c_j} \cdot \frac{T_{j,k}}{T_j} \cdot g_j - \frac{c_{j,k}}{c_k} \cdot \frac{T_{j,k}}{T_k} \cdot g_k \tag{12.12}$$

Die Syzygy-Bedingung ist identisch mit der im allgemeinen Fall:

Definition 12.2.6 (Syzygy-Bedingung mit $\mathrm{SP}_K(G)$ und $\mathrm{SP}_{\mathbb{Z}}(G)$)
$G \subseteq K[x_1, x_2, \ldots, x_n]$ erfüllt die Syzygy-Bedingung mit $\mathrm{SP}_K(G)$, *wenn jedes $h \in \mathrm{SP}_K(G)$ von folgender Form ist:*

$$h = \sum_i f_i \cdot g_i \qquad \text{mit } f_i \in R,\ g_i \in G \text{ und } \mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(h) \tag{12.13}$$

$G \subseteq \mathbb{Z}[x_1, x_2, \ldots, x_n]$ erfüllt die Syzygy-Bedingung mit $\mathrm{SP}_{\mathbb{Z}}(G)$, *wenn jedes $h \in \mathrm{SP}_{\mathbb{Z}}(G)$ von der Form* (12.13) *ist.*

Wir zeigen, daß genau dann G eine Gröbner-Basis ist, wenn G die Syzygy-Bedingung mit $\mathrm{SP}_K(G)$ bzw. $\mathrm{SP}_{\mathbb{Z}}(G)$ erfüllt:

Satz 12.2.7
$G \subseteq K[x_1, x_2, \ldots, x_n]$ ist Gröbner-Basis genau dann, wenn G die Syzygy-Bedingung mit $\mathrm{SP}_K(G)$ erfüllt.

Beweis. Wir zeigen beide Richtungen:

„⇒“ Da $\mathrm{SP}_K(G) \subseteq G$, erfüllt $\mathrm{SP}_K(G)$ die Syzygy-Eigenschaft.

„⇐“ Beweis durch Widerspruch: Sei $f \in (G)$

$$f = \sum_i f_i g_i \qquad f_i \in R, g_i \in G \tag{12.14}$$

mit $M := \max_i \mathrm{Hterm}(f_i \cdot g_i)$ minimal bezüglich der Ordnung $\leq_A$. Angenommen, es gelte $M >_A \mathrm{Hterm}(f)$. Nach Voraussetzung erfüllt G die Syzygy-Bedingung mit $\mathrm{SP}_{\mathbb{Z}}(G)$, wenn jedes $h \in \mathrm{SP}_{\mathbb{Z}}(G)$ von der Form ist:

$$h = \sum_i f_i \cdot g_i \qquad \text{mit } f_i \in R,\ g_i \in G \text{ und } \mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(h)$$

Sei o.B.d.A. $\{1, 2, \ldots, k\} := \{i \mid \mathrm{Hterm}(f_i \cdot g_i) = M\}$. Da sich die führenden Terme wegheben, ist $k \geq 2$. Für $c_i := \mathrm{Hcoef}(g_i)$ und $s_i := \mathrm{Hcoef}(f_i)$ definiere die Vektoren:

$$\vec{c} := (c_1, c_2, \ldots, c_k) \qquad \vec{s} := (s_1, s_2, \ldots, s_k)$$

Da sich in (12.14) die führenden Terme wegheben, erhalten wir:

$$\vec{c} \cdot \vec{s}^{\mathsf{T}} = \sum_{i=1}^{k} c_i s_i = 0$$

Wie im Beweis zu Satz 12.2.3 auf Seite 146 gilt mit $T_i = \mathrm{Hterm}(g_i)$:

$$f = \underbrace{\sum_{\mathrm{Hterm}(f_i \cdot g_i) = M} s_i \cdot \frac{M}{T_i} \cdot g_i}_{=:h} + \sum_i f_i' \cdot g_i \tag{12.15}$$

mit

$$f_i' := \begin{cases} \mathrm{Tail}(f_i) & \text{falls } \mathrm{Hterm}(f_i \cdot g_i) = M \\ f_i & \text{sonst} \end{cases}$$

Beachte: $\mathrm{Hterm}\,(f_i' \cdot g_i) <_A M$. Da nach Annahme $\mathrm{Hterm}(f) <_A M$, folgt $\mathrm{Hterm}(h) <_A M$.

Behauptung 1: Das Polynom h ist von der Form

$$h = \sum_{i=2}^{k} a_i \cdot S(g_1, g_k) \cdot S_i$$

mit $a_i \in K$, $S_i \in \mathrm{PP}(x_1, x_2, \ldots, x_n)$ und $T_{1,i} \cdot S_i = M$.

BEWEIS. Die Vektoren $b_2, b_3, \ldots, b_k \in K^k$ mit

$$\begin{bmatrix} b_2 \\ b_3 \\ \vdots \\ b_{k-1} \\ b_k \end{bmatrix} := \begin{bmatrix} \frac{c_{1,2}}{c_1} & -\frac{c_{1,2}}{c_2} & 0 & \cdots & \cdots & 0 \\ \frac{c_{1,3}}{c_1} & 0 & -\frac{c_{1,3}}{c_3} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{c_{1,k-1}}{c_1} & 0 & \cdots & 0 & -\frac{c_{1,k-1}}{c_{k-1}} & 0 \\ \frac{c_{1,k}}{c_1} & 0 & \cdots & \cdots & 0 & -\frac{c_{1,k}}{c_k} \end{bmatrix} \tag{12.16}$$

bilden eine Basis von $V := K^k \cap \operatorname{span}(\vec{c})^{\perp}$, des $(k-1)$-dimensionalen Vektorraums der senkrecht auf $\vec{c}$ stehenden Vektoren aus K^k. Da K ein Körper ist, existieren die Inversen. Zum Vergleich: Das S-Polynom $S(g_1, g_i)$ aus $\mathrm{SP}_K(G)$ lautet:

$$S(g_1, g_i) = \frac{c_{1,i}}{c_1} \cdot \frac{T_{1,i}}{T_1} g_1 - \frac{c_{1,i}}{c_i} \cdot \frac{T_{1,i}}{T_i} g_i$$

Da $\vec{s} = (s_1, s_2, \ldots, s_k)$ in V liegt, gibt es Koeffizienten $a_2, a_3, \ldots, a_k \in K$ mit:

$$(s_1, s_2, \ldots, s_k) = \sum_{i=2}^{k} b_i a_i$$

Wähle $S_i \in \mathrm{PP}(x_1, x_2, \ldots, x_n)$ so, daß $T_{1,i} \cdot S_i = M$. Nach Definition von h aus (12.15) gilt:

$$h = \sum_{\mathrm{Hterm}(f_i \cdot g_i) = M} s_i \cdot \frac{M}{T_i} \cdot g_i$$

Aus der Wahl von S_i folgt:

$$h = \sum_{i=2}^{k} a_i \cdot S(g_1, g_i) \cdot S_i$$

Dies war zu zeigen. □

Es ist:

$$S(g_1, g_i) = \frac{c_{1,i}}{c_1} \cdot \frac{T_{1,i}}{T_1} g_1 - \frac{c_{1,i}}{c_i} \cdot \frac{T_{1,i}}{T_i} g_i$$

Beide Summanden haben den führenden Term $T_{1,i}$ und führenden Koeffizienten $c_{1,i}$, so daß sich dieser weghebt. Wegen $T_{1,i} \cdot S_i = M$ gilt daher:

$$\mathrm{Hterm}(S(g_1, g_i) \cdot S_i) <_A M \qquad \text{für } i = 2, 3, \ldots, k$$

Weil G die Syzygy-Bedingung mit $\mathrm{SP}_{\mathbb{Z}}(G)$ erfüllt, ist h von der Form

$$h = \sum_{g_i \in G} \bar{f}_i \cdot g_i \qquad \text{mit } \mathrm{Hterm}(\bar{f}_i \cdot g_i) \leq_A \mathrm{Hterm}\,(S(g_1, g_i) \cdot S_i) <_A M$$

Dies ist ein Widerspruch zur Minimalität von M.

■

Analog zu Satz 12.2.7 gilt für den Koeffizientenring $\mathbb{Z}$:

Satz 12.2.8
$G \subseteq \mathbb{Z}[x_1, x_2, \ldots, x_n]$ ist Gröbner-Basis genau dann, wenn G die Syzygy-Bedingung mit $\mathrm{SP}_{\mathbb{Z}}(G)$ erfüllt.

Beweis. Siehe [Mishra93, Lemma 3.2.3]. ■

Beachte: Die Basis aus (12.16) ist keine Basis des ganzzahligen Gitters

$$L := \left\{ \vec{x} \in \mathbb{Z}^k \;\middle|\; \vec{x} \cdot \vec{c} = 0 \right\}$$

mit $c \in \mathbb{Z}^k$. Wir betrachten den Fall $k = 3$ mit $\vec{c} := (2, 3, 5)$. Die Zeilenvektoren

$$\begin{bmatrix} 6/2 & -6/3 & 0 \\ 10/2 & 0 & -10/5 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 0 \\ 5 & 0 & -2 \end{bmatrix}$$

bilden keine Basis des Gitters L, da der Vektor $\vec{x} = (0, 5, -3) \in L$ nicht als ganzzahlige Linearkombination dargestellt werden kann.

Kapitel 13

Reduktion und Gröbner-Basen

In diesem Kapitel formulieren wir Reduktions-Algorithmen für Gröbner-Basen, die wir im Kapitel 12 eingeführt haben. Sei S ein kommutativer Ring mit 1 und $R := S[x_1, x_2, \ldots, x_n]$ der Ring der Polynome in den Variablen $x_1, x_2, \ldots, x_n$ mit Koeffizienten in S.

13.1 Grundbegriffe

Wir führen einen Reduktionsbegriff für $f \in R \setminus \{0\}$ ein:

Definition 13.1.1 (Reduzierbar und Retrakt modulo G)
Sei $G \subseteq R$ Gröbner-Basis. $f \in R \setminus \{0\}$ heißt reduzierbar (reduzibel) modulo G, falls

$$\mathrm{Hmono}(f) \in \mathrm{Head}(G)$$

Falls insbesondere

$$\mathrm{Hmono}(f) = \sum_i a_i p_i \,\mathrm{Hmono}(g_i)$$

mit $a_i \in S$, $g_i \in G$, $p_i \in PP(x_1, x_2, \ldots, x_n)$ und für alle i gilt: $p_i \cdot \mathrm{Hterm}(g_i) = \mathrm{Hterm}(f)$. Dann heißt

$$h = f - \sum_i a_i p_i g_i$$

Retrakt von f modulo G.

Für den Retrakt h von f ist $\mathrm{Hterm}(h) <_A \mathrm{Hterm}(f)$. Wir schreiben $f \xrightarrow{G} h$, falls h der Retrakt von f modulo G ist. Es bezeichne $\xrightarrow[*]{G}$ den reflexiven (d.h. das Polynom darf auf sich selbst reduziert werden), transitiven (d.h. beliebig viele Reduktionsschritte) Abschluß von $\xrightarrow{G}$.

Definition 13.1.2 (Normalformen $\mathrm{NF}_G(f)$)
Die Menge der $\mathrm{NF}_G(f)$ der Normalformen von f modulo G besteht aus allen Polynomen $h \in R$ mit:

$$f \xrightarrow[*]{G} h$$

und h ist nicht reduzierbar modulo G.

Zu jedem $f \in R$ existiert mindestens eine Normalform:

Satz 13.1.3
Sei $G \subseteq R$. Für alle $f \in R$ ist $\mathrm{NF}_G(f) \neq \emptyset$.

Beweis. Angenommen, $\mathrm{Hterm}(f)$ sei minimal bezüglich $<_A$ mit $\mathrm{NF}_G(f) = \emptyset$. Fallunterscheidung:

- 1. Fall: $f \xrightarrow{G} h$, $\mathrm{Hterm}(h) <_A \mathrm{Hterm}(f)$. Wegen der Minimalität ist $\mathrm{NF}_G(f) \supseteq \mathrm{NF}_G(h) \neq \emptyset$ – Widerspruch.
- 2. Fall: f ist nicht reduzibel modulo G, d.h. $f \in \mathrm{NF}_G(f)$ – Widerspruch.

■

Wir erhalten eine weitere Charakterisierung von Gröbner-Basen:

Satz 13.1.4
Zu $G \subseteq R$ und $I = (G)$ sind folgende Aussagen äquivalent:

a) G ist Gröbner-Basis, d.h. $\mathrm{Head}(I) = \mathrm{Head}(G)$.

b) $\forall f \in (G): \quad f \xrightarrow[*]{G} 0$

c) $\forall f \in \mathrm{SP}(G): \quad f \xrightarrow[*]{G} 0$

Beweis. Wir führen einen Ringbeweis:

- „$a) \Rightarrow b)$“

  Induktion über $\mathrm{Hterm}(f)$, $f \in (G)$, bezüglich $<_A$.

  - Verankerung: $0 \xrightarrow[*]{G} 0$
  - Wegen $\mathrm{Hmono}(f) \in \mathrm{Head}(I)$ gibt es ein $h \in (G)$ mit $f \xrightarrow{G} h$. Wegen $\mathrm{Hterm}(h) <_A \mathrm{Hterm}(f)$ gilt nach Induktionsannahme: $h \xrightarrow[*]{G} 0$. Also:

    $$f \xrightarrow{G} h \xrightarrow[*]{G} 0$$

- „$b) \Rightarrow a)$“

  Aus $f \xrightarrow[*]{G} 0$ folgt durch Induktion über die Anzahl der Reduktionsschritte, daß f von der Form ist:

  $$f = \sum_i f_i g_i \qquad f_i \in R, g_i \in G \text{ mit } \mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(f)$$

  Nach Satz 12.1.12 folgt, daß G Gröbner-Basis von I ist.

- „$c) \Rightarrow a)$“

  Sei $f \in \mathrm{SP}(G)$. Wegen $f \xrightarrow[*]{G} 0$ ist jedes f von der Form

  $$f = \sum_i f_i g_i \qquad f_i \in R, g_i \in G \text{ mit } \mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(f),$$

  d.h. G erfüllt die Syzygy-Bedingung. Nach Satz 12.2.3 ist G Gröbner-Basis von I.

Wir haben die Aussage des Satzes bewiesen. ■

Bezeichnung 13.1.5 (Modul $[G]_S$)
Zu $G \subseteq R = S[x_1, x_2, \ldots, x_n]$ bezeichne $[G]_S$ die Menge der endlichen Summen:

$$[G]_S := \left\{ \sum_{g_i \in G} s_i \cdot g_i \;\middle|\; s_i \in S \right\} \subseteq R$$

$[G]_S$ ist ein S-Modul. Zur Erinnerung: Sei S ein Ring und $(M, +)$ eine abelsche Gruppe, so daß S auf M linear durch

$$\circ : S \times M \to M \quad \text{mit} \quad (s, x) \mapsto s \circ x$$

operiert, so ist M ein S-Modul. „Linear“ bedeutet, daß für alle $s, t \in S$ und $x, y \in M$ gilt:

$$\begin{aligned} s \circ (x +_M y) &= (s \circ x) +_M (s \circ y) \\ (s +_S t) \circ x &= (s \circ x) +_M (s \circ y) \\ (s \cdot_S t) \circ x &= s \circ (t \circ x) \\ 1 \circ x &= x \end{aligned}$$

Die Reduktions-Algorithmen werden Basen für $[\mathrm{SP}(G)]_S$ bestimmen. Dies setzt voraus, daß diese endlich erzeugt sind. Da wir uns auf Noether'sche Ringe beschränken, gilt diese Voraussetzung nach folgendem Lemma:

Lemma 13.1.6
Für einen Noether'schen Ring S und $G \subseteq R$ endlich, ist $[\mathrm{SP}(G)]_S$ endlich erzeugt.

Beweis. Siehe Proposition 2.4.2 in [Mishra93]. ■

Definition 13.1.7 (Noether'scher Modul)
Ein S-Modul heißt Noether'sch, wenn jeder Submodul endlich erzeugt ist.

Es gilt:

Lemma 13.1.8
Für einen Noether'scher Ring S ist S^n mit $n \in \mathbb{N}$ ein Noether'scher Modul.

Beweis. Siehe Proposition 2.4.1 in [Mishra93]. ■

Satz 13.1.9
Zu endlichem $G \subseteq R = S[x_1, x_2, \ldots, x_n]$ mit $I = (G)$ und S Noether'scher Ring sind folgende Aussagen äquivalent:

1. *G ist Gröbner-Basis, d.h.* $\mathrm{Head}(I) = \mathrm{Head}(G)$
2. *Für jede Basis $h_1, h_2, \ldots, h_m$ des S-Moduls $[\mathrm{SP}(G)]_S$ gilt: $h_i \xrightarrow[*]{G} 0$ für $i = 1, 2, \ldots, m$.*

Beweis. Wir zeigen beide Richtungen:

„⇒“ Nach Satz 12.1.12 ist jedes $f \in I$ von der Form

$$f = \sum_i f_i g_i \qquad f_i \in R,\ g_i \in G \text{ mit } \mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(f) \tag{13.1}$$

Insbesondere gilt dies für jedes Polynom $h \in \mathrm{SP}(G)$ und alle $h \in [\mathrm{SP}(G)]_S$, denn aus

$$h = \sum_j s_j h_j \qquad s_j \in S,\ h_j \in \mathrm{SP}(G)$$

folgt durch Einsetzen der h_j die Form (13.1). Aus Satz 13.1.4 erhalten wir die Behauptung.

„⇐“ Wegen $h_i \xrightarrow[*]{G} 0$ ist h_i von der Form (13.1). Dann ist auch jedes $f \in [\mathrm{SP}(G)]_S$ von der Form (13.1). Nach Satz 12.2.3 ist G eine Gröbner-Basis.

■

13.2 Reduktionsalgorithmen

Wir formulieren einen Algorithmus zur Reduktion von Polynomen. Wir definieren die vorausgesetzen Eigenschaften:

Definition 13.2.1 (Computable Ring/Körper)
Ein Ring S heißt computable, wenn es eine injektive Abbildung $\imath : S \to \{0,1\}^*$ gibt, unter der die Operationen $+, \cdot$ und $s \mapsto -s$ berechenbar sind. Ein Körper K heißt computable, wenn K als Ring computable ist und die Operation $r \mapsto r^{-1}$ auf $K \setminus \{0\}$ berechenbar ist.

Definition 13.2.2 (Detachable (abtrennbarer) Ring)
Ein Ring S heißt detachable (abtrennbar), wenn zu gegebenen Koeffizienten $s, s_1, s_2, \ldots, s_k \in S$ ein Algorithmus zum Entscheiden (ob eine Lösung existiert) und gegebenen Lösen der lineare Gleichung

$$\sum_{i=1}^{k} x_i s_i = s$$

in $x_1, x_2, \ldots, x_k \in S$ existiert.

Definition 13.2.3 (Syzygy-solvable Ring)
Ein Ring S heißt Syzygy-solvable, wenn es einen Algorithmus gibt, der zu $s, s_1, s_2, \ldots, s_k \in S$ eine Basis des S-Moduls $\ker \phi \subseteq S^k$

$$\phi : (x_1, x_2, \ldots, x_k) \mapsto \sum_{i=1}^{k} x_i s_i$$

berechnet.

Definition 13.2.4 (Strongly computable Ring)
Ein Ring S heißt strongly computable, falls S Noether'sch, computable, detachable und Syzygy-solvable ist.

Für einen Noether'schen Ring S folgt, daß ein S-Modul eine endliche Basis hat. Beispiele für strongly computable Ringe sind:

- $\mathbb{Z}$, $\mathbb{Q}$
- Endliche Körper GF $\left[p^k\right]$
- Mit S auch $R := S[x_1, x_2, \ldots, x_n]$

13.2.1 Head-Reduktion

Sei S ein strongly computable Ring, $G \subseteq R$ eine endliche Gröbner-Basis des Ideals (G) und allgemein $f \in R = S[x_1, x_2, \ldots, x_n]$. Die Korrektheit des Reduktionsschrittes (Algorithmus 13.2.1) folgt aus:

Lemma 13.2.5
Sei $f \in R \setminus \{0\}$ und $\emptyset \neq G \subseteq R$. Setze:

$$G_f := \{g \in G \,:\, \mathrm{Hterm}(g) \textit{ teilt } \mathrm{Hterm}(f)\}$$

Dann gilt:

$$\mathrm{Hmono}(f) \in \mathrm{Head}(G) \quad \Longleftrightarrow \quad \mathrm{Hcoef}(f) \in (\{\mathrm{Hcoef}(g) \,:\, g \in G_f\}) \subseteq S$$

Beweis. Wir zeigen beide Richtungen:

„⇒“ Sei

$$\mathrm{Hmono}(f) = \sum_{g_i \in G} a_i \cdot p_i \cdot \mathrm{Hmono}(g_i)$$

mit $a_i \in S$ und $p_i \in \mathrm{PP}(x_1, x_2, \ldots, x_n)$. Dann ist:

$$\mathrm{Hcoef}(f) = \sum_{\substack{g_i \text{ mit:} \\ \mathrm{Hterm}(f) = p_i \cdot \mathrm{Hterm}(g_i)}} a_i \cdot \mathrm{Hcoef}(g_i)$$

Wegen $\{g_i \,:\, \mathrm{Hterm}(f) = p_i \cdot \mathrm{Hterm}(g_i)\} \subseteq G_f$ folgt die Behauptung.

„⇐“ Sei

$$\mathrm{Hcoef}(f) = \sum_{g_i \in G_f} a_i \cdot \mathrm{Hcoef}(g_i)$$

mit $a_i \in S$. Wir erhalten durch Multiplikation mit $\mathrm{Hterm}(f)$:

$$\mathrm{Hmono}(f) = \underbrace{\sum_{g_i \in G_f} \left(a_i \cdot \frac{\mathrm{Hterm}(f)}{\mathrm{Hterm}(g_i)} \right) \cdot \mathrm{Hmono}(g_i)}_{\in \mathrm{Head}(G)} \qquad \text{mit } a_i \in S$$

■

Algorithmus 13.2.2 gibt mit Hilfe des Unterprogrammes 13.2.1 zu gegebenem f und G eine (aber nicht alle) Normalform aus.

Algorithmus 13.2.1 Reduktionsschritt (f, G)

EINGABE: $f \in R$ und $G \subseteq R$

1. IF $f \neq 0$ AND $\mathrm{Hcoef}(f) \in (\{\mathrm{Hcoef}(g) \,:\, g \in G_f\})$ THEN

1.1. Löse: $\mathrm{Hcoef}(f) = \sum_{g_i \in G_f} a_i \cdot \mathrm{Hcoef}(g_i)$ mit $a_i \in S$

1.2. $f := f - \sum_{g_i \in G_f} a_i \cdot \frac{\mathrm{Hterm}(f)}{\mathrm{Hterm}(g_i)} \cdot g_i$

END if

AUSGABE: f

Algorithmus 13.2.2 Reduktion (f, G)

EINGABE: $f \in R$ und $G \subseteq R$ mit $G \neq \emptyset$

1. $h := f$

2. WHILE $h \neq 0$ AND $\mathrm{Hcoef}(h) \in (\{\mathrm{Hcoef}(g) \,:\, g \in G_f\})$ DO

$h :=$ Reduktionsschritt(h, G) /∗ Algorithmus 13.2.1 ∗/

END while

AUSGABE: Eine Normalform $h \in \mathrm{NF}_G(f)$

13.2.2 Gröbner-Basis-Algorithmus

Algorithmus 13.2.3 liefert ein Erzeugendensystem zu $[\mathrm{SP}(G)]_S$. Falls der Koeffizientenring S ein Körper oder $\mathbb{Z}$ ist, genügt es, nur die F mit $|F| = 2$ zu betrachten (vergleiche Algorithmus 13.2.5).

Betrachten wir Algorithmus 13.2.4, der zu $h_1, h_2, \ldots, h_k \in R$ eine Gröbner-Basis des Ideals $I := (h_1, h_2, \ldots, h_k)$ bestimmt:

- Der Algorithmus terminiert. Das Ideal $\mathrm{Head}(G) \subseteq I$ wächst in jeder Iteration echt. Sei $f \in S(G)$. Dann gilt:

$$\mathrm{Hmono}(f) \notin \mathrm{Head}(G) \quad \Longleftrightarrow \quad \mathrm{Head}(G) \subsetneq \mathrm{Head}\,(G \cup \{f\})$$

  (weil f nicht reduzierbar modulo G ist). Da R Noether'sch ist, bricht die aufsteigende Folge der Ideale $\mathrm{Head}(G) \subseteq R$ ab.

- Korrektheit: Bei Abbruch gilt für eine Basis $b_1, b_2, \ldots, b_{\bar{k}}$ von $[\mathrm{SP}(G)]_S$:

$$G \subseteq I, \quad (G) = (I) \quad \text{und} \quad b_i \xrightarrow[*]{G} 0$$

  Nach Satz 13.1.9 ist G Gröbner-Basis von I, d.h. $\mathrm{Head}(I) = \mathrm{Head}(G)$.

Zwar liefert Algorithmus zu $h_1, h_2, \ldots, h_k \in R$ eine Gröbner-Basis des Ideals $I := (h_1, h_2, \ldots, h_k)$, allerdings ist das Verfahren im allgemeinen nicht effizient (vergleiche Gradschranken in Kapitel 13.4 auf Seite 162).

Falls der Koeffizientenring ein Körper oder $\mathbb{Z}$ ist, haben wir in Kapitel 12.2 gesehen, daß die Menge der S-Polynome $\mathrm{SP}_K(G)$ bzw. $\mathrm{SP}_{\mathbb{Z}}(G)$ nur aus Polynomen gebildet werden (Satz 12.2.7 und

Algorithmus 13.2.3 Konstruktion eines Erzeugendensystems zu $[\mathrm{SP}(G)]_S$

EINGABE: $G = \{g_1, g_2, \ldots, g_k\} \subseteq R$

1. FOR all $F \subseteq G$ DO

1.1. $M_F := \left\{ (s_1, s_2, \ldots, s_k) \in S^k \;\middle|\; \sum_{g_i \in F} s_i \cdot \mathrm{Hcoef}(g_i) = 0 \right\}$

/* M_F ist ein S-Modul vom Rang $k_F \geq |F| - 1$. */

1.2. Konstruiere Basis $B_F := \left\{ s^{(1)}, \ldots, s^{(k_F)} \right\}$ mit $s^{(i)} = \left(s_1^{(i)}, s_2^{(i)}, \ldots, s_k^{(i)} \right) \in S^k$.

/* Da nach Voraussetzung der Ring Syzygy-solvable ist, existiert ein solcher Algorithmus zur Konstruktion einer Basis. */

1.3. $m_F = \mathrm{kgV}\left(\{\mathrm{Hterm}(g_i) \;:\; g_i \in F\}\right)$

END for

2. Gib folgendes Erzeugendensystems zu $[\mathrm{SP}(G)]_S$ aus:

$$\left\{ \sum_{g_i \in F} s_i^{(j)} \cdot g_i \cdot \frac{m_F}{\mathrm{Hterm}(g_i)} \;\middle|\; F \subseteq G,\ j = 1, 2, \ldots, k_F \right\}$$

Algorithmus 13.2.4 Konstruktion einer Gröbner-Basis

EINGABE: $h_1, h_2, \ldots, h_k \in R$ mit $I = (h_1, h_2, \ldots, h_k)$

1. $G := \{h_1, h_2, \ldots, h_k\}$

2. Konstruiere Basis $b_1, b_2, \ldots, b_{\bar{k}}$ des S-Moduls $[\mathrm{SP}(G)]_S \subseteq I$

3. FOR $i = 1, 2, \ldots, \bar{k}$ DO Konstruiere Normalform $f_i \in \mathrm{NF}_G(b_i)$

4. $S(G) := \left\{ f_i \;:\; f_i \neq 0,\ i = 1, 2, \ldots, \bar{k} \right\}$

5. IF $S(G) \neq \emptyset$ THEN $G := G \cup S(G)$; GOTO 2

AUSGABE: Gröbner-Basis G von $I := (h_1, h_2, \ldots, h_k)$

12.2.8). Wir können daher den Algorithmus zur Konstruktion einer Gröbner-Basis vereinfachen (siehe [BeWe93, Kapitel 5.3]).

13.2.3 Einfache Anwendungen

In diesem Abschnitt lernen wir zwei Anwendungen der Reduktions-Algorithmen kennen. Betrachten wir zunächst die folgende Aufgabe (Membership-Problem):

- Gegeben sind $f, g_1, g_2, \ldots, g_m \in R$.
- Entscheide, ob $f \in (g_1, g_2, \ldots, g_m)$, d.h. ob $f = \sum_{i=1}^{m} h_i g_i$ mit $h_i \in R$ lösbar ist.

Algorithmus 13.2.5 Konstruktion einer Gröbner-Basis für Koeffizientenkörper oder $\mathbb{Z}$

EINGABE: $h_1, h_2, \ldots, h_k$ mit $I = (h_1, \ldots, h_k)$ Ideal in $K[x_1, \ldots, x_n]$ bzw. $\mathbb{Z}[x_1, \ldots, x_n]$

1. $G := \{h_1, h_2, \ldots, h_k\}$

2. $B = \{(g_1, g_2) \in G \times G \mid g_1 \neq g_2\}$

3. WHILE $B \neq \emptyset$ DO

3.1. Wähle $(g_1, g_2) \in B$

3.2. $B := B \setminus \{(g_1, g_2)\}$

3.3. $f := S(g_1, g_2)$ /* S-Polynom zu g_1, g_2 */

3.4. Konstruiere Normalform $f' \in \mathrm{NF}_G(f)$

3.5. IF $f' \neq 0$ THEN

$$B := B \cup \{(f', g) \mid g \in G\}$$
$$G := G \cup \{f'\}$$

END if

END while

AUSGABE: Gröbner-Basis G von $I := (h_1, h_2, \ldots, h_k)$

Wir berechnen eine Gröbner-Basis G zu $(g_1, g_2, \ldots, g_m)$. Bestimme ein $h \in \mathrm{NF}_G(f)$. Es gilt:

$$h = 0 \iff f \in (g_1, g_2, \ldots, g_m)$$

Wir beweisen beide Richtungen der Äquivalenz:

„$\Rightarrow$“ Klar.

„$\Leftarrow$“ Nach Satz 13.1.4 existiert eine Ableitungsfolge $f \xrightarrow[*]{G} 0$. Angenommen, die Ableitung ende mit h. Aus $f \in (g_1, g_2, \ldots, g_m)$ folgt:

$$\mathrm{NF}_G(f) \subseteq (g_1, g_2, \ldots, g_m)$$

Es ist $h <_A f$, auf h können wir erneut Satz 13.1.4 anwenden.

Betrachten wir eine zweite Anwendung:

- Gegeben sind $f_1, f_2, \ldots, f_{m'}, g_1, g_2, \ldots, g_m \in R$.
- Entscheide, ob $(f_1, f_2, \ldots, f_{m'}) = (g_1, g_2, \ldots, g_m)$.

Diese Aufgabe können wir mit der ersten Anwendung lösen, denn offenbar:

$$(f_1, f_2, \ldots, f_{m'}) \subseteq (g_1, g_2, \ldots, g_m) \iff [f_i \in (g_1, g_2, \ldots, g_m) \quad \text{für } i = 1, 2, \ldots, m']$$

Für weitere Anwendungen der Reduktions-Algorithmen verweisen wir auf [Mishra93, Kapitel 3.7] und [BeWe93, Kapitel 6].

13.3 Reduktionstheorie mit Eindeutigkeit

Für eine Reduktionstheorie mit Eindeutigkeit müssen wir stärkere Forderungen an die Gröbner-Basis stellen. Wir definieren Gröbner-Basen mit speziellen Eigenschaften:

Definition 13.3.1 (Selbst-reduzierte Gröbner-Basis)
Sei G Gröbner-Basis zu $I = (G)$. G heißt selbst-reduziert, falls:

$$\forall g \in G: \quad \mathrm{NF}_{G \setminus \{g\}}(g) = \{g\}$$

Also ist g nicht reduzierbar modulo $G \setminus \{g\}$.

Wir führen den Begriff minimaler Gröbner-Basen ein.

Definition 13.3.2 (Minimale Gröbner-Basis)
Eine Gröbner-Basis G heißt minimal, falls für jedes $g \in G$ gilt, daß $G \setminus \{g\}$ keine Gröbner-Basis von (G) ist.

Der folgende Satz zeigt die Beziehung zwischen minimaler und selbst-reduzierter Gröbner-Basis zu einem Ideal:

Satz 13.3.3
Sei G Gröbner-Basis zum Ideal $I = (G)$ mit $0 \notin G$. G ist genau dann minimal, wenn G selbst-reduziert ist.

Beweis. Wir zeigen beide Richtungen durch Widerspruchsbeweise:

„⇒“ Angenommen, G sei minimal und nicht selbst-reduziert. Dann gibt es $g \in G$, daß reduzierbar modulo $(G \setminus \{g\})$ ist. Also:

$$\mathrm{Hmono}(g) \in \mathrm{Head}\,(G \setminus \{g\})$$

Somit:

$$\mathrm{Head}(G) = \mathrm{Head}\,(G \setminus \{g\}) + \underbrace{(\mathrm{Hmono}(g))}_{\subset \mathrm{Head}(G \setminus \{g\})} = \mathrm{Head}\,(G \setminus \{g\})$$

Dies ist ein Widerspruch zur Minimalität von G, da $G \setminus \{g\}$ eine Gröbner-Basis von (G) ist.

„⇐“ Angenommen, G sei selbst-reduziert und nicht minimal. Dann gibt es ein $g \in G$, so daß $G \setminus \{g\}$ Gröbner-Basis von (G) ist. Also:

$$\mathrm{Head}(G) = \mathrm{Head}\,(G \setminus \{g\})$$

Wegen $\mathrm{Hmono}(g) \in \mathrm{Head}\,(G \setminus \{g\})$ ist g reduzierbar modulo $G \setminus \{g\}$ — Dies ist ein Widerspruch, da G nach Voraussetzung selbst-reduziert ist.

■

Bezeichnung 13.3.4 (Menge der Monome $M(f)$)
Zu $f = \sum_{i=1}^{n} a_i p_i \in R$ mit $a_i \in K \setminus \{0\}$ und $p_i \in PP(x_1, x_2, \ldots, x_n)$ sei

$$M(f) := \{a_1 p_1, a_2 p_2, \ldots, a_n p_n\}$$

die Menge der Monome von f.

Definition 13.3.5 (Vollständig reduzierbar* modulo G)
Zu $G \subseteq R$ ist $f \in R$ (vollständig) reduzierbar modulo G, wenn:*

$$M(f) \cap \mathrm{Head}(G) \neq \emptyset$$

Aus f reduzierbar modulo G folgt, daß f vollständig reduzierbar modulo G ist.

Definition 13.3.6 (Reduktionsschritt*)
Wir schreiben $f \xrightarrow{G} * r$ genau dann, wenn es $a \cdot p \in M(f)$ sowie $a_1, a_2, \ldots, a_m \in K \setminus \{0\}$ und $g_1, g_2, \ldots, g_m \in G$ gibt mit:

$$\begin{aligned} a \cdot p &= \sum_{i=1}^{m} a_i \cdot p_i \cdot \mathrm{Hmono}(g_i) \\ p &= p_i \cdot \mathrm{Hmono}(g_i) && i = 1, 2, \ldots, m \\ f &= \sum_{i=1}^{m} a_i \cdot p_i \cdot g_i + r \end{aligned}$$

Mit $\xrightarrow[*]{G} *$ bezeichnen wir den reflexiven und transitiven Abschluß von $\xrightarrow{G} *$.

Definition 13.3.7 (Normalformen $\mathrm{NF}^*_G(f)$)
Mit $\mathrm{NF}^*_G(f)$ bezeichnen wir die Menge der Normalformen zu f bezüglich $\xrightarrow{G} *$.

Satz 13.3.8
Sei K ein Körper. Für eine Gröbner-Basis $G \subseteq R := K[x_1, x_2, \ldots, x_n]$ gilt für alle $f \in R$: $|\mathrm{NF}^*_G(f)| = 1$.

Beweis. Angenommen, es gäbe $r, r' \in \mathrm{NF}^*_G(f)$ mit $r \neq r'$. Dann ist $r - r' \in (G)$ und $r - r'$ ist von der Form:

$$r - r' = \sum_{g_i \in G} f_i \cdot g_i \qquad \text{mit } \mathrm{Hterm}(f_i \cdot g_i) \leq_A \mathrm{Hterm}(r - r') \tag{13.2}$$

Sei $p := \mathrm{Hterm}(r - r')$ und $a \cdot p = \mathrm{Hmono}(r - r')$. O.B.d.A. sei p Term von r, d.h. $c \cdot p \in M(r)$. Wegen (13.2) gibt es ein $g_i \in G$ mit $\mathrm{Hterm}(g_i) \mid p$. Aus $c \cdot p \in \mathrm{Head}(G)$ folgt:

$$M(r) \cap \mathrm{Head}(G) \neq \emptyset$$

Dies ist ein Widerspruch, da r reduzierbar* ist. ■

Der Beweis gilt allgemein für beliebige Koeffizientenringe, falls alle $g \in G$ monisch sind, d.h. $\mathrm{Hcoef}(g) = 1$: In diesem Fall ist $c \cdot p$ im Ideal $\mathrm{Head}(G)$ darstellbar. Wir definieren:

Definition 13.3.9 (Monische Gröbner-Basis)
Eine Gröbner-Basis heißt monisch, falls für alle $g \in G$ gilt: $\mathrm{Hcoef}(g) = 1$.

Mit dieser Definition folgt aus Satz 13.3.8:

Korollar 13.3.10
Sei S ein beliebiger Ring und $G \subseteq R = S[x_1, x_2, \ldots, x_n]$ eine monische Gröbner-Basis. Dann gilt für alle $f \in R$: $|\mathrm{NF}^*_G(f)| = 1$.

Korollar 13.3.11
Sei S ein beliebiger Ring und $G \subseteq R$ eine monische Gröbner-Basis. Dann gibt es in jeder Restklasse $f + (G)$ eine Normalform bezüglich $\xrightarrow{G} *$, d.h. $\mathrm{NF}^*(f) = \mathrm{NF}^*(f + g)$ mit $g \in (G)$.

Nach Korollar 13.3.10 existiert genau ein $r \in \mathrm{NF}^*(f)$, das wir als Standardrest von f bezeichnen:

Definition 13.3.12 (Standardrest modulo G)
Wir nennen $r \in \mathrm{NF}^*(f)$ Standardrest von f modulo G.

Wir definieren reduzierte Gröbner-Basen:

Definition 13.3.13 (Reduzierte Gröbner-Basis)
Eine Gröbner-Basis $G \subseteq R$ ist reduziert, falls G monisch und selbst-reduziert ist, d.h.:*

$$\forall g \in G: \quad \mathrm{NF}^*_{G \setminus \{g\}}(g) = \{g\}$$

Es gilt:

Satz 13.3.14
Sei K ein Körper. Jedes Ideal $I \subseteq R := K[x_1, x_2, \ldots, x_n]$ hat eine eindeutig bestimmte, reduzierte Gröbner-Basis.

Algorithmus 13.3.1 Reduzierte Gröbner-Basis

EINGABE: Gröbner-Basis $G = \{g_1, g_2, \ldots, g_m\}$ von I

1. $G := \left\{ \frac{g_1}{\mathrm{Hcoef}(g_1)}, \frac{g_2}{\mathrm{Hcoef}(g_2)}, \ldots, \frac{g_m}{\mathrm{Hcoef}(g_m)} \right\}$ /* Normiere $g_1, g_2, \ldots, g_m$ */

2. FOR $i = 1, 2, \ldots, m$ DO

2.1. Berechne $g_i' \in \mathrm{NF}^*_{G \setminus \{g_i\}}(g_i)$

2.2. IF $g_i' \neq 0$ THEN $g_i := g_i'$ ELSE $G := G \setminus \{g_i\}$

END for i

AUSGABE: Reduzierte Gröbner-Basis G von I

Beweis. Wir weisen die Korrektheit von Algorithmus 13.3.1 und die Eindeutigkeit der Reduktion nach:

- Die Ausgabe des Algorithmus' ist eine reduzierte Gröbner-Basis G von $(g_1, g_2, \ldots, g_m)$, denn:
  1. $\mathrm{Head}(G)$ bleibt bei jeder Iteration erhalten. Fallunterscheidung:
     - Falls $\mathrm{Hmono}(g_i) \neq \mathrm{Hmono}(g_i')$ ist, existiert ein $g_j \in G \setminus \{g_i\}$ mit:
     $$\mathrm{Hterm}(g_j) \mid \mathrm{Hterm}(g_i)$$
     Also ist g_i reduzierbar modulo g_j und $\mathrm{Head}(G) = \mathrm{Head}(G \setminus \{g_i\})$.
     - Falls $\mathrm{Hmono}(g_i) = \mathrm{Hmono}(g_i')$ ist, bleibt $\mathrm{Head}(G)$ erhalten.
  2. Die Polynome in G sind offenbar nach jeder Iteration monisch.

3. $\forall g \in G : g \in \mathrm{NF}^*_{G\setminus\{g\}}(g)$. Zum Zeitpunkt der Berechnung von g'_i gilt:

$$g'_i \in \mathrm{NF}^*_{G\setminus\{g'_i\}}(g'_i)$$

Wird später ein g_j mit $j > i$ durch ein g'_j ersetzt, dann bleibt $\mathrm{Hmono}(g_j)$ erhalten. Also bleibt g'_i irreduzibel* modulo $G \setminus \{g_i\}$. Falls g_j entfernt wird, bleibt g'_i irreduzibel modulo $G \setminus \{g_i\}$.

- Eindeutigkeit: Angenommen, G und H seien reduzierte Basis von $I \subseteq R$. Wähle

$$g \in G \triangle H = (G \cup H) \setminus (G \cap H) \qquad \text{mit } \mathrm{Hterm}(g) = \min\{\mathrm{Hterm}(f) \,:\, f \in G \triangle H\}$$

O.B.d.A. sei $g \in G$. Weil H Gröbner-Basis ist, gilt $\mathrm{Hmono}(g) \in \mathrm{Head}(H)$. Da der Koeffizientenring ein Körper ist, gibt es ein $h \in H$ mit

$$\mathrm{Hterm}(h) \mid \mathrm{Hterm}(g)$$

Wegen $h \neq g$ und $\mathrm{NF}^*_{G\setminus\{g\}}(g) = \{g\}$ gilt $h \in H \setminus G$ (sonst könnten wir g reduzieren). Fallunterscheidung:

1. $\mathrm{Hterm}(h) <_A \mathrm{Hterm}(g)$: Dies ist ein Widerspruch zur Minimalität von g, da $h \in G \triangle H$.
2. $\mathrm{Hterm}(h) = \mathrm{Hterm}(g)$: Sei $f = g - h$. Da g und h monisch sind, gilt:

$$\mathrm{Hterm}(f) <_A \mathrm{Hterm}(g) \quad \text{und} \quad \mathrm{Hterm}(f) <_A \mathrm{Hterm}(h)$$

O.B.d.A. sei $\mathrm{Hterm}(f)$ ein Term von g. Wegen $f \in I$:

$$\exists p \in G \setminus \{g\} : \quad \mathrm{Hterm}(p) \mid \mathrm{Hterm}(f)$$

Also ist $M(g) \cap \mathrm{Head}(G \setminus \{g\}) \neq \emptyset$ und g reduzierbar modulo $G \setminus \{g\}$ – Widerspruch.

■

Die reduzierte Gröbner-Basis ist eindeutig bestimmt:

Satz 13.3.15
Sei S ein beliebiger Ring. Jedes Ideal $I \subseteq R := S[x_1, x_2, \ldots, x_n]$ besitzt höchstens eine reduzierte Gröbner-Basis.

Beweis. Die Polynome der reduzierten Gröbner-Basis sind monisch. Die Eindeutigkeit folgt aus dem Beweis zu Satz 13.3.14. ■

13.4 Gradschranken

In diesem Abschnitt fassen wir Resultate für die Grade der Polynome der Gröbner-Basen zusammen. Sei S ein Körper und $R := S[x_1, x_2, \ldots, x_n]$. Definiere:

1. $D(n, d)$ sei die kleinste Zahl $D \in \mathbb{N}$ mit:

   Für jede zulässige Ordnung und jedes Ideal $I \subseteq R$, daß von Polynomen vom Grad maximal d erzeugt wird, existiert eine Gröbner-Basis, deren Elemente Grad maximal D haben.

2. $I(n, d)$ sei die kleinste Zahl $I \in \mathbb{N}$ mit:

   Für alle zulässigen Ordnungen und alle $f \in (g_1, g_2, \ldots, g_n)$ mit $\mathrm{grad}(f) \leq d$ existieren $h_1, h_2, \ldots, h_m \in R$ mit: $\mathrm{grad}(h_i) \leq I$ und $\sum_i h_i g_i = f$.

3. $S(n,d)$ sei kleinste Zahl $S \in \mathbb{N}$:

   Der Modul $\{h_1, h_2, \ldots, h_m \in R^n : \sum_i h_i g_i = 0\}$ hat eine Basis bestehend aus Polynomen, deren Grad durch S nach oben beschränkt ist.

Mishra listet in [Mishra93] die folgenden Gradschranken auf:

- Obere Schranken:

M. Giusti (1984)	:	$S(n,d)$	$\leq$	$D(n,d)$	
D. Lazard (1992)	:	$S(n,d)$	$\leq$	$I(n,d)$	
		$I(n,d)$	$\leq$	$S(n,d)^{\mathcal{O}(n)}$	
G. Hermann (1926)	:	$S(n,d)$	$\leq$	$d + 2(md)^{2^{n-1}}$	mit $m := \lvert G \rvert$
D. Lazard (1992)	:	$S(n,d)$	$\leq$	$d^{2^{(\log 3/\log 4)\cdot n}}$	
		$I(n,d)$	$\leq$	$d^{2^{(\log 3/\log 4)\cdot n \cdot o(n)}}$	
T.W. Dubé (1989)	:	$D(n,d)$	$\leq$	d^{2^n}	

- Untere Schranken:

E.W. Mayr und A.R. Meyer (1982)	:	$I(n,d), S(n,d)$	$\geq$	$d^{2^{n'}}$	mit $n' \approx \frac{n}{10}$
C.K. Yap (1991)	:	$I(n,d), S(n,d)$	$\geq$	$2^{2^{\bar{n}}}$	mit $\bar{n} \approx \frac{n}{2}$

D. Brownawell (1987) zeigt:

$$1 \in (g_1, g_2, \ldots, g_m) \iff \exists h_1, h_2, \ldots, h_m \in R : \operatorname{grad}(h_i) \leq d^{\mathcal{O}(n)}, \quad 1 = \sum_{i=1}^{m} g_i h_i \tag{13.3}$$

M. Guisti und J. Heintz zeigen, es gibt Schaltkreise mit den Operationen „+“ und „·“ mit

- Größe $m^{\mathcal{O}(1)} d^{\mathcal{O}(n)}$ und
- Tiefe $n^{\mathcal{O}(1)} \left(\log_2(md)\right)^{\mathcal{O}(1)}$,

welche zu gegebenen $g_1, g_2, \ldots, g_m \in R$ die Polynome $h_1, h_2, \ldots, h_m \in R$ aus (13.3) berechnen.

13.5 Lösen von polynomialen Gleichungssystemen

Wir lernen eine weitere Anwendung der Reduktionsalgorithmen für Gröbner-Basen kennen: Wir werden polynomiale Gleichungssysteme lösen.

13.5.1 Dreicksform

Das Ziel ist der Entwurf eines Algorithmus' zur Bestimmung der Lösungsmenge eines beliebigen Systems multivariabler polynomialer Gleichungen:

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ f_r(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

Zu berechnen ist die Menge aller Punkte $x_1, x_2, \ldots, x_n$, in denen alle Polynome $f_1, f_2, \ldots, f_r$ den Wert 0 annehmen (verschwinden):

$$\{(x_1, x_2, \ldots, x_n) \mid f_i(x_1, x_2, \ldots, x_n) = 0 \text{ für } i = 1, 2, \ldots, r\}$$

Für Polynome vom Grad 1 gibt es ein bekanntes Verfahren: Ein lineares Gleichungssystem löst man mit dem Gauß-Eliminationsverfahren. Im allgemeinen Fall tritt der Gröbner-Basis-Algorithmus aus Kapitel 13.2.2 an die Stelle der Gauß-Elimination. Anstatt des Pivot-Schrittes werden S-Polynome und Reduktionsprozesse verwendet.

Der Gröbner-Basis-Algorithmus liefert eine „Dreiecksmenge“ von Polynomen. Es werden Klassen von polynomialen Systemen bestimmt, so daß die letzte Klasse nur x_n enthält, die vorletzte x_{n-1}, x_n usw. Bei gegebener Dreiecksmenge kann man die Lösungsmenge leicht bestimmen.

Definition 13.5.1 (Dreiecksform)
Sei $G \subseteq S[x_1, x_2, \ldots, x_n]$ eine endliche Menge von Polynomen über dem Ring S. Wir betrachten folgende Partitionierung von G in $n+1$ Klassen:

$$\begin{aligned} G_0 &= (G \cap S[x_1, x_2, \ldots, x_n]) \setminus S[x_2, x_3, \ldots, x_n] \\ G_1 &= (G \cap S[x_2, x_3, \ldots, x_n]) \setminus S[x_3, x_4, \ldots, x_n] \\ &\vdots \\ G_{n-1} &= (G \cap S[x_n]) \setminus S \\ G_n &= G \cap S \end{aligned}$$

G_i ist die Menge aller Polynome in G, die $x_{i+1}, x_{i+2}, \ldots, x_n$ enthalten, aber kein x_j mit $j \leq i$. Speziell ist G_n die Menge aller konstanten Polynome in G. Die Folge $G_0, G_1, \ldots, G_n$ von Polynom-Mengen heißt Dreiecksform von G.

Definition 13.5.2 (Strikte Dreiecksform)
Sei K Körper und $G \subseteq K[x_1, x_2, \ldots, x_n]$ endliche Teilmenge. Eine Dreiecksform $G_0, G_1, \ldots, G_n$ von G heißt strikte Dreiecksform, falls:

1. $(G_n) = (G \cap K) = \{0\}$
2. *Für $i = 0, 1, \ldots, n-1$ existiert ein $g_i \in G_i$, das ein Monom der Form ax_{i+1}^d mit $a \in K^*$ und $d \in \mathbb{N}$ enthält.*

Sei eine Menge G von Polynomen gegeben. Falls die Erzeugenden des Ideals (G) auf strikte Dreiecksform gebracht werden können, gibt es eine endliche, nicht-leere Menge von gemeinsamen Nullstellen der Polynome in G. Jedes Element aus dem Ideal (G) muß bei den gemeinsamen Nullstellen von G verschwinden. Daher hat jede Basis von G die gleichen, gemeinsamen Nullstellen. Intuitiv: Wegen der Bedingung

$$(G_n) = (G \cap K) = \{0\}$$

wissen wir, daß die Menge nichtleer sein kann, da sonst in $(G \cap K)$ eine Kostante ungleich 0 läge.

Beispiel 13.5.3 (Strikte Dreiecksform)
Betrachte die polynomialen Gleichungssysteme $G', G'', G''' \subseteq \mathbb{C}[x_1, x_2]$ in Dreicksform:

$$\begin{aligned} G' &= \{x_1x_2 - x_2, x_2^2 - 1\} \\ G'' &= \{x_1x_2 - x_2, x_2^2\} \\ G''' &= \{x_1x_2 - x_2\} \end{aligned}$$

Welches der Systeme G', G'', G''' hat nur endlich viele gemeinsame Nullstellen? Aus

$$G'_0 = G''_0 = G'''_0 = \{x_1x_2 - x_2\}$$

erhalten wir diese Information nicht.

- Im Fall von G' erhalten wir aus $x_2^2 - 1 = 0$, daß $x_2 = \pm 1$ sein muß. Durch Einsetzen in $x_1x_2 - x_2 = 0$ erhalten wir, daß G' genau die zwei gemeinsamen Nullstellen $(1, 1)$ und $(1, -1)$ hat. Tatsächlich hat G' eine Basis in strikter Dreiecksform:

$$G' = \left(x_1 - 1, x_2^2 - 1\right)$$

- Im Fall von G'' erhalten wir aus $x_2^2 = 0$, daß $x_2 = 0$ sein muß. Eingesetzt in $x_1x_2 - x_2 = 0$ erhalten wir, daß $(\xi, 0)$ mit $\xi \in \mathbb{C}$ die gemeinsamen Nullstellen von G'' sind. G'' hat keine Basis in strikter Dreiecksform.

- Im Fall von G''' erhalten wir $x_2(x_1 - 1) = 0$, daß $(1, \zeta)$ mit $\zeta \in \mathbb{C}$ und $(\xi, 0)$ mit $\xi \in \mathbb{C}$ die gemeinsamen Nullstellen von G''' sind. In der Tat hat G''' keine Basis in strikter Dreiecksform.

$\diamond$

Definition 13.5.4 (Eliminationsideal)
Sei $I \subseteq S[x_1, x_2, \ldots, x_n]$ ein Ideal im Polynomring $S[x_1, x_2, \ldots, x_n]$. Wir nennen

$$I_i := I \cap S[x_{i+1}, x_{i+2}, \ldots, x_n]$$

das i-tes Eliminationsideal von I für $i = 0, 1, \ldots, n-1$.

Zu einem gegebenen Gleichungssystem

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ f_r(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

möchten wir eine Basis $G = \{g_1, g_2, \ldots, g_s\}$ des Ideals $I = (f_1, f_2, \ldots, f_r)$ so konstruieren, daß für die Dreiecksform von G gilt:

$$\bigcup_{j=i}^{n} G_j \text{ ist eine Basis des } i\text{-ten Eliminationsideals } I_i$$

Wegen $I_i \subseteq I_{i-1}$ folgt, daß die Lösungsmenge des Systems der Polynome $\bigcup_{j=i}^{n} G_j$ die Lösungsmenge des Systems $\bigcup_{j=i-1}^{n} G_j$ enthält. Die berechnete Basis G wird eine Gröbner-Basis von I sein, und auch die Mengen $\bigcup_{j=i}^{n} G_j$ werden Gröbner-Basen von I_j bezüglicher der gleichen, zulässigen lexikographischen Ordnung sein.

Definition 13.5.5 (Verallgemeinerte lexikographische Ordnung $>_L$)
Wir betrachten den Ring $S[X, Y] = S[x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n]$. Seien $>_X$ und $>_Y$ zulässige Ordnungen auf $PP(X)$ bzw. $PP(Y)$. Eine verallgemeinerte lexikographische Ordnung $>_L$ auf $PP(X, Y)$ wird definiert durch:

$$pq >_L p'q' \qquad \text{falls } p >_X p' \text{ oder } p =_X p' \text{ und } q >_Y q'$$

Dabei ist $p, p' \in PP(X)$ und $q, q' \in PP(Y)$. Diese Ordnung ist zulässig.

Satz 13.5.6
Sei G Gröbner-Basis von I bezüglich $>_L$ in $S[X,Y]$. Dann ist $G \cap S[Y]$ Gröbner-Basis von $I \cap S[Y]$ bezüglich $>_Y$ in $S[Y]$.

Beweis. Nach Definition von $>_L$ gilt für $f \in S[X,Y]$:

$$\text{Hmono}_L(f) \in S[Y] \quad \Longleftrightarrow \quad f \in S[Y] \tag{13.4}$$

Ein Polynom, dessen Headterm in $\text{PP}(Y)$ ist, enthält kein Potenzprodukt mit $x_i \in X$. Damit:

$$\begin{aligned}
\text{Head}_Y(G \cap S[Y]) &= \text{Head}_L(G \cap S[Y]) && \text{(wegen (13.4))} \\
&= \text{Head}_L(G) \cap \text{Head}(S[Y]) \\
&= \text{Head}_L(I) \cap \text{Head}(S[Y]) && \text{(da } G \text{ Gröbner-Basis von } I\text{)} \\
&= \text{Head}_L(I \cap S[Y]) \\
&= \text{Head}_Y(I \cap S[Y]) && \text{(wegen (13.4))}
\end{aligned}$$

Wegen $G \subseteq I$ gilt auch:

$$(G \cap S[Y]) \subseteq (I \cap S[Y])$$

$G \cap S[Y]$ ist ebenfalls eine Gröbner-Basis für $I \cap S[Y]$ bezüglich $>_Y$ in $S[Y]$. ■

Korollar 13.5.7
Sei G Gröbner-Basis von I bezüglich $>_{\text{lex}}$ in $S[x_1, x_2, \ldots, x_n]$ (mit $x_1 >_{\text{lex}} x_2 >_{\text{lex}} > \cdots >_{\text{lex}} x_n$). Dann gilt für $i = 0, 1, \ldots, n$:

1. *$G \cap S[x_{i+1}, x_{i+2}, \ldots, x_n]$ ist Gröbner-Basis von $I \cap S[x_{i+1}, x_{i+2}, \ldots, x_n]$ bezüglich $>_{\text{lex}}$ in $S[x_1, x_2, \ldots, x_n]$.*
2. *Äquivalent: $\bigcup_{j=i}^{n} G_j$ ist eine Gröbner-Basis für das i-te Eliminationsideal.*

13.5.2 Algebraische Geometrie

Sei L ein algebraisch abgeschlossener Körper, d.h. für alle $f \in L[x] \setminus L$ existiert ein $x \in L$ mit $f(x) = 0$. Für $d = \text{grad}(f)$ gibt es $x_1, x_2, \ldots, x_d \in L$ mit:

$$f(x) = \prod_{i=1}^{d} (x - x_i)$$

Wir bezeichnen mit L^n den n-dimensionalen Vektorraum über L.

Definition 13.5.8 (Varietät, Nullstellenmenge)
Sei $\mathcal{V} : \text{Pot}(L[x_1, x_2, \ldots, x_n]) \to \text{Pot}(L^n)$ die Abbildung mit:

$$F \mapsto \{p \in L^n \ : \ \forall f \in F : f(p) = 0\}$$

Dann heißt $\mathcal{V}(F)$ die Varietät (Nullstellenmenge) zu F.

Definition 13.5.9 (Ideal)
Sei $\mathcal{I} : \text{Pot}(L^n) \to \text{Pot}(L[x_1, x_2, \ldots, x_n])$ die Abbildung mit:

$$X \mapsto \{f \in L[x_1, x_2, \ldots, x_n] \ : \ X \subseteq \mathcal{V}(F)\} = \{f \in L[x_1, x_2, \ldots, x_n] \ : \ \forall p \in X : f(p) = 0\}$$

Dann heißt $\mathcal{I}(X)$ Ideal zu $X \subseteq L^n$.

Beachte: $\mathcal{I}(X) \subseteq L[x_1, x_2, \ldots, x_n]$ ist ein Ideal, da:

1. Aus $f, g \in \mathcal{I}(X)$ folgt $f + g \in \mathcal{I}(X)$.
2. Aus $f \in \mathcal{I}(X)$ folgt für alle $g \in L[x_1, x_2, \ldots, x_n]$, daß $f \cdot g \in \mathcal{I}(X)$.

Definition 13.5.10 (Algebraisch)
$X \subseteq L[x_1, x_2, \ldots, x_n]$ heißt algebraisch, wenn $X = \mathcal{V}(F)$ für ein $F \subseteq L[x_1, x_2, \ldots, x_n]$.

Es gilt:

Satz 13.5.11
Sei $F \subseteq L[x_1, x_2, \ldots, x_n]$ ein endliches, erzeugendes System von $I = (F)$. Dann gilt $\mathcal{V}(F) = \mathcal{V}(I)$.

Beweis. Wir zeigen beide Inklusionen:

„$\supseteq$“ Allgemein: Aus $F \subseteq F'$ folgt $\mathcal{V}(F) \supseteq \mathcal{V}(F')$. Da $F \subseteq I$, ist:

$$\mathcal{V}(F) \supseteq \mathcal{V}(I)$$

„$\subseteq$“ Sei $F = \{f_1, f_2, \ldots, f_r\}$. Aus $p \in \mathcal{V}(F)$ folgt:

$$f_1(p) = f_2(p) = \cdots = f_r(p) = 0$$

Dann gilt für alle $f = \sum_{i=1}^{r} h_i f_i$ mit $h_1, h_2, \ldots, h_r \in L[x_1, x_2, \ldots, x_n]$, daß $f(p) = 0$ ist. Weil jedes $f \in I$ eine solche Darstellung hat, folgt: Für alle $f \in I$ gilt $f(p) = 0$. Insgesamt erhalten wir $p \in \mathcal{V}(I)$.

■

Ein bekannter Satz aus der Algebra ist:

Satz 13.5.12 (Hilbert'scher Nullstellensatz)
Sei L ein algebraisch abgeschlossener Körper. $I \subseteq L[x_1, x_2, \ldots, x_n]$ Ideal. Dann gilt:

$$I = L[x_1, x_2, \ldots, x_n] = (1) \qquad \Longleftrightarrow \qquad \mathcal{V}(I) \neq \emptyset$$

Beweis. Siehe Beweis zu Theorem 4.3.5 in [Mishra93]. ■

Definition 13.5.13 (Radikal eines Ideals)
Sei $I \subseteq L[x_1, x_2, \ldots, x_n]$ ein Ideal. Definiere das Radikal zum Ideal I:

$$\sqrt{I} := \{f \in L[x_1, x_2, \ldots, x_n] \mid \exists q \in \mathbb{N} : f^q \in I\}$$

Definition 13.5.14 (Lösbares System von Polynomen)
Sei $F \subseteq L[x_1, x_2, \ldots, x_n]$ ein System von Polynomen. F heißt lösbar, wenn $\mathcal{V}(F) \neq \emptyset$, und endlich lösbar, wenn $\mathcal{V}(F) \neq \emptyset$, $\mathcal{V}(F)$ endlich sind.

Satz 13.5.15 (Allgemeiner Hilbert'scher Nullstellensatz)
Sei L ein algebraisch abgeschlossener Körper und $I \subseteq L[x_1, x_2, \ldots, x_n]$ Ideal. Dann gilt ($\circ$ ist die Konkatenation der Funktionen):

$$\mathcal{I} \circ \mathcal{V}(I) = \sqrt{I}$$

Beweis. Da es sich bei L um einen Körper handelt, ist jedes Ideal in L endlich erzeugt (siehe Hilbert'schen Basissatz 12.1.15 auf Seite 144). Sei $I = (f_1, f_2, \ldots, f_r)$.

„$\supseteq$“ Sei $f \in \sqrt{I}$. Es gibt nach Definition ein $q \in \mathbb{N}$ mit $f^q \in I$. Es gilt für alle $p \in \mathcal{V}(I)$, daß $f^q(p) = 0$. Wir erhalten, da L keine Nullteiler enthält:

$$\forall p \in \mathcal{V}(I): \quad f(p) = 0$$

Daraus folgt $f \in \mathcal{I} \circ \mathcal{V}(I)$.

„$\subseteq$“ Sei $f \in \mathcal{I} \circ \mathcal{V}(I)$. Zu zeigen ist, daß $f \in \sqrt{I}$. Wir nehmen an, $f \neq 0$, weil sonst die Aussage trivialerweise gilt. Betrachte die Polynome $f_1, f_2, \ldots, f_r, 1 - zf$ in $L[x_1, x_2, \ldots, x_n, z]$.

BEHAUPTUNG 1: $\mathcal{V}\big(\{f_1, f_2, \ldots, f_r, 1 - zf\}\big) = \emptyset$

BEWEIS. Angenommen, $\mathcal{V}\big(\{f_1, f_2, \ldots, f_r, 1 - zf\}\big) \neq \emptyset$. Wähle:

$$p =: (\xi_1, \xi_2, \ldots, \xi_n, \xi) \in \mathcal{V}\big(\{f_1, f_2, \ldots, f_r, 1 - zf\}\big)$$

Dann gilt:

$$\underbrace{(1 - zf)(p)}_{\text{Polynom } 1 - zf \text{ an Stelle p}} = 1 - \xi \cdot f(\xi_1, \xi_2, \ldots, \xi_n) = 0$$

Da $(\xi_1, \xi_2, \ldots, \xi_n)$ aber eine Nullstelle von f ist, gilt gleichzeitig $(1 - zf)(p) = 1$ — Widerspruch. □

Nach dem Hilbert'schen Nullstellensatz 13.5.12 mit $I = (f_1, f_2, \ldots, f_r, 1 - zf)$ folgt, da nach Behauptung 1 $\mathcal{V}(I) = \emptyset$:

$$1 \in (f_1, f_2, \ldots, f_r, 1 - zf) = L[x_1, x_2, \ldots, x_n, z]$$

Daher existieren $g_1, g_2, \ldots, g_r, g \in L[x_1, x_2, \ldots, x_n, z]$ mit:

$$1 = g_1 f_1 + g_2 f_2 + \ldots g_r f_r + g(1 - zf) \tag{13.5}$$

Setze $z := \frac{1}{f}$. Wir erhalten aus (13.5) mit $g'_1, g'_2, \ldots, g'_r \in L[x_1, x_2, \ldots, x_n]$ und $q_i := \operatorname{grad}_z(g_i)$:

$$1 = \frac{g'_1}{f^{q_1}} \cdot f_1 + \frac{g'_2}{f^{q_2}} \cdot f_2 + \cdots + \frac{g'_r}{f^{q_r}} \cdot f_r + g \cdot 0 \tag{13.6}$$

Definiere:

$$g''_i = g'_i \cdot f^{q - q_i} \in L[x_1, x_2, \ldots, x_n] \qquad i = 1, 2, \ldots, r$$

Es folgt mit $q := \max\{q_1, q_2, \ldots, q_r\}$ aus (13.6):

$$f^q = g''_1 \cdot f_1 + g''_2 \cdot f_2 + \cdots + g''_r \cdot f_r \in L[x_1, x_2, \ldots, x_n]$$

Aus $f^q \in L[x_1, x_2, \ldots, x_n]$ folgt $f \in \sqrt{I} = \{f \in L[x_1, x_2, \ldots, x_n] \mid \exists q \in \mathbb{N} : f^q \in I\}$.

■

Korollar 13.5.16
Sei $F \subseteq L[x_1, x_2, \ldots, x_n]$ und G Gröbner-Basis zu (F). F ist genau dann unlösbar, wenn ein $c \in (L \cap G)$ mit $c \neq 0$ existiert.

Beweis. Wir zeigen beide Implikationen:

„⇐“ Sei $c(L \cap G)$ mit $c \neq 0$. Es ist $c \in (F)$, da G Gröbner-Basis zu (F). Das konstante Polynom c hat keine Nullstelle.

„⇒“ Sei $\mathcal{V}(F) \neq \emptyset$. Nach dem Hilbert'schen Nullstellensatz (Satz 13.5.12) mit $I = (F) = (G)$ ist $1 \in (G)$. Also:

$$1 \xrightarrow[*]{G} 0$$

Aus $\mathrm{Hmono}(1) \cap \mathrm{Head}(G) = \emptyset$ folgt die Existenz eines $c \in (G \cap L)$ mit $c \neq 0$.

■

Aus Korollar 13.5.16 erhalten wir Algorithmus 13.5.1, um zu entscheiden, ob $F \subseteq L[x_1, x_2, \ldots, x_n]$ lösbar ist.

Algorithmus 13.5.1 Entscheidungsalgorithmus für Lösbarkeit

EINGABE: $F \subseteq L[x_1, x_2, \ldots, x_n]$, L algebraisch abgeschlossen

1. Berechne Gröbner-Basis G zu (F)
2. IF ($\nexists c \in (G \cap L) : c \neq 0$) THEN gib „true“ aus ELSE gib „false“ aus

AUSGABE: „true“ genau dann, wenn $\mathcal{V}(F) \neq 0$

Definition 13.5.17 (Unabhängige Variable, Dimension eines Ideals)
Sei K Körper und $F \subseteq K[x_1, x_2, \ldots, x_n]$. Die Variablen $x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(l)}$ sind unabhängig bezüglich I, wenn:

$$I \cap K\left[x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(l)}\right] = (0)$$

Das heißt: Es gibt keine nichttriviale Relation zu den Variablen $x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(l)}$, die in dem Ideal $I \setminus \{0\}$ ist. Die Null kann nur als Nullpolynom dargestellt werden. Wir definieren $\dim I$ *als die Maximalzahl unabhängiger Variablen bezüglich I.*

Beispiel 13.5.18 (Dimension eines Ideals)
Es gilt $\dim K[x_1, x_2, \ldots, x_n] = 0$ und $\dim((x_i)) = n - 1$. ◇

Es gilt:

Satz 13.5.19
Sei $F \subseteq L[x_1, x_2, \ldots, x_n]$ und G eine Gröbner-Basis zu (F) bezüglich $<_{\mathrm{lex}}$. Folgende Aussagen sind äquivalent:

1. *F ist endlich lösbar.*
2. $\dim((F)) = 0$ *und* $(F) \neq L[x_1, x_2, \ldots, x_n]$.
3. *G ist in strikter Dreiecksform (vergleiche Definition 13.5.2 auf Seite 164).*

Beweis. Wir zeigen die Äquivalenz durch einen Ringbeweis:

- F ist endlich lösbar $\Rightarrow$ $\dim(F) = 0$ und $(F) \neq (1)$

  Weil F endlich lösbar ist, gilt:

  $$(F) \cap L = (0) \neq (1) \tag{13.7}$$

  Seien

  $$\mathcal{V}(F) =: \{(\xi_{i,1}, \xi_{i,2}, \ldots, \xi_{i,n}) \mid i = 1, 2, \ldots, m\} \subseteq L^n$$

  die gemeinensamen Nullstellen von F. Definiere:

  $$f_j(x_j) := \prod_{i=1}^{m} (x_j - \xi_{i,j}) \qquad j = 1, 2, \ldots, n$$

  Es gilt: $f_j(x_j)$ ist ein Polynom vom Grad m in x_j, das auf allen gemeinsamen Nullstellen $\mathcal{V}(F)$ verschwindet. Nach Konstruktion ist $f_j \in \mathcal{I} \circ \mathcal{V}(F)$. Nach dem allgemeinen Hilbert'schen Nullstellensatz (Satz 13.5.15) gilt:

  $$\exists q \in \mathbb{N} : \quad f_j^q \in (F)$$

  Somit gilt $(F) \cap L[x_j] \neq 0$ für $j = 1, 2, \ldots, n$. Wir erhalten $\dim((F)) = 0$, denn es gibt keine unabhängigen Variablen. Da nach (13.7) gilt

  $$(0) \subsetneq (F) \subsetneq (1) = L[x_1, x_2, \ldots, x_n],$$

  ist $(F) \neq L[x_1, x_2, \ldots, x_n]$.

- $\dim((F)) = 0$ und $(F) \neq L[x_1, x_2, \ldots, x_n]$ $\Rightarrow$ G ist in strikter Dreiecksform

  Wegen $\dim((F)) = 0$ und $(F) \cap L = (0)$ (sonst wäre $1 \in (F)$ und $(F) = L[x_1, x_2, \ldots, x_n]$), folgt:

  $$(F) \cap L[x_j] \neq \emptyset \qquad j = 1, 2, \ldots, n$$

  Sei $f_j(x_j) \in ((F) \cap L[x_j])$. Da G eine Gröbner-Basis zu (F) ist, gilt für ein $D_j \in \mathbb{N}$ und $a_j \in L$:

  $$\text{Hmono}\,(f(x_j)) = a_j x_j^{D_j} \in \text{Head}(G)$$

  Es ist $a_j \neq 0$, da sonst $a_j x_j^{D_j}$ nicht das Head-Monom wäre. Wegen $(F) \cap L = (0)$ ist $D_j > 0$. Wir erhalten:

  $$\exists g_i \in G : \quad \text{Hterm}(g_j) = x_j^{d_j} \qquad 0 < d_j \leq D_j$$

  Da G eine Gröbner-Basis bezüglich der zulässigen Ordnung $>_{\text{lex}}$ ist, folgt:

  $$g_j \in L[x_j, x_{j+1}, \ldots, x_n] \setminus L[x_{j+1}, x_{j+2}, \ldots, x_n]$$

  Wir erhalten, daß für $i = 1, 2, \ldots, n$ ein $g_{i+1} \in G_i$ existiert (vergleiche Definition 13.5.1 auf Seite 164), so daß g_{i+1} ein Monom der Form $a x_{i+1}^d$ mit $a \in L \setminus \{0\}$ und $d > 0$ enthält. G, eine Gröbner-Basis zu (F) bezüglich $<_{\text{lex}}$, kann in strikter Dreiecksform ausgedrückt werden.

- G ist in strikter Dreiecksform $\Rightarrow$ F ist endlich lösbar

  Sei $G_0, G_1, \ldots, G_n$ die strikte Dreiecksform von G, d.h.:

  1. $(G_n) = (G \cap L) = \{0\}$

2. Für $i = 0, 1, \ldots, n-1$ existiert ein $g_i \in G_i$, das ein Monom der Form ax_{i+1}^d mit $a \in L \setminus \{0\}$ und $d \in \mathbb{N}$ enthält.

Sei $I := (F) = (G)$. Es gilt:

$$\mathcal{V}(I) = \mathcal{V}(F) = \mathcal{V}(G)$$

Es genügt zu zeigen, daß I endlich lösbar ist. Wegen $(G \cap L) = I \cap L = \{0\}$ ist $1 \notin I$ und I lösbar. Wir zeigen durch Induktion über i, daß für $i = 0, 1, \ldots, n-1$ das i-te Eliminationsideal I_i nur endlich viele Nullstellen hat. Aus Korollar 13.5.7 wissen wir, daß

$$I_i = \left(\bigcup_{j=i}^{n} G_j \right)$$

und $\bigcup_{j=i}^{n} G_j$ in strikter Dreiecksform ist.

– Induktionsverankerung für $i = n - 1$:

  G_{n-1} besteht nur aus monovariablen Polynomen in x_n. Da G in strikter Dreiecksform ist, gibt es ein Polynom $p(x_n) \in G_{n-1}$ mit maximalem (minimalem?) Grad d_n. Das Polynom $p(x_n)$ hat höchstens d Nullstellen. Wir suchen die gemeinsamen Nullstellen von I_{n-1}. Wegen $p(x_n) \in I_{n-1}$ hat I_{n-1} höchstens d Nullstellen, also nur endlich viele.

– Induktionsschluß von $i + 1$ auf i:

  Nach Induktionsannahme hat das $(i + 1)$-te Eliminationsideal I_{i+1} nur endlich viele Nullstellen. Sei D_{i+2} die Anzahl. Definiere Projektion $\Pi : A^{n-i} \to A^{n-i-1}$, wobei $A := L[x_1, x_2, \ldots, x_n]$:

  $$(\xi_{i+1}, \xi_{i+2}, \ldots, \xi_n) \mapsto (\xi_{i+2}, \xi_{i+3}, \ldots, \xi_n)$$

  Wir unterteilen die Nullstellen $\mathcal{V}(I_i)$ des i-ten Eliminationsideals I_i in Äquivalenzklassen bezüglich der folgenden Äquivalenzrelation „$\sim$“. Seien $P, Q \in \mathcal{V}(I_i)$:

  $$P \sim Q \qquad \Longleftrightarrow \qquad \Pi(P) = \Pi(Q)$$

  Man überlege sich, daß für Ideal $A \subseteq L[y_1, y_2, \ldots, y_n]$ und $J := A \cap L[y_2, y_3, \ldots, y_n]$ gilt [Mishra93, Proposition 4.3.3]:

  $$\Pi\left(\mathcal{V}(I)\right) \subseteq \mathcal{V}(J)$$

  Nach Induktionsannahme und obiger Beobachtung ist die Anzahl der Äquivalenzklassen endlich, sogar kleiner oder gleich D_{i+2}. Sei

  $$p(x_{i+1}, x_{i+2}, \ldots, x_n) \in \bigcup_{j=i}^{n} G_j$$

  ein Polynom, das ein Monom der Form $ax_{i+1}^{d_{i+1}}$ mit $a \in L \setminus \{0\}$ und $d_{i+1} > 0$ enthält. Wir wählen d_{i+1} maximal. Sei die Äquivalenzklasse von P, einer gemeinsamen Nullstelle von I_i:

  $$[P]_\sim = \{Q \mid \Pi(Q) = (\xi_{i+1}, \xi_{i+2}, \ldots, \xi_n)\}$$

  Dann ist ξ mit

  $$Q = (\xi, \xi_{i+2}, \xi_{i+3}, \ldots, \xi_n) \in [P]_\sim$$

  eine Nullstelle des monovariablen Polynoms $p\left(x_{i+1}, \xi_{i+2}, \xi_{i+3}, \ldots, \xi_n\right)$. Daher:

  $$|[P]_\sim| \leq d_{i+1},$$

  und I_i hat nur endlich viele Nullstellen, nämlich maximal $d_{i+1} \cdot D_{i+2}$.

Aus obiger Argumentation erhalten wir als obere Schranke $d_1 d_2 \dots d_n$ für die Anzahl gemeinsamer Nullstellen eines Systems von Polynomen, wobei d_i der höchste Grad eines Terms der Form $x_i^{d_i}$ der Polynome in G_{i-1} ist.

■

Aus den Ideen dieses Abschnitts kann man einen Algorithmus zur Berechnung der gemeinsamen Nullstellen herleiten. Dieser bestimmt für $i = n, n-1, \dots, 1$ die gemeinsamen Nullstellen des i-ten Eliminationsideals und bildet aus diesen die gemeinsamen Nullstellen des $(i-1)$-ten Eliminationsideals. Dazu bestimmt er die Nullstellen eines monovariablen Polynoms. Den Algorithmus findet man in Mishras Buch [Mishra93] im Kapitel 4.4.

Algorithmenverzeichnis

Index

Literaturverzeichnis

[Babai86] L. Babai (1986): **On Lovász' Lattice Reduction and the nearest Lattice Point Problem**, Combinatorica, Band 6, Seiten 1–13.

[Barnes59] E.S. Barnes (1959): **The Contruction of perfect and extreme Forms II**, Acta Arithmetica, Band 5, Seiten 205-222.

[BaKa84] A. Bachem und R. Kannan (1984): **Lattices and the Basis Reduction Algorithm**, Technischer Report, Carnegie-Mellon-Universität (USA).

[BeWe93] Th. Becker und V. Weispfennig (1993): **Gröbner Bases — a computational Approach to commutative Algebra**, Graduate Texts in Mathematics, Band 141, Springer-Verlag, Berlin/Heidelberg.

[Blich14] H.F. Blichfeldt (1914): **A new Principle in the Geometry of Numbers with some Applications**, Transaction of the American Mathematical Society, Band 15, Seiten 227–235.

[Blich29] H.F. Blichfeldt (1929): **The Minimum Value of quadratic Forms and the closet Packing of Sphere**, Mathematische Annalen, Band 101, Seiten 366-389.

[Blich35] H.F. Blichfeldt (1935): **The minimum Value of positive Quadratic Forms in six, seven and eight Variables**, Mathematische Zeitschrift, Band 39, Seiten 1–15.

[Bb65] B. Buchenberger (1965): **Ein Algorithmus zum Auffinden der Basiselemente des Restklassenrings nach einem nulldimensionalen Polynomideal**, Dissertation, Fachbereich Mathematik, Universität Insbruck (Österreich).

[Cassels71] J.W.S. Cassels (1971): **An Introduction to the Geometry of Numbers**, Springer-Verlag, Berlin/Heidelberg.

[Cohen93] H. Cohen (1993): **A Course in Computational Algebraic Number Theory**, Graduate Texts in Mathematics, Band 138, Springer-Verlag, Berlin/Heidelberg.

[CoSl88] J.H. Conway und N.J. Sloane (1988): **Sphere Packings, Lattices and Groups**, Springer-Verlag, New York.

[CJLOSS92] M.J. Coster, A. Joux, B.A. LaMacchina, A.M. Odlyzko, C.P. Schnorr und J. Stern (1992): **An improved low-density Subset Sum Algorithm**, Computational Complexity, Band 2, Seiten 111–128.

[CR88] B. Chor und R.L. Rivest (1988): **A Knapsack type Public Key Cryptosystem based on Arithmetic in finite Fields**, IEEE Transaction Information Theory, Band IT-34, Seiten 901–909.

[Di1842] G.L. Dirichlet (1842): **Verallgemeinerung eines Satzes aus der Lehrere von Kettenbrüchen nebst einigen Anwendungen auf die Theorie der Zahlen**, Bericht über die zur Bekanntmachung geeigneter Verhandlungen der Königlich Preussischen Akademie der Wissenschaften zu Berlin, Seiten 93–95.

[Dåmgard89] I.B. Dåmgard (1989): **A Design Principle for Hash Functions**, Advances in Cryptology — Proceedings EuroCrypt '89, Lecture Notes in Computer Science, Band 435 (1990), Springer-Verlag, Berlin/Heidelberg, Seiten 416–427.

[Dantzig63] G.B. Dantzig (1963): **Linear Programming and Extensions**, Princeton University Press, Princeton, New Jersey (dt. Übersetzung „Lineare Programmierung und Erweiterungen" 1966 im Springer-Verlag, Berlin/Heidelberg, erschienen).

[DKT87] P.D. Domich, R. Kannan und L.E. Trotter (1987): **Hermite normal Form Computation using modulo Determinant Arithmetic**, Mathematics of Operation Research, Band 12, Nr. 1 (Februar), Seiten 50–59.

[EmBoas81] P. van Emde Boas (1981): **Another $\mathcal{NP}$-complete Partition Problem and the Complexity of Computing short Vectors in a Lattice**, Technischer Report 81-04, Fachbereich Mathematik der Universität Amsterdam.

[Euchner91] M. Euchner (1991): **Praktische Algorithmen zur Gitterreduktion und Faktorisierung**, Diplomarbeit, Fachbereich Informatik der Johann-Wolfgang-Goethe-Universität, Frankfurt/Main.

[Feller68] W. Feller (1968): **An Introduction to Probability Theory and its Application**, Band I, 3. Auflage, John Wiley & Sons, New York.

[Frieze86] A.M. Frieze (1986): **On the Lagarias-Odlyzko Algorithm for the Subset Sum Problem**, SIAM Journal on Computing, Band 15, Nr. 2, Seiten 536–539.

[GaSi78] J. von zur Gathen und M. Sieveking (1978): **A Bound on Solution of linear Integer Equations and Inequations**, Proceedings of the American Mathematical Society, Band 72, Seiten 155–158.

[GaJo79] M.R. Garey und D.S. Johnson (1979): **Computer and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completness**, W.H. Freeman and Company, San Francisco.

[Gauß1801] C.F. Gauß (1801): **Disquisitiones Arithmeticae**, Gerhard Fleischer, Leipzig. Deutsche Übersetzung (1889): „Untersuchung über höhere Arithmetik", Springer-Verlag, Berlin/Heidelberg.

[GrLek87] M. Gruber und C.G. Lekkerkerker (1987): **Geometry of Numbers**, 2. Auflage, North-Holland, Amsterdam.

[GLLS88] M. Grötschel, L. Lovász und A. Schrijver (1988): **Geometric Algorithms and combinatorial Optimization**, Algorithms and Combinatorics, Band 2, Springer-Verlag, Berlin/Heidelberg.

[HaMcC91] J. Hafner und K. McCurley (1991): **Asymptotic Fast Triangulation of Matrices over Ring**, SIAM Journal on Computing, Band 20, Nr. 6, Seiten 1068–1083.

[HJLS89] J. Håstad, B. Just, J.C. Lagarias und C.P. Schnorr (1989): **Polynomial Time Algorithms for Finding Integer Relations among real Numbers**, SIAM Journal on Computing, Band 18, Nr. 5, Seiten 859–881.

[Helfrich85] B. Helfrich (1985): **Algorithms to construct Minkowski reduced and Hermite reduced Lattice Bases**, Theoretical Computer Science, Band 41, Seiten 125–139.

[Hermite1850] C. Hermite (1850): **Extraits de lettres de M. Ch. Hermite à M. Jacobi sur differents objets de la théorie des nombres, Deuxième lettre**, Reine Angewandte Mathematik, Band 40, Seiten 279–290.

[Hlawka44] E. Hlawka (1944): **Zur Geometrie der Zahlen**, Mathematische Zeitschrift, Band 49, Seiten 285–312.

[Hörner94] H.H. Hörner (1994): **Verbesserte Gitterbasenreduktion; getestet am Chor-Rivest-Kryptosystem und an allgemeinen Rucksackproblemen**, Diplomarbeit, Fachbereich Mathematik der Johann-Wolfgang-Goethe-Universität, Frankfurt/Main.

[John48] F. John (1948): **Extremum Problems with Inequalities as subsidiary Conditions**, in K.O. Friedrichs, O.E. Neugebauer und J.J. Stoker (Ed.): „Studies and Essays presented to R. Courant on his 60th Birthday Januar 8, 1948“, Interscience Publisher, New York, Seiten 187–204.

[JoSt94] A. Joux und J. Stern (1994): **Lattice Reduction: A Toolbox for the Cryptanalyst**, Technischer Report, DGA/CELAR, Bruz (Frankreich). Eingereicht bei Journal of Cryptology.

[KaLe78] G.A. Kabatiansky und V.I. Levenshtein (1978): **Bounds for Packings on a Sphere and in Space**, Problems of Information Transmission, Band 14, Seiten 1–17.

[Kaib91] M. Kaib (1991): **The Gauß Lattice Basis Reduction succeeds with any Norm**, Proceedings of Fundamentals of Computation Theory (FCT '91), Springer Lecture Notes in Computer Science, Band 591, Seiten 275–286.

[Kaib94] M. Kaib (1994): **Gitterbasenreduktion für beliebige Normen**, Dissertation, Fachbereich Mathematik der Johann-Wolfgang-Goethe-Universität, Frankfurt/Main.

[KaSchn96] M. Kaib und C.P. Schnorr (1996): **The Generalized Gauss Reduction Algorithm**, Journal of Algorithms, Band 21, Nr. 3 (November), Seiten 565–578.

[KaBa79] R. Kannan und A. Bachem (1979): **Polynomial Algorithm for Computing the Smith and the Hermite Normal Form of an Integer Matrix**, SIAM Journal on Computing, Band 8, Seiten 499–507.

[Kannan87] R. Kannan (1987): **Minkowski's Convex Body Theorem and Integer Programming**, Mathematics of Operation Research, Band 12, Nr. 3 (August), Seiten 415–440.

[Karma84] M. Karmarkar (1984): **A new Polynomial-Time Algorithm for Linear Programming**, Combinatorica, Band 4, Seiten 373–395.

[Khach79] L.G. Khachiyan (1979): **A Polynomial Algorithm in Linear Programming**, Soviet Mathmatics Doklady, Band 20, Seiten 191–194.

[Khach80] L.G. Khachiyan (1980): **Polynomial Algorithms in Linear Programming**, U.S.S.R. Computational Mathematics and Mathematical Physics, Band 20, Seiten 53–72.

[KoZo1872] A. Korkine und G. Zolotareff (1872): **Sur les formes quadratique positive quaternaires**, Mathematische Annalen, Band 5, Seiten 366-389.

[KoZo1873] A. Korkine und G. Zolotareff (1873): **Sur les formes quadratique**, Mathematische Annalen, Band 6, Seiten 366–389.

[KoZo1877] A. Korkine und G. Zolotareff (1877): **Sur les formes quadratique positive**, Mathematische Annalen, Band 11, Seiten 242–292.

[Knuth71] D.E. Knuth (1971): **The Art of Computer Programming**, Fundamental Algorithms, Band I, Addison-Wesley, Reading.

[LARIFARI] M. Kaib, R. Mirwald, C. Rössner, H.H. Hörner, H. Ritter (1994): **Programmieranleitung für LARIFARI — Version 13.07.1994**, Fachbereiche Mathematik und Informatik der Johann-Wolfgang-Goethe-Universität, Frankfurt/Main.

[La1773] J.L. Lagrange (1773): **Recherches d'arithmétique**, Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres, Berlin, Seiten 265–312.

[Lang93] S. Lang (1993): **Algebra**, 3. Auflage, Addison-Wesley, Reading.

[LLS90] J.C. Lagarias, H.W. Lenstra und C.P. Schnorr (1990): **Korkin-Zolotarev Bases and successive Minima of a Lattice and its reciprocal lattice**, Combinatorica, Band 10, Seiten 333–348.

[LaOd85] J.C. Lagarias und A.M. Odlyzko (1985): **Solving low-density Subset Sum Problems**, Journal of ACM, Band 32, Nr. 1, Seiten 229–246.

[LLL82] A.K. Lenstra, H.W. Lenstra und L. Lovász (1982): **Factoring Polynomials with Rational Coefficients**, Springer Mathematische Annalen, Band 261, Seiten 515–534.

[Lenstra83] H.W. Lenstra (1983): **Integer Programming in a fixed Number of Variables**, Mathematics of Operation Research, Band 8, Nr. 4 (November), Seiten 538–548

[Lováz86] L. Lovász (1986): **An algorithmic Theory of Numbers, Graphs and Convexity**, CBMS-NSF Regional Conference Series in Applied Mathematics, Band 50, SIAM Publications, Philadelphia.

[LoSc92] L. Lovász und H. Scarf (1992): **The Generalized Basis Reduction Algorithm**, Mathematics of Operation Research, Band 17, Nr. 3 (August), Seiten 751–764.

[MaOd90] J.E. Mazo und A.M. Odlyzko (1990): **Lattice Points in high-dimensional Sphere**, Monatsheft Mathematik, Band 110, Seiten 47–61.

[Mink1896] H. Minkowski (1896): **Geometrie der Zahlen**, erste Auflage, Teubner-Verlag, Leipzig.

[Mink1911] H. Minkowski (1911): **Gesammelte Abhandlungen**, Band I und II, Teubner-Verlag, Leipzig.

[Mishra93] B. Mishra (1993): **Algorithmic Algebra**, Texts and Monographs in Computer Science, Springer-Verlag, New-York.

[Orton1994] G. Orton (1994): **A multiple-iterated Trapdoor for dense compact Knapsacks**, Advances in Cryptology — Proceedings EuroCrypt '94, Lecture Notes in Computer Science, Band 950 (1995), Springer-Verlag, Berlin/Heidelberg, Seiten 112–130.

[PaSchn87] A. Paz und C.P. Schnorr (1987): **Approximating Integer Lattices by Lattices with cyclic Factor Group**, 14.th International Colloquium on Automata, Languages and Programming (ICALP), Lecture Notes in Computer Science, Band 267, Springer-Verlag, Berlin/Heidelberg, Seiten 386–393.

[Rieger78] G.J. Rieger (1978): **über die mittlere Schrittzahl bei Divisionsalgorithmen**, Mathematische Nachrichten, Band 82, Seiten 157–180.

[Ritter96] H. Ritter (1996): **Breaking Knapsack Cryptosystems by ℓ_∞-norm Enumeration**, Proceedings of the 1.st International Conference on the Theory and Applications of Cryptography — PragoCrypt '96, CTU Publishing House, Prag, Seiten 480–492.

[Ritter97] H. Ritter (1997): **Aufzählung kurzer Gittervektoren in allgemeiner Norm**, Dissertation, Fachbereich Mathematik der Johann-Wolfgang-Goethe-Universität, Frankfurt/Main.

[Rogers64] C.A. Rogers (1964): **Packing and Covering**, Cambridge University Press, Cambridge.

[Schnorr87] C.P. Schnorr (1987): **A Hierarchy of polynomial time Lattice Basis Reduction Algorithms**, Theoretical Computer Science, Band 53, Seiten 201–224.

[Schnorr88] C.P. Schnorr (1988): **A more efficient Algorithm for Lattice Basis Reduction**, Journal of Algorithms, Band 9, Seiten 47–62.

[Schnorr91a] C.P. Schnorr (1991): **Gittertheorie und ganzzahlige Optimierung**, Skript zur Vorlesung, Johann-Wolfgang-Goethe-Universität, Frankfurt/Main.

[Schnorr91b] C.P. Schnorr (1991): **Factoring Integers and Computing Discrete Logarithms via Diophantine Approximation**, Advances in Cryptology – Proceedings EuroCrypt '91, Lecture Notes in Computer Science, Band 547, Springer-Verlag, Berlin/Heidelberg, Seiten 171–181.

[SchnEu91] C.P. Schnorr und M. Euchner (1991): **Lattice Basis Reduction: improved Algorithms and solving Subset Sum Problems**, Proceedings of Fundamentals of Computation Theory (FCT '91), Lecture Notes in Computer Science, Band 591, Springer-Verlag, Berlin/Heidelberg, Seiten 68–85.

[Schnorr93] C.P. Schnorr (1993): **Factoring Integers and Computing Discrete Logarithms via Diophantine Approximation**, Advances in Computational Complexity, Ed. Jim-Yi Cai, AMS DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Band 13, Seiten 171–182.

[Schnorr94a] C.P. Schnorr (1994): **Block Reduced Lattice Bases and Successive Minima**, Combinatorics, Probability and Computing, Band 3, Seiten 507–522.

[Schnorr94b] C.P. Schnorr (1994): **Gittertheorie und Kryptographie**, Ausarbreitung, Johann-Wolfgang-Goethe-Universität, Frankfurt/Main.

[SchnHö95] C.P. Schnorr und H.H. Hörner (1995): **Attacking the Chor-Rivest Cryptosystem by improved Lattice Reduction**, Advances in Cryptology — Proceedings EuroCrypt '95, Lecture Notes in Computer Science, Band 921, Springer-Verlag, Berlin/Heidelberg, Seiten 1–12.

[Schrijver86] A. Schrijver (1986): **Theory of Linear and Integer Programming**, Wiley-Interscience Series in discrete Mathematics and Optimization, John Wiley & Son Ltd.

[Seysen93] M. Seysen (1993): **Simultaneous Reduction of a Lattice and its reciprocal Basis**, Combinatorica, Band 13, Seiten 363–376.

[Siegel89] C.L. Siegel (1989): **Lectures on the Geometry of Numbers**, Springer-Verlag, Berlin/Heidelberg.

[Smith1861] H.J.S. Smith (1861): **On Systems of linear indeterminate Equations and Congruences**, Philosophical Transaction of the Royal Society of London, Band 151, Seiten 293–326.

[SpStr76] E. Specker und V. Strassen (1976): **Komplexität von Entscheidungsproblemem**, Lecture Notes in Computer Science, Band 43, Springer-Verlag, Berlin/Heidelberg.

[Vetchin82] N.M. Vetchinkin (1982): **Uniqueness of Classes of positive quadratic Forms on which Values of the Hermite Constants are attained for** $6 \leq n \leq 8$, Proceedings of the Steklov Institute of Mathematics, Nr. 3, Seiten 37–95.

[Watson66] G.L. Watson (1966): **On the Minimum of a positiv Quadratic Form in** n **(**$n \leq 8$**) Variables (Verification of Blichfeldt's Calculations)**, Proceeedings of the Cambrigde Philosophical Society (Mathematical and Physical Science), Band 62, Seite 719.

[Ye91] Y. Ye (1991): **Potential Reduction Algorithm for Linear Programming**, Mathematical Programming, Band 51, Seiten 239–258.