Lattice cryptography: isomorphisms & dual attacks

Pulles, L.N.

Chapter 2

Preliminaries

The sets of complex, natural, rational and real numbers, and integers, are denoted by $\mathbb{C}, \mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{Z}$, respectively. The natural numbers, $\mathbb{N}$ include 0. The logarithm in base $2, e$ and 10, are denoted by $\lg, \ln$ and $\log$, respectively. We may write $[n] = \{1, 2, \ldots, n\}$. Boldfaced uppercase and lowercase letters denote matrices and column vectors, respectively. More notations can be found in the back matter of this thesis in the List of Symbols.

2.1 Elementary theory

For later use, we shortly explain some terminology, and things related to geometry and probability theory.

A square matrix $\mathbf{A}$ is called *unitriangular* if it is upper triangular, i.e., $\mathbf{A}_{ij} = 0$ for $1 \leq j < i \leq n$, and its diagonal consists of only ones.

2.1.1 Terminology

In this thesis, we will encounter heuristics. A *heuristic* is a simplification of a mathematical analysis that is not fully precise or rationalised but nevertheless decent enough of an approximation. It will be made clear whenever a heuristic is used.

Moreover, any result that use one or more heuristics will be called a *heuristic claim*. A *heuristic justification* then shows how the heuristic

claim is derived from the used heuristic(s), or motivates why it is generally believed to be true.

For instance, in Chapters 3 and 4, we derive various heuristic claims regarding random lattices from Heuristic 2.41 and/or Heuristic 2.42. We use heuristics because it would be very difficult to unconditionally prove a more precise version of these claims using a formally defined notion of random lattices.

2.1.2 Geometric objects

In this work, we consider Euclidean space $\mathbb{R}^n$ with the standard ℓ^2-norm induced by the dot product,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\mathsf{T} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i \ .$$

The ℓ^2-norm, or simply norm, of a point $\mathbf{x} \in \mathbb{R}^n$ is $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.

More generally, one may consider for any number $p \in \mathbb{Z}_{\geq 1}$ the ℓ^p-norm of a point $\mathbf{x} \in \mathbb{R}^n$ is given by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} \mathbf{x}_i^p \right)^{1/p} \ ,$$

and the ℓ^∞-norm of $\mathbf{x}$ is given by $\|\mathbf{x}\|_\infty = \max\{|\mathbf{x}_1|, \dots |\mathbf{x}_n|\}$.

The *Minkowski sum* of a subset $S \subseteq \mathbb{R}^n$ and vector $\mathbf{t} \in \mathbb{R}^n$ is denoted by $S + \mathbf{t} = \{\mathbf{s} + \mathbf{t} \in \mathbb{R}^n \mid \mathbf{s} \in S\}$, and the *Minkowski sum* of subsets $S, T \subseteq \mathbb{R}^n$ is denoted by $S + T = \cup_{\mathbf{t} \in T}(S + \mathbf{t})$. Given any $c \in \mathbb{R}$ and subset $S \subseteq \mathbb{R}^n$, we write $cS = \{c\mathbf{s} \in \mathbb{R}^n \mid \mathbf{s} \in S\}$.

For $n \in \mathbb{Z}_{\geq 2}$, the $(n-1)$-dimensional sphere is denoted by $\mathcal{S}^{n-1}$ and consists of all points $\mathbf{x} \in \mathbb{R}^n$ of norm 1. The *unit circle* $\mathcal{S}^1$ can be seen as a subgroup of $\mathbb{C}^*$ by considering the map $\mathbb{R}^2 \to \mathbb{C}, (x, y) \mapsto x + iy$. The n-dimensional closed ball is denoted by $\mathcal{B}^n$ and consists of all points $\mathbf{x}$ of norm $\mathbf{x} \leq 1$. For any $r \in \mathbb{R}_{>0}$, we have

$$\mathrm{Vol}_n(r\mathcal{B}^n) = \frac{r^n \pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} \geq (2r/\sqrt{n})^n \ , \tag{2.1}$$

where Γ denotes the *Gamma function*, which satisfies $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$. The lower bound follows from $[-r/\sqrt{n}, r/\sqrt{n}]^n \subseteq r\mathcal{B}^n$.

2.1.3 Probability theory

The probability of an event E happening is denoted by $\Pr[E]$. For a random variable X, we denote its *expectation value* by $\mathbb{E}[X]$; its *variance* by $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; its *standard deviation* (std. dev.) by $\sigma_X = \sqrt{\mathbb{V}[X]}$.

For a probability distribution $\mathcal{D}$ on $\mathbb{R}$, we denote its support by $\operatorname{Supp}\mathcal{D}$, the *cumulative distribution function* (CDF) by $x \mapsto \Pr_{X\leftarrow\mathcal{D}}[X \leq x]$, and the *survival function* (SF) by $\Pr_{X\leftarrow\mathcal{D}}[X \geq x] = 1 - \Pr_{X\leftarrow\mathcal{D}}[X < x]$. The *probability density function* (PDF) at $x \in \mathbb{R}$ is the derivative of $\Pr_{X\leftarrow\mathcal{D}}[X \leq x]$ with respect to x.

The *uniform distribution on a set* X is denoted by $\mathrm{U}(X)$.

The n-dimensional *Gaussian mass with parameter* $\sigma \in \mathbb{R}_{>0}$ is denoted by

$$\rho_\sigma \colon \mathbb{R}^n \to \mathbb{R}_{>0}, \quad \mathbf{x} \mapsto \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right) .$$

The *Gaussian, or normal distribution, with centre* $\mu \in \mathbb{R}$ *and parameter* $\sigma \in \mathbb{R}_{>0}$, is denoted by $\mathcal{N}(\mu, \sigma^2)$, and its PDF is given by $\frac{1}{\sigma\sqrt{2\pi}}\,\rho_\sigma(x-\mu)$. Note the Gaussian $\mathcal{N}(\mu, \sigma^2)$ has expected value μ and standard deviation σ. The *standard Gaussian* is $\mathcal{N}(0, 1)$. Given $n \geq 1$, *the* n*-dimensional centred Gaussian with parameter* $\sigma \in \mathbb{R}_{>0}$ is denoted by $\mathcal{N}_n(0, \sigma)$ and is the joint distribution of points $\mathbf{x} \in \mathbb{R}^n$ where each coefficient is independently sampled according to $\mathcal{N}(0, \sigma)$.

The *error function* is given by $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,\mathrm{d}t$. The CDF of the Gaussian $\mathcal{N}(\mu, \sigma^2)$ is expressible in terms of the error function as follows:

$$\Pr_{X\leftarrow\mathcal{N}(\mu,\sigma^2)}[X \leq x] = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \qquad (x \in \mathbb{R}). \tag{2.2}$$

Concretely, we have

$$\Pr_{X\leftarrow\mathcal{N}(\mu,\sigma^2)}[\mu - k\sigma \leq X \leq \mu + k\sigma] = \operatorname{erf}\left(\frac{k}{\sqrt{2}}\right) \approx \begin{cases} 68\% & k = 1, \\ 95\% & k = 2, \\ 99.7\% & k = 3. \end{cases}$$

The *complementary error function* is given by $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$, and has the following asymptotic decay rate.

Lemma 2.1 ([AS64, 7.1.24])**.** *For all $x > 0$,*

$$\operatorname{erfc}(x) \leq \frac{1}{\sqrt{\pi}x} \cdot e^{-x^2}.$$

Theorem 2.2 (central limit)**.** *Let $X_1, X_2, \ldots$ be* independent and identically distributed *(i.i.d.) random variables, each having mean μ and variance σ^2. As n tends to infinity, the random variable*

$$\sqrt{n} \cdot \left(\frac{X_1 + \cdots + X_n}{n} - \mu \right) ,$$

converges in distribution towards $\mathcal{N}(0, \sigma^2)$.

In concrete settings, we heuristically expect that above random variable is closely distributed as Gaussian for large n.

Heuristic 2.3 (central limit heuristic)**.** *Given some large number n and i.i.d. random variables $X_1, \ldots, X_n$, each having mean μ and variance σ^2, the random variable $(X_1 + \cdots + X_n)/n$ has distribution $\mathcal{N}(\mu, \sigma^2/n)$.*

Statistical distance and Rényi divergence

Definition 2.4. Let P, Q be two probability distributions on S. The *statistical distance, or total variation distance, between P and Q*, is

$$\Delta(P, Q) = \sup_{T \subseteq S} |P(T) - Q(T)| \ .$$

If S is discrete, we have $\Delta(P, Q) = \frac{1}{2} \sum_{x \in S} |P(x) - Q(x)|$.

The Rényi divergence [Rén61, vEH14] is used in security proofs to provide tighter reductions compared to a proof using statistical distance [Pre17, BLR+18].

Definition 2.5. Let P, Q be two probability distributions such that their supports satisfy $\operatorname{Supp}(P) \subseteq \operatorname{Supp}(Q)$. The *Rényi divergence of P from Q at order $a \in (1, \infty)$*, is given by

$$\mathrm{R}_a(P \parallel Q) = \left(\sum_{x \in \operatorname{Supp}(P)} \frac{P(x)^a}{Q(x)^{a-1}} \right)^{1/(a-1)} . \tag{2.3}$$

In the literature, one may find the Rényi divergence being defined as the logarithm of R_a in Equation (2.3).

In contrast to statistical distance or 'max-log distance' [MW17], Rényi divergence does not provide a metric, as it is not symmetric in its arguments and does not satisfy the triangle inequality.

If the Rényi divergence is finite, which it will be for our applications, we think of it as a value $1 + \delta$ for $\delta \geq 0$; the smaller δ is the closer the two distributions. Note that for any fixed P, Q on which the Rényi divergence is finite, $\mathrm{R}_a(P \parallel Q)$ is nondecreasing as a function of a, i.e. a larger order a defines a stricter notion of closeness.

Rényi divergence satisfies the following three properties.

Proposition 2.6. *Let $a \in (1, \infty)$ be an order. Two probability distributions P, Q satisfy:*

- *(data processing inequality) for any function f we have*

$$\mathrm{R}_a\Big(P^f \,\Big\|\, Q^f\Big) \leq \mathrm{R}_a(P \parallel Q) \quad ,$$

  in which P^f and Q^f denotes the distribution of $f(X)$ induced by sampling $X \leftarrow P$ and $X \leftarrow Q$, respectively.

- *(probability preservation) for any event $E \subseteq \mathrm{Supp}(P)$ we have*

$$Q(E) \geq P(E)^{a/(a-1)} / \mathrm{R}_a(P \parallel Q) \quad .$$

Moreover, (multiplicativity) for families of distributions $(P_i)_{i \in I}, (Q_i)_{i \in I}$, we have

$$\mathrm{R}_a\left(\prod_{i \in I} P_i \,\middle\|\, \prod_{i \in I} Q_i\right) = \prod_{i \in I} \mathrm{R}_a(P_i \parallel Q_i) \quad .$$

Often P is considered as an inexact realisation of some ideal distribution Q, and the event E as some adversary winning a search game; in this setting the probability preservation inequality allows us to bound from above the probability of an adversary winning the search game using distribution P. We also consider the order a as a parameter that we may optimise over.

2.1.4 Fourier transform

Definition 2.7. The *Fourier transform* of a function $f\colon \mathbb{R}^n \to \mathbb{R}$, for which $\int_{\mathbb{R}^n} f(\mathbf{x})^2 \mathrm{d}\mathbf{x}$ is a finite value, is the function $\widehat{f}\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$\widehat{f}(\mathbf{y}) = \int_{\mathbb{R}^n} e^{-2\pi i \langle \mathbf{x},\mathbf{y}\rangle} f(\mathbf{x}) \mathrm{d}\mathbf{x},$$

for all $\mathbf{y} \in \mathbb{R}^n$.

For example, the Fourier transform of ρ_σ is given by,

$$\widehat{\rho_\sigma}(\mathbf{y}) = \left(\sigma\sqrt{2\pi}\right)^n \rho_{1/(2\pi\sigma)}(\mathbf{y}).$$

The well-known Fourier inversion theorem states that, under certain convergence conditions, one may recover f from $\widehat{f}$ by:

$$f(\mathbf{x}) = \int_{\mathbb{R}^n} e^{2\pi i \langle \mathbf{x},\mathbf{y}\rangle} \widehat{f}(\mathbf{y}) \mathrm{d}\mathbf{y}.$$

If f is a *radial function*, i.e., there exists $g\colon \mathbb{R} \to \mathbb{R}$ such that $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, then $\widehat{f}$ is also a radial function, and vice versa [SW71, Thm. 3.3].

2.2 Lattice

The main subject of this thesis is the theory of Euclidean lattices. For a thorough introduction, see [Sie89, Cas96].

Definition 2.8. A *lattice* Λ is a discrete additive subgroup of $\mathbb{R}^n$.

Here, discrete means that the induced topology on Λ is the discrete topology, i.e., there is an open ball $r\mathcal{B}^n$ $(r > 0)$ around the origin such that $r\mathcal{B}^n \cap \Lambda = \{\mathbf{0}\}$ holds. In particular, there is a minimal distance between lattice points. For example, $\mathbb{Z}^n$ is a lattice. Figure 1.4 depicts two lattices in two-dimensional space. Important properties of lattices are rank, volume, successive minima and the covering radius, which we will cover below in this sequence.

Let us use $\mathrm{span}(S)$ to denote the $\mathbb{R}$-*linear span of a set* $S \subset \mathbb{R}^n$.

Definition 2.9 (rank)**.** The *rank* of a lattice $\Lambda \subset \mathbb{R}^n$ is denoted by $\mathrm{rk}(\Lambda)$, and equals the dimension of $\mathrm{span}(\Lambda)$. We say Λ is *full-rank* if $\mathrm{span}(\Lambda) = \mathbb{R}^n$, or equivalently the dimension of $\mathrm{span}(\Lambda)$ equals n.

Later, Proposition 2.59 will show that any lattice can be taken without loss of generality to be of full-rank after possibly applying an orthogonal transformation to it.

Definition 2.10 (volume)**.** The *volume*, or rather covolume, of a rank-k lattice $\Lambda \subseteq \mathbb{R}^n$ is $\det(\Lambda) = \mathrm{Vol}_k(\mathrm{span}(\Lambda)/\Lambda)$. A lattice Λ is called *unit-volume* if $\det(\Lambda) = 1$.

We may normalise a rank-k lattice Λ to have unit-volume, by considering the lattice

$$\frac{1}{\det(\Lambda)^{1/k}} \cdot \Lambda \ ,$$

in which each point in Λ is multiplied by $1/\det(\Lambda)^{1/k}$.

The volume is a measure of global sparsity of lattice points. Equivalently, $1/\det(\Lambda)$ measures the density of lattice points in Euclidean space, as the following proposition shows.

Proposition 2.11. *Let $\Lambda \subset \mathbb{R}^n$ be a full-rank lattice, and $V \subset \mathbb{R}^n$ be a measurable set. Then,*

$$\mathop{\mathbb{E}}_{\mathbf{t} \leftarrow \mathrm{U}(\mathbb{R}^n/\Lambda)} \left[|(V + \mathbf{t}) \cap \Lambda| \right] = \frac{\mathrm{Vol}_n(V)}{\det(\Lambda)} \ .$$

Definition 2.12 (successive minima)**.** For $i \in [\mathrm{rk}(\Lambda)]$, the *$i$th successive minimum of* Λ is given by

$$\lambda_i(\Lambda) = \inf \{ r \in \mathbb{R}_{>0} \mid \dim \mathrm{span}(\Lambda \cap r\mathcal{B}^n) \geq i \} \ ,$$

or equivalently it is the smallest $r > 0$ such that there are at least i points in Λ of norm $\leq r$ that are $\mathbb{R}$-linearly independent. In particular, the *first minimum*, $\lambda_1(\Lambda)$, is the length of a shortest nonzero vector of Λ.

Note that we have $0 < \lambda_1(\Lambda) \leq \lambda_2(\Lambda) \leq \cdots \leq \lambda_k(\Lambda) < \infty$. For example, $\Lambda = \mathbb{Z}^n$ satisfies $\lambda_1(\Lambda) = \cdots = \lambda_n(\Lambda) = 1$.

Example 2.13. Given $n \geq 4$, the lattice $\Lambda = \mathbb{Z}^n + \mathbb{Z} \cdot (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})^\mathsf{T}$ has the same successive minima as $\mathbb{Z}^n$ and contains the sublattice $\mathbb{Z}^n$ of index $[\Lambda : \mathbb{Z}^n] = 2$.

Example 2.14. For $k \in \mathbb{Z}_{\geq 1}$, the *root lattice of type A* is given by

$$A_k = \left\{ \mathbf{x} \in \mathbb{Z}^{k+1} \,\middle|\, \sum_i \mathbf{x}_i = 0 \right\} \ .$$

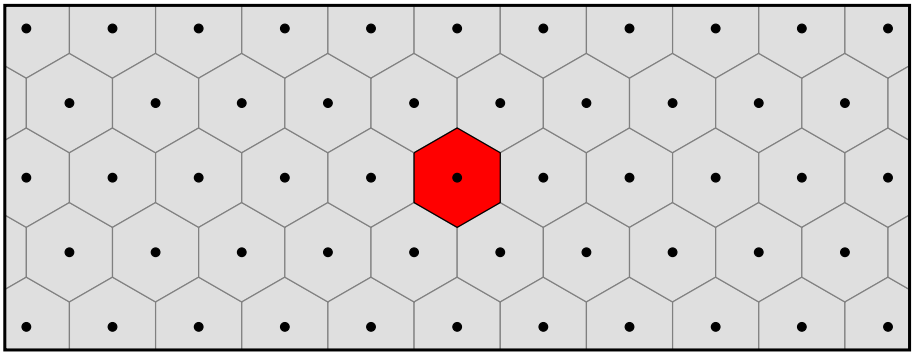


Figure 2.1: Voronoi diagram of the hexagonal lattice. The Voronoi cell, $\mathcal{V}(\Lambda)$, is highlighted in red.

The lattice A_k has rank k, volume $\sqrt{k+1}$, successive minima $\lambda_1(A_k) = \cdots = \lambda_k(A_k) = \sqrt{2}$, and $k(k+1)$ shortest vectors of this length [CS98].

Minkowski [Min1896] showed in 1896 that the first minimum can be bounded using $\det(\Lambda)$.

Theorem 2.15 (Minkowski's convex body theorem)**.** *Let $\Lambda \subset \mathbb{R}^n$ be a full-rank lattice and let $S \subseteq \mathbb{R}^n$ be of volume $\mathrm{Vol}_n(S) > 2^n \det(\Lambda)$, such that $-S \subseteq S$, and for all $\mathbf{x}, \mathbf{y} \in S$ and $\lambda \in [0,1]$ we have $\lambda\mathbf{x}+(1-\lambda)\mathbf{y} \in S$. Then, $S \cap \Lambda$ contains a nonzero lattice point.*

Proof. See Cassels [Cas96, III]. □

Minkowski's first theorem easily follows.

Theorem 2.16 (Minkowski's first)**.** *A full-rank lattice $\Lambda \subset \mathbb{R}^n$ satisfies*

$$\lambda_1(\Lambda) \leq 2 \cdot \frac{\det(\Lambda)^{1/n}}{\mathrm{Vol}_n(\mathcal{B}^n)^{1/n}} \ .$$

Proof. Consider $S = r\mathcal{B}^n$ for $r > 2 \cdot (\det(\Lambda)/\mathrm{Vol}_n(\mathcal{B}^n))^{1/n}$. As we have $\mathrm{Vol}_n(S) > 2^n \det(\Lambda)$, there exists $\mathbf{v} \in S \cap \Lambda \setminus \{\mathbf{0}\}$ so $\lambda_1(\Lambda) \leq r$. □

By placing balls of radius $\frac{1}{2}\lambda_1(\Lambda)$ centred at each lattice point we get a so-called lattice packing. A unit-volume lattice maximising its first minimum yields the densest lattice packing through this construction.

Table 2.1: Known Hermite constants

n	1	2	3	4	5	6	7	8	9	24
γ_n	1	$\sqrt{4/3}$	$\sqrt[3]{2}$	$\sqrt{2}$	$\sqrt[5]{8}$	$\sqrt[6]{\frac{64}{3}}$	$\sqrt[7]{64}$	2	2	4

Definition 2.17. *Hermite's constant* is given by

$$\gamma_n = \sup_{\Lambda \in \mathcal{X}_n} \lambda_1(\Lambda)^2 \ ,$$

where $\mathcal{X}_n$ is the family of rank-n unit-volume lattices.

By combining Theorem 2.16 and Equation (2.1), we obtain

$$\sqrt{\gamma_n} \leq 2\,\mathrm{GH}(n) \leq \sqrt{n}.$$

The values of γ_n are known for $n \leq 8$ [Bli29, Bli35], for $n = 24$ [CK09] and for $n = 9$ [DvW25], and can be found in Table 2.1. Hermite's constant in dimension 9 was determined using extensive machine-aided computation. A remarkable fact is that in dimension 8 and 24 the densest sphere packing is one in which the sphere centres form a lattice [Via17, CKM+17].

Definition 2.18 (Voronoi cell)**.** The *Voronoi cell of* Λ is denoted by $\mathcal{V}(\Lambda)$ and consists of all points that are not closer to any lattice point other than to the origin:

$$\mathcal{V}(\Lambda) = \{\mathbf{x} \in \mathrm{span}(\Lambda) \mid \forall \mathbf{v} \in \Lambda\colon \ \|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{v}\|\} \ .$$

Note that we have $\det(\Lambda) = \mathrm{Vol}_k(\mathcal{V}(\Lambda))$ for any rank-k lattice Λ. Figure 2.1 shows that the Voronoi cell provides a tiling of Euclidean space, also known as a Voronoi diagram, named after Voronoï [Vor1908a, Vor1908b].

Definition 2.19 (covering radius)**.** The *covering radius* of a lattice Λ is denoted by $\mu(\Lambda)$ and equals the furthest distance from any point to Λ:

$$\mu(\Lambda) = \sup_{\mathbf{x} \in \mathrm{span}(\Lambda)} \left\{ \min_{\mathbf{v} \in \Lambda} \{\|\mathbf{x} - \mathbf{v}\|\} \right\} \ .$$

The function $d(\mathbf{x}, \Lambda) = \min_{\mathbf{v} \in \Lambda} \|\mathbf{x} - \mathbf{v}\|$ is Λ-periodic, i.e., for all $\mathbf{v} \in \Lambda$ we have $d(\mathbf{x}, \Lambda) = d(\mathbf{x} + \mathbf{v}, \Lambda)$. Hence, one may take this supremum over all $\mathbf{x} \in \mathcal{V}(\Lambda)$ instead. As the origin is the closest lattice point to any point in $\mathcal{V}(\Lambda)$, the covering radius equals

$$\mu(\Lambda) = \max\{\|\mathbf{x}\| \mid \mathbf{x} \in \mathcal{V}(\Lambda)\} \ .$$

2.2.1 Lattice basis

In order to algorithmically work with a lattice Λ, we describe a lattice by a so-called generating set, or a so-called basis.

Definition 2.20 (generating set and basis)**.** A *generating set* for a lattice $\Lambda \subset \mathbb{R}^n$ is a finite list of vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m \in \Lambda$ such that for all $v \in \Lambda$ there exist $c_1, \ldots, c_m \in \mathbb{Z}$ such that $\mathbf{v} = c_1\mathbf{b}_1 + \cdots + c_m\mathbf{b}_m$ holds.

A *basis* $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_m] \in \mathbb{R}^{n \times m}$ is a matrix such that $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are $\mathbb{R}$-linearly independent. We say $\mathbf{B}$ is *a basis for* Λ whenever $\mathbf{b}_1, \ldots, \mathbf{b}_m$ is a generating set for Λ. Given a basis $\mathbf{B} \in \mathbb{R}^{n \times m}$, *the lattice generated by* $\mathbf{B}$ is denoted by

$$\mathcal{L}(\mathbf{B}) = \{\mathbf{y} \in \mathbb{R}^n \mid \exists \mathbf{x} \in \mathbb{Z}^m \colon \mathbf{y} = \mathbf{B} \cdot \mathbf{x}\}.$$

A generating set for Λ requires at least $\mathrm{rk}(\Lambda)$ many vectors, and a basis for Λ contains exactly $k = \mathrm{rk}(\Lambda)$ column vectors. It is easily verifiable that $\mathcal{L}(\mathbf{B})$ is a lattice, and $\mathbf{B}$ is a basis for $\mathcal{L}(\mathbf{B})$. The set of square bases $n = m$ corresponds to $\mathrm{GL}_n(\mathbb{R})$. The following proposition shows that every lattice Λ has a basis.

Proposition 2.21 ([Kan83, Prop. 1.1])**.** *Let $\Lambda \subset \mathbb{R}^n$ be a lattice, and $\mathbf{v} \in \Lambda$ be a nonzero lattice point. If $x\mathbf{v} \notin \Lambda$ for all $x \in (0, 1)$, then there exists a basis $\mathbf{B}$ for Λ, such that $\mathbf{B}$ contains $\mathbf{v}$.*

Corollary 2.22. *Every lattice Λ can be described by a basis $\mathbf{B}$ such that $\mathbf{b}_1$ is a shortest vector.*

Proof. Let $\mathbf{v} \in \Lambda$ be a shortest vector of Λ. Suppose $x\mathbf{v} \in \Lambda$ would hold for some $x \in (0, 1)$. In this case, we would have $\|x\mathbf{v}\| = x\|\mathbf{v}\| < \|\mathbf{v}\|$, which would contradict that $\mathbf{v}$ is a shortest vector of Λ. Thus, we may use Proposition 2.21, which gives a basis $[\mathbf{b}_1, \ldots, \mathbf{b}_k]$ for Λ such that $\mathbf{b}_i = \mathbf{v}$ for some $i \in [k]$. By swapping $\mathbf{b}_1$ and $\mathbf{b}_i$ if $i > 1$, we obtain the corollary. □

Given a lattice Λ, there is always a basis $\mathbf{B}$ for a *sublattice* of Λ such that for all $i \in [\mathrm{rk}(\Lambda)]$, we have $\mathbf{b}_i \in \Lambda$ and $\|\mathbf{b}_i\| = \lambda_i(\Lambda)$, but this is not always a basis for the lattice Λ itself. For example, given the lattice Λ in Example 2.13, such a basis is $\mathbf{B} = \mathbf{I}_n(\mathbb{Z})$, which is a basis for the sublattice $\mathbb{Z}^n \subset \Lambda$ whenever $n > 4$ but not for Λ itself.

Lemma 2.23 ([MG02, Lem. 7.1]). *There exists a polynomial-time algorithm that, on input a basis* $\mathbf{B} \in \mathbb{R}^{n\times k}$ *of a lattice* Λ *and* k *linearly independent lattice vectors* $\mathbf{S} \subset \mathcal{L}(\mathbf{B})$ *such that* $\|\mathbf{s}_1\| \leq \|\mathbf{s}_2\| \leq \cdots \leq \|\mathbf{s}_k\|$ *holds, outputs a basis* $\mathbf{R}$ *for* Λ *such that* $\|\mathbf{r}_i\| \leq \|\mathbf{s}_i\| \cdot \max\{1, \sqrt{i}/2\}$ *holds for all* $i = 1, \ldots, k$.

Corollary 2.24 ([MG02, Cor. 7.2]). *Given a lattice* $\Lambda \subset \mathbb{R}^n$*, there exists a basis* $\mathbf{B}$ *for* Λ *such that* $\|\mathbf{b}_i\| \leq \lambda_i(\Lambda) \cdot \max\{1, \sqrt{i}/2\}$ *holds for all* $i = 1, \ldots, \mathrm{rk}(\Lambda)$.

There are some lists of vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m$ that do not form a generating set. Namely, if such a list satisfies an $\mathbb{R}$-linear relation, it may not generate a discrete set. For example, the set $\mathbb{Z} + \sqrt{2}\mathbb{Z}$ is dense in $\mathbb{R}$ and hence not a lattice.

There is an efficient algorithm that, upon input a generating set for some lattice Λ, outputs a basis for Λ [MG02, Sec. 2.1].

Sublattices and basis transformations

If $\mathbf{B}$ is a basis for a rank-k lattice Λ, and $\mathbf{M} \in \mathbb{Z}^{k\times m}$ is a basis with $m \in [k]$, then $\mathbf{BM}$ is a basis for the sublattice $\mathcal{L}(\mathbf{BM})$, which is of rank m.

Two lattices $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{R}^{n\times k}$ generate the same lattice if and only if there exists a unimodular matrix $\mathbf{U} \in \mathrm{GL}_k(\mathbb{Z})$ such that $\mathbf{B}_2 = \mathbf{B}_1\mathbf{U}$ holds. This is an easy consequence of the above because a matrix $\mathbf{U}$ is invertible over the integers whenever $\det(\mathbf{U}) = \pm 1$.

Given a basis $\mathbf{B}$ for Λ, we have $\det(\Lambda) = \sqrt{\det(\mathbf{B}^\mathsf{T}\mathbf{B})}$, and if Λ is full-rank, $\det(\Lambda) = |\det \mathbf{B}|$. The quantity $\det(\mathbf{B}^\mathsf{T}\mathbf{B})$ is invariant under $\mathrm{GL}_n(\mathbb{Z})$. Namely, because unimodular matrices have determinant ± 1, we have $\det\Big((\mathbf{B}_1\mathbf{U})^\mathsf{T} \cdot \mathbf{B}_1\mathbf{U}\Big) = (\det \mathbf{U})^2 \cdot \det\Big(\mathbf{B}_1^\mathsf{T} \cdot \mathbf{B}_1\Big) = \det(\mathbf{B}_1^\mathsf{T}\mathbf{B}_1)$.

The group $\mathrm{GL}_k(\mathbb{Z})$ is of infinite cardinality when $k \geq 2$, and thus, any lattice of rank 2 and above, admits an infinite number of bases.

Orthogonal transformations

Given an orthogonal transformation $\mathbf{O} \in \mathrm{O}_n(\mathbb{R})$ and a basis $\mathbf{B} \in \mathbb{R}^{n\times k}$ for $\Lambda \subset \mathbb{R}^n$, then $\mathbf{OB}$ is a basis for the rotated lattice $\mathbf{O}\Lambda = \{\mathbf{Ov} : \mathbf{v} \in \Lambda\}$.

Remark 2.25. Technically, an orthogonal transformation is a combination of a rotation (having determinant 1) and a reflection. However, for ease

of writing, we may say 'rotation' or 'isometry' to mean an orthogonal transformation.

Definition 2.26 (isomorphic lattice)**.** A lattice Λ is *isomorphic* to another lattice Λ' if there exists an orthogonal transformation $\mathbf{O} \in \mathrm{O}_n(\mathbb{R})$ such that $\Lambda' = \mathbf{O} \cdot \Lambda$ holds.

2

Lattice isomorphism is an equivalence relation as $\mathrm{O}_n(\mathbb{R})$ is a group. We choose to call such a map an isomorphism as it preserves the inner product space, i.e., for all $\mathbf{x}, \mathbf{y} \in \Lambda$ we have

$$\langle \mathbf{Ox}, \mathbf{Oy} \rangle = \mathbf{x}^\mathsf{T}\mathbf{O}^\mathsf{T}\mathbf{Oy} = \mathbf{x}^\mathsf{T}\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle \quad .$$

Recovering the orthogonal transformation is the Lattice Isomorphism Problem (LIP).

Problem 2.27 (Lattice Isomorphism Problem, LIP)**.** Let Λ_1, Λ_2 be two lattices that are isomorphic to one another. Given on input a basis $\mathbf{B}_1$ for Λ_1 and a basis $\mathbf{B}_2$ for Λ_2, find $\mathbf{O} \in \mathrm{O}_n(\mathbb{R})$ satisfying

$$\Lambda_2 = \mathbf{O} \cdot \Lambda_1 \quad .$$

Note that an orthogonal transformation $\mathbf{O}$ satisfies above equation if and only if there exists $\mathbf{U} \in \mathrm{GL}_k(\mathbb{Z})$ such that $\mathbf{B}_2 = \mathbf{OB}_1\mathbf{U}$.

Automorphisms

A lattice automorphism is an isomorphism from a lattice to itself.

Definition 2.28 (lattice automorphism)**.** Let $\Lambda \subset \mathbb{R}^n$ be a lattice. A *lattice automorphism of* Λ is an orthogonal matrix $\mathbf{O} \in \mathrm{O}_n(\mathbb{R})$ such that $\mathbf{O}\Lambda = \Lambda$ holds. The *group of automorphisms* of Λ is denoted by $\mathrm{O}(\Lambda)$ and consists of all lattice automorphisms of Λ.

Given a basis $\mathbf{B}$ for Λ, any automorphism provides a unimodular matrix $\mathbf{U} \in \mathrm{GL}_k(\mathbb{Z})$ such that $\mathbf{OB} = \mathbf{BU}$, since $\mathbf{OB}$ is another basis for Λ.

Any lattice automorphism group contains the 'trivial automorphisms' $\mathbf{I}_n(\mathbb{R})$ and $-\mathbf{I}_n(\mathbb{R})$.

Remark 2.29. Suppose Λ_1 is isomorphic to the lattice $\Lambda_2 = \mathbf{O}_{12}\Lambda_1$ for some $\mathbf{O}_{12} \in \mathrm{O}_n(\mathbb{R})$. First, there is a group isomorphism, $\mathrm{O}(\Lambda_1) \cong \mathrm{O}(\Lambda_2)$,

where a group isomorphism is given by $\mathbf{O}_1 \mapsto \mathbf{O}_{12}\mathbf{O}_1\mathbf{O}_{12}^{\mathsf{T}}$. Second, there is a bijection

$$\mathrm{O}(\Lambda_1) \to \{\mathbf{O} \in \mathrm{O}_n(\mathbb{R}) \mid \mathbf{O}\Lambda_1 = \Lambda_2\}\,, \quad \mathbf{O}_1 \mapsto \mathbf{O}_{12}\mathbf{O}_1 \ ,$$

where the image is the set of solutions to Problem 2.27. Hence, a solution to LIP is only unique up to right multiplication by an automorphism of Λ_1, and, similarly, unique up to left multiplication by an automorphism of Λ_2.

Example 2.30. The automorphisms of the *integer lattice*, $\mathbb{Z}^n$, are precisely the *signed permutations*: a permutation of the coordinates followed by a possible sign change on each coordinate independently. Abstractly we have

$$\mathrm{O}(\mathbb{Z}^n) \cong \{\pm 1\}^n \rtimes S_n \ ,$$

and the group contains $2^n \cdot n!$ elements.

Normal forms

There are two normal forms of interest: Hermite normal form and Smith normal form.

The Hermite normal form is useful for sublattices of $\mathbb{Z}^n$, and by scaling, to any lattice $\Lambda \subset \mathbb{Q}^n$.

Definition 2.31 (HNF)**.** A nonzero matrix $\mathbf{A} \in \mathbb{Z}^{n\times k}$ is said to be *in column-style Hermite normal form* (HNF) if there exist so-called 'pivots' $(p_i)_{i=1}^r$ for some $r \le k$ such that $1 \le p_1 < p_2 < \cdots < p_r \le n$ and:

(1) if $j > r$ then $\mathbf{A}_{ij} = 0$
(2) for all $j \le r$ and $i < p_j$, we have $\mathbf{A}_{ij} = 0$,
(3) $\forall j \in [r]\colon \mathbf{A}_{p_j,j} > 0$, and
(4) $\forall k < j \le r\colon 0 \le \mathbf{A}_{p_j,k} < \mathbf{A}_{p_j,j}$.

The Hermite normal form can be generalised from $\mathbb{Z}$ to a principal ideal domain [SL96].

Example 2.32. A square full-rank matrix $\mathbf{A}$ is in HNF, if $r = k$, $\forall i \in$

$[k]\colon p_i = i$, and it is of the form

$$\begin{pmatrix} \mathbf{A}_{11} & 0 & \cdots & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \mathbf{A}_{k1} & \mathbf{A}_{k2} & \cdots & \mathbf{A}_{kk} \end{pmatrix},$$

Proposition 2.33. *For any matrix $\mathbf{A} \in \mathbb{Z}^{n\times k}$, there exists a matrix $\mathbf{U} \in \mathrm{GL}_k(\mathbb{Z})$ such that $\mathbf{AU}$ is in HNF. If $\mathbf{A}$ is full-rank, then $\mathbf{U}$ is unique.*

Corollary 2.34. *Every lattice $\Lambda \subseteq \mathbb{Z}^n$ has a unique basis $\mathbf{B} \in \mathbb{Z}^{n\times k}$ in HNF.*

By reducing a matrix $\mathbf{A}$ using the action of $\mathrm{GL}_n(\mathbb{Z})$ on both left and right of $\mathbf{A}$, we obtain the Smith normal form.

Definition 2.35 (SNF)**.** A matrix $\mathbf{A} \in \mathbb{Z}^{n\times k}$ is in *Smith normal form* (SNF) if $\mathbf{A}_{11} \mid \mathbf{A}_{22} \mid \cdots \mid \mathbf{A}_{rr}$ for some $r \le k$ is a divisor chain of positive integers, and $\mathbf{A}$ is zero except for $\mathbf{A}_{ii}$ with $i \le r$,

Proposition 2.36. *For any matrix $\mathbf{A} \in \mathbb{Z}^{n\times k}$, there exist $\mathbf{U} \in \mathrm{GL}_n(\mathbb{Z})$ and $\mathbf{V} \in \mathrm{GL}_k(\mathbb{Z})$ such that $\mathbf{UAV}$ is in SNF.*

2.2.2 Random lattice

To use lattices in cryptography, we ideally want to generate lattices using randomness. There are uncountably many full-rank lattices of volume 1 in any dimension $n \ge 2$.

As was shown by Siegel [Sie45], there is a natural Haar measure μ_n on the collection of full-rank, unit-volume lattices, which may be identified by the set $\mathcal{X}_n = \mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$. Hence, there exists a probability distribution to sample random lattices according to the so-called invariant distribution μ_n.

Although such a distribution may have elegant theoretical properties, it is practically desirable to have a large enough distribution that is efficiently sampleable from. Moreover, integer or rational lattices may be easier to work with. In this direction, Goldstein and Mayer give an algorithm for generating dimension-n unit-volume lattices [GM03], which is Algorithm 2.1 below. We will explain shortly after Theorem 2.37 that

Algorithm 2.1. SampleRandom: sampler for Goldstein–Mayer lattices

Input: dimension n
Output: unit-volume lattice in dimension n

1: sample a large integer N
2: sample uniformly from the finite collection of sublattices $\Lambda \subseteq \mathbb{Z}^n$ with volume N
3: **return** $N^{-1/n} \cdot \Lambda$

Step 2 of Algorithm 2.1 is easily implemented if one only generates prime numbers in Step 1, while it may be hard for composite N.

For sufficiently large N, the output of Algorithm 2.1 is closely distributed to Siegel's definition of random lattices.

Theorem 2.37 ([GM03, Thm. 2.1]). *Let $A \subseteq \mathcal{X}_n$ be any measurable set such that its boundary has zero measure with respect to μ_n, and let $\mathcal{L}_{n,N}$ be the collection of full-rank sublattices of $\mathbb{Z}^n$ of volume N. Then,*

$$\lim_{N\to\infty} \frac{1}{|\mathcal{L}_{n,N}|} \cdot \sum_{\Lambda\in\mathcal{L}_{n,N}} \mathbf{1}_A(N^{-1/n}\Lambda) = \mu(A) \ .$$

For prime q, by Corollary 2.34, the collection $\mathcal{L}_{n,q}$ consists of lattices generated by a basis of the form

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ x_1 & \cdots & x_{i-1} & q & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} ,$$

for some $i \in [n]$ and $x_1, \ldots, x_{i-1} \in \{0, 1, \ldots, q-1\}$.

For a particular $i \in [n]$, there are q^{i-1} possible options for $x_1, \ldots, x_{i-1}$. Hence, for large q, most lattices in $\mathcal{L}_{n,q}$ are generated by a basis of the

form

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ x_1 & \cdots & x_{n-1} & q \end{pmatrix} . \tag{2.4}$$

2

In this thesis, we are interested in q-ary lattices.

Definition 2.38 (q-ary lattice)**.** Let $q \geq 2$ be a prime number. A *q-ary lattice* is of the form

$$\mathcal{L}_q(\mathbf{A}) = \{\mathbf{x} \in \mathbb{Z}^m \mid \exists \mathbf{y} \in \mathbb{Z}^n \colon \mathbf{x} \equiv \mathbf{A}\mathbf{y} \pmod q\} \ ,$$

where $n, m \in \mathbb{Z}_{\geq 1}$ and $\mathbf{A} \in (\mathbb{Z}/q\mathbb{Z})^{m \times n}$.

The distribution of *random q-ary lattices* of dimensions (m, n) is given by $\mathcal{L}_q(\mathbf{A})$ where $\mathbf{A} \leftarrow \mathrm{U}((\mathbb{Z}/q\mathbb{Z})^{m \times n})$.

Given $n \leq m$, a sample $\mathbf{A} \leftarrow \mathrm{U}((\mathbb{Z}/q\mathbb{Z})^{m \times n})$ is full-rank with probability approximately $1/q$, in which case we have $[\mathcal{L}_q(\mathbf{A}) : q\mathbb{Z}^m] = q^n$ and $[\mathbb{Z}^m : \mathcal{L}_q(\mathbf{A})] = q^{m-n}$. Intuitively, every column of $\mathbf{A}$ increases the density of $\mathcal{L}_q(\mathbf{A})$ by a factor q.

Alternatively, one may consider the so-called *small-secret embedding* [BG14], i.e.,

$$\mathcal{L}'_q(\mathbf{A}) = \left\{ \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \in \mathbb{Z}^{m+n} \;\middle|\; \mathbf{x}_1 \equiv \mathbf{A} \cdot \mathbf{x}_2 \pmod q \right\} = \mathcal{L}_q\left(\begin{bmatrix} \mathbf{A} \\ \mathbf{I}_n \end{bmatrix} \right) .$$

A q-ary lattice of the form $\mathcal{L}_q(\mathbf{A})$ is equivalent to a small-secret embedding if $\mathbf{A} \in (\mathbb{Z}/q\mathbb{Z})^{m \times n}$ is full-rank. Namely, in this case, by permuting the rows of $\mathbf{A}$, we may obtain a matrix $\mathbf{A}' = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}$ such that the square matrix $\mathbf{A}_2$ is invertible, which in turn shows:

$$\mathcal{L}'_q\left(\mathbf{A}_1\mathbf{A}_2^{-1}\right) = \mathcal{L}_q(\mathbf{A}') \cong \mathcal{L}_q(\mathbf{A}) \ .$$

The class of *orthogonal q-ary lattices* of the form

$$\mathcal{L}_q^{\perp}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{Z}^m \mid \mathbf{x} \cdot \mathbf{A} \equiv \mathbf{0} \pmod q\} \ , \tag{2.5}$$

are related to q-ary lattices by the so-called notion of duality, which will be explained in Section 2.3.

Ajtai proves that this class exhibits a *worst case to average case reduction* for some lattice problems [Ajt96, Ajt99]. In other words, if there exists a polynomial time algorithm $\mathcal{A}$ that solves a lattice problem upon input a random lattice sampled from this class, then there is a polynomial time algorithm $\mathcal{B}$ that solves that lattice problem upon input any lattice of said distribution.

For more information on random lattices, see [AEN19].

Gaussian heuristic

Siegel's mean value theorem [Sie45] implies that for any bounded, compact set $S \subseteq \mathbb{R}^n$, we have

$$\mathop{\mathbb{E}}_{\Lambda \sim \mu_n} [|S \cap \Lambda \setminus \{\mathbf{0}\}|] = \mathrm{Vol}_n(S) \ . \tag{2.6}$$

Definition 2.39. For $n \in \mathbb{Z}_{\geq 1}$, let us denote the radius of the unit-volume ball by

$$\mathrm{GH}(n) = \mathrm{Vol}_n(\mathcal{B}^n)^{-1/n} \ .$$

By considering $S = \mathrm{GH}(n)\mathcal{B}^n$, we expect one nonzero lattice point in S (and its negation) by Equation (2.6). In addition, $\lambda_1(\Lambda)$ is very concentrated around $\mathrm{GH}(n)$ [Rog56, Söd11].

Theorem 2.40 ([LN25, Thm. 6])**.** *A Haar-random lattice $\Lambda \leftarrow \mu_n$ satisfies*

$$\left| \frac{\lambda_1(\Lambda)}{\mathrm{GH}(n)} - 1 \right| \leq \frac{\log \log n}{n} \ ,$$

with probability at least $1 - o(1)$ as $n \to \infty$.

These concentration results allow us to approximate $\lambda_1(\Lambda)$ for many lattices encountered in cryptography (although not all), even when these lattices are not entirely random. Given a lattice Λ and a bounded set $S \subset \mathbb{R}^n$, the *Gaussian heuristic* states

$$|\Lambda \cap S| \approx \mathrm{Vol}_n(S) / \det(\Lambda) \ ,$$

which is a heuristic version of Equation (2.6). The Gaussian heuristic leads to the following two heuristics regarding the length of short vectors.

Heuristic 2.41. *A random lattice $\Lambda \subset \mathbb{R}^n$, has $\lambda_1(\Lambda) = \mathrm{GH}(n) \cdot \det(\Lambda)^{1/n}$.*

Note Theorem 2.16 states $\lambda_1(\Lambda) \leq 2 \cdot \mathrm{GH}(n) \cdot \det(\Lambda)^{1/n}$. Moreover, applying Stirling's approximation on Equation (2.1) leads to

$$\mathrm{GH}(n) \sim \sqrt{n/(2\pi e)} \qquad (\text{as } n \to \infty).$$

2

It is known that $\mathrm{GH}(n) \approx \sqrt{n/(2\pi e)}$ for $n \geq 50$ [Che13].

Heuristic 2.42. *Given a random lattice $\Lambda \subset \mathbb{R}^n$ and some $r > 1$, we have*

$$\left|\left\{\mathbf{v} \in \Lambda \;\middle|\; \|\mathbf{v}\| \leq r \cdot \mathrm{GH}(n)\det(\Lambda)^{1/n}\right\}\right| \approx r^n.$$

In particular, the ith shortest lattice point $\mathbf{v}$ has a length of approximately $\|\mathbf{v}\| \approx \mathrm{GH}(n)\det(\Lambda)^{1/n} i^{1/n}$.

There exist constructions of lattices where the Gaussian heuristic does not hold. However, for our applications, i.e., random q-ary lattices with sufficiently large q or projected sublattices considered during lattice reduction, Heuristic 2.41 accurately describes $\lambda_1(\Lambda)$.

2.2.3 Hard problems

In this section, we state the lattice problems that are deemed to be computationally hard to solve. The fundamental hard problem in lattice-based cryptography is that of finding the shortest vector.

Our problem statements say input and/or output contain lattices and/or lattice points. Computationally, these are actually described respectively by a rational basis and/or a rational coefficient vector expressing a point as a linear combination of basis vectors of certain precision.

Problem 2.43 (shortest vector, SVP)**.** Given upon input a lattice $\Lambda \subset \mathbb{R}^n$, output a vector $\mathbf{v} \in \Lambda$ such that $\|\mathbf{v}\| = \lambda_1(\Lambda)$.

In practice, the lattice is specified by a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$, and the goal is to find coefficients $\mathbf{c}_1, \ldots, \mathbf{c}_k$ such that $\mathbf{v} = \sum_{i=1}^{k} c_i \mathbf{b}_i$ is a shortest vector.

Problem 2.44 (closest vector, CVP)**.** Given upon input a lattice Λ and a vector $\mathbf{t} \in \mathbb{R}^n$, find a vector $\mathbf{v} \in \Lambda$ such that $\|\mathbf{v} - \mathbf{t}\| \leq \|\mathbf{w} - \mathbf{t}\|$ for all $\mathbf{w} \in \Lambda$.

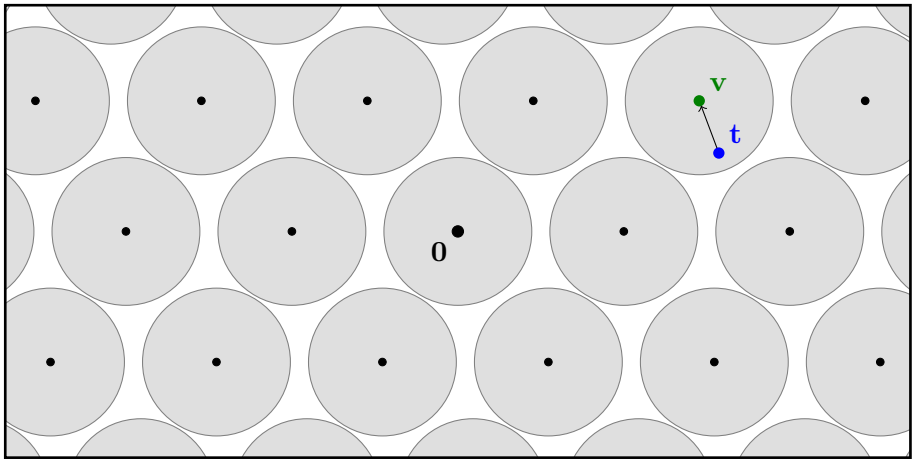


Figure 2.2: Search bounded distance decoding with $\alpha = 0.49$

Van Emde Boas [vEB81] proved NP-completeness of the decisional variants of two computational problems: (1) SVP with respect to ℓ^∞-norm, and (2) CVP, and conjectured SVP with respect to ℓ^2-norm is NP-complete as well. Remark that he phrased (1) and (2) the 'closest vector' and 'nearest vector' problem, respectively, which may be confusing with modern terminology. Ajtai [Ajt98] showed that SVP with respect to ℓ^2-norm is NP-hard for randomised reductions.

Bounded distance decoding

The *bounded distance decoding problem* (BDD), visualised in Figure 2.2, is a variant of CVP in which the distance from the target $\mathbf{t}$ to the lattice is guaranteed to be bounded in norm. The following computational problems are considered hard for specific parameters.

Problem 2.45 (search-BDD$_\alpha$, lattice form)**.** Given upon input a parameter $\alpha > 0$, a lattice $\Lambda \subset \mathbb{R}^n$ and a target $\mathbf{t} \in \mathbb{R}^n$, such that $\mathbf{t} = \mathbf{v} + \mathbf{e}$ holds for some $\mathbf{v} \in \Lambda$ and $\|\mathbf{e}\| \leq \alpha\lambda_1(\Lambda)$, output $\mathbf{v}$.

By considering $\mathbf{t}$ modulo the lattice and demanding $\mathbf{t} - \mathbf{v}$ as a result, we get the syndrome form.

Problem 2.46 (search-BDD$_\alpha$, syndrome form)**.** Given upon input a parameter $\alpha > 0$, a lattice $\Lambda \subset \mathbb{R}^n$ and a target $\bar{\mathbf{t}} = \mathbf{e} + \Lambda \in \mathbb{R}^n/\Lambda$, such that $\|\mathbf{e}\| \leq \alpha\lambda_1(\Lambda)$ holds, output $\mathbf{e}$.

Concretely, in the syndrome form of search-BDD, one is given a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ of Λ, and a target $\mathbf{t} = \mathbf{B} \cdot \mathbf{c}$ that is expressed in terms of $\mathbf{B}$ using coefficients $\mathbf{c} \in \mathbb{R}^k$ in the interval $[0, 1)$.

Note it is difficult to compute $\lambda_1(\Lambda)$ for an arbitrary lattice by the hardness of SVP, so it is hard to verify a candidate $\mathbf{v} \in \Lambda$, i.e., to check $\|\mathbf{t} - \mathbf{v}\| \leq \alpha\lambda_1(\Lambda)$. For random lattices, however, this is possible as we have $\lambda_1(\Lambda) \approx \mathrm{GH}(\Lambda)$ by Heuristic 2.41.

Remark 2.47. When search-BDD is instantiated with $\alpha < \frac{1}{2}$, it is guaranteed that there is a unique lattice point close enough to $\mathbf{t}$. Indeed, suppose $\mathbf{v}_1 \in \Lambda$ and $\mathbf{v}_2 \in \Lambda$ satisfy $\|\mathbf{v}_i - \mathbf{t}\| < \lambda_1(\Lambda)/2$ $(i = 1, 2)$. By the triangle inequality, we have $\|\mathbf{v}_1 - \mathbf{v}_2\| \leq \|\mathbf{v}_1 - \mathbf{t}\| + \|\mathbf{t} - \mathbf{v}_2\| < \lambda_1(\Lambda)$, so $\mathbf{v}_1 = \mathbf{v}_2$.

The following heuristic shows that there is still a unique lattice point close enough with high probability in average case search-BDD$_\alpha$ instances using $\alpha < 1$, a random lattice and $\mathbf{e} \leftarrow \mathrm{U}(\alpha\,\mathrm{GH}(n)\mathcal{B}^n)$.

Heuristic Claim 2.48. *Let $\Lambda \subset \mathbb{R}^n$ be a random unit-volume lattice, $\alpha \in (0, 1)$, and $R = \alpha\,\mathrm{GH}(n)$. The probability that a target $\mathbf{t} \leftarrow \mathrm{U}(R\mathcal{B}^n)$ is at a distance of at most R from some* nonzero *lattice point $\mathbf{v} \in \Lambda$ is at most $O(n\sqrt{n})\alpha^n$.*

This claim can be justified using Heuristic 2.42, and an upper bound by Micciancio and Voulgaris on spherical domes [MV10, Lem. 4.1].

Heuristic Justification. Note that only lattice points $\mathbf{v} \in \Lambda \setminus \{\mathbf{0}\}$ are relevant with $\|\mathbf{v}\| \leq 2R = 2\alpha\,\mathrm{GH}(n)$ by the triangle inequality. For such a $\mathbf{v} \in \Lambda \setminus \{\mathbf{0}\}$, we are interested in $\mathrm{Vol}_n\left(R\mathcal{B}^n \cap (\mathbf{v} + R\mathcal{B}^n)\right)$, which is twice the volume of the spherical dome $\{\mathbf{t} \in R\mathcal{B}^n \,|\, \langle \mathbf{t}, \mathbf{v}\rangle \geq \frac{1}{2}\|\mathbf{v}\|^2\}$. Set $\alpha = \|\mathbf{v}\|/2R$. This spherical dome is contained in a cylinder with base $R\sqrt{1-\alpha^2} \cdot \mathcal{B}^{n-1}$ and height $R(1-\alpha)$, which has volume at most $R^n(1-\alpha^2)^{n/2}\,\mathrm{Vol}_{n-1}(\mathcal{B}^{n-1})$. One can show that $\mathrm{Vol}_{n-1}(\mathcal{B}^{n-1}) \leq \frac{\sqrt{en}}{2}\,\mathrm{Vol}_n(\mathcal{B}^n)$ holds, which implies

$$\mathrm{Vol}_n(R\mathcal{B}^n \cap (\mathbf{v} + R\mathcal{B}^n)) \leq O(\sqrt{n})\alpha^n\left(1-\alpha^2\right)^{n/2}.$$

Heuristic 2.42 predicts approximately ℓ^n lattice points in a ball of radius $\ell\,\mathrm{GH}(n)$. By using this estimate for $\ell \in (1, 2\alpha)$, the volume of all

the spherical domes is roughly

$$\int_1^{2\alpha} n\ell^{n-1} O(\sqrt{n})\alpha^n \left(1 - \frac{\ell^2}{4\alpha^2}\right)^{n/2} \mathrm{d}\ell \leq O(n\sqrt{n})\alpha^n \int_1^{2\alpha} \left(\ell^2 - \frac{\ell^4}{4\alpha^2}\right)^{n/2} \mathrm{d}\ell.$$

The integrand reaches its maximum α^n at $\ell = \sqrt{2}\alpha$, so the above equation can be bounded from above by $O(n\sqrt{n})\alpha^{2n}(2\alpha - 1)$. For the desired probability, we consider the ratio of volumes, which is at most $O(n\sqrt{n})\alpha^{2n}/\operatorname{Vol}_n(R\mathcal{B}^n) = O(n\sqrt{n})\alpha^n$. △

Alternatively, there is also a decisional version of BDD.

Problem 2.49 (decision-BDD)**.** Given upon input a parameter $\alpha > 0$, a lattice $\Lambda \subset \mathbb{R}^n$ and a target $\mathbf{t} \in \mathbb{R}^n$, output 'yes' if there exists $\mathbf{v} \in \Lambda$ such that $\|\mathbf{t} - \mathbf{v}\| \leq \alpha\lambda_1(\Lambda)$, and otherwise 'no'.

In this thesis, we are interested in average case variants of search-BDD and decision-BDD. Here, one should output 'yes' on input a target $\mathbf{t}$ drawn from a *BDD distribution* χ on $\mathbb{R}^n$ such that almost all samples have a norm around $\alpha\lambda_1(\Lambda)$ for some $\alpha > 0$. At the same time, one should output 'no' on input a target $\mathbf{t} \leftarrow \mathrm{U}(\mathbb{R}^n/\Lambda)$ drawn from the uniform distribution. Since the BDD and uniform distribution may overlap, average case BDD cannot always be solved. Rather, a successful algorithm both rarely outputs 'yes' when $\mathbf{t}$ is sampled uniformly (*false positive*), and rarely outputs 'no' when $\mathbf{t}$ is sampled from χ (*false negative*). The statistical distance between the BDD and uniform distribution is large if α is small (e.g., $\alpha = \frac{1}{2}$) and n is large, in which case a successful algorithm may exist, albeit not necessarily efficient.

Problem 2.50 (average case search-BDD$_\alpha$)**.** Let $\alpha > 0$ and let χ be a distribution on $\mathbb{R}^n$ such that $\|\mathbf{t}\| \approx \alpha\lambda_1(\Lambda)$ is likely if $\mathbf{t} \leftarrow \chi$. Given upon input α, a lattice $\Lambda \subset \mathbb{R}^n$ and a target $\mathbf{t} \in \mathbb{R}^n$ such that $\mathbf{t} = \mathbf{v} + \mathbf{e}$ holds for some $\mathbf{v} \in \Lambda$, and $\mathbf{e} \leftarrow \chi$, output $\mathbf{v}$.

Problem 2.51 (average case decision-BDD$_\alpha$)**.** Let $\alpha > 0$ and let χ be a distribution on $\mathbb{R}^n$ such that $\|\mathbf{t}\| \approx \alpha\lambda_1(\Lambda)$ is likely if $\mathbf{t} \leftarrow \chi$. Given upon input α, a lattice $\Lambda \subset \mathbb{R}^n$ and a target $\mathbf{t} \in \mathbb{R}^n$ sampled either from $\chi \pmod{\Lambda}$ or $\mathrm{U}(\mathbb{R}^n/\Lambda)$, output 'yes' in the former case and 'no' in the latter case.

Usually, in search-BDD and decision-BDD, the error distribution χ is a uniform distribution on a sphere or ball of some radius R, or a (discrete, see Section 2.4) Gaussian with some parameter σ. The hardness of search-BDD and decision-BDD then depends on the radius R or parameter σ, i.e., the bigger R or σ is, the larger the average error norm is and thus the harder the problem instances are.

2

Unique shortest vector

In certain lattices, the shortest vector is of much smaller norm than other vectors that are not parallel to it, in which case SVP is not harder than for a general lattice.

Problem 2.52 (unique shortest vector, $\mathrm{UniqueSVP}_\gamma$)**.** Given upon input $\gamma \in \mathbb{R}_{\geq 1}$ and a lattice Λ with the promise $\lambda_2(\Lambda) > \gamma\lambda_1(\Lambda)$, output a vector $\mathbf{v} \in \Lambda$ such that $\|\mathbf{v}\| = \lambda_1(\Lambda)$.

Lyubashevsky and Micciancio [LM09] show two polynomial-time reductions: (1) from $\mathrm{BDD}_{1/2\gamma}$ to $\mathrm{UniqueSVP}_\gamma$ given any $\gamma \geq 1$, and (2) from $\mathrm{UniqueSVP}_\gamma$ to $\mathrm{BDD}_{1/\gamma}$ given any γ polynomial in n. Hence, given $\alpha < \frac{1}{2}$, BDD_α is essentially as hard as $\mathrm{UniqueSVP}_\gamma$ for some $\gamma \in [\frac{1}{2\alpha}, \frac{1}{\alpha}]$.

Learning with errors

It is common in cryptography to base security on the hardness of the learning with errors problem, which was introduced by Regev in 2005 [Reg09]. Learning with errors is basically the BDD problem using q-ary lattices.

Problem 2.53 (learning with errors, LWE)**.** Let $\mathbf{A} \leftarrow \mathrm{U}((\mathbb{Z}/q\mathbb{Z})^{m\times n})$ be a random q-ary matrix, χ_s be a secret distribution on $(\mathbb{Z}/q\mathbb{Z})^n$, and χ_e an error distribution on $\mathbb{Z}^m$. Given upon input the pair $(\mathbf{A}, \mathbf{b})$ where $\mathbf{b} \equiv \mathbf{A}\mathbf{s} + \mathbf{e} \pmod q$ is computed after sampling $\mathbf{s} \leftarrow \chi_\mathrm{s}$ and $\mathbf{e} \leftarrow \chi_\mathrm{e}$, output the vector $\mathbf{s}$.

A standard choice for the error distribution is the discrete Gaussian (see Section 2.4) $\chi_\mathrm{e} = \mathcal{D}_{\mathbb{Z}^m,\sigma}$ with parameter $\sigma > 0$. In a worst-case analysis of LWE, one determines, given any $\mathbf{s} \in (\mathbb{Z}/q\mathbb{Z})^n$, the difficulty of recovering $\mathbf{s}$. On the other hand, in cryptographic constructions, such as ML-KEM and ML-DSA, the secret $\mathbf{s}$ always has small norm, for example by taking $\chi_\mathrm{s} = \chi_\mathrm{e}$ and/or restricting secret coefficients to the set $\{-3, -2, \ldots, 2, 3\}$, which is the case for ML-KEM and ML-DSA.

2.2.4 Projection

First, we recall some linear algebra and discuss Gram–Schmidt orthogonalisation and QR decomposition. These give rise to projected sublattices, which are used in basis reduction to map a high dimensional lattice to a lower dimensional one.

Definition 2.54 (orthogonality)**.** A vector $\mathbf{v}$ is *orthogonal to a vector* $\mathbf{w} \in \mathbb{R}^n$ if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$, and *orthogonal to a subspace* $V \subseteq \mathbb{R}^n$ if $\forall \mathbf{w} \in V\colon \langle \mathbf{v}, \mathbf{w} \rangle = 0$.

Given a subspace $V \subseteq \mathbb{R}^n$, the *orthogonal subspace* $V^\perp$ consists of all $\mathbf{v} \in \mathbb{R}^n$ orthogonal to V. Moreover, $\pi_V\colon \mathbb{R}^n \to V$ is the projection onto V, and $\pi_V^\perp = \pi_{V^\perp}$ is the projection orthogonally away from V.

Note for all $\mathbf{x} \in \mathbb{R}^n$, we have $\mathbf{x} = \pi_V(\mathbf{x}) + \pi_V^\perp(\mathbf{x})$. Moreover, as $\pi_V(\mathbf{x})$ is orthogonal to $\pi_V^\perp(\mathbf{x})$, by Pythagoras' theorem, we have

$$\|\mathbf{x}\|^2 = \|\pi_V(\mathbf{x})\|^2 + \|\pi_V^\perp(\mathbf{x})\|^2 \ , \tag{2.7}$$

so in particular $\|\pi_V(\mathbf{x})\| \leq \|\mathbf{x}\|$, and $\|\pi_V^\perp(\mathbf{x})\| \leq \|\mathbf{x}\|$.

Given a basis, we are interested in the subspaces generated by subarrays of a basis $\mathbf{B}$, and the orthogonal subspaces of these.

Gram–Schmidt orthogonalisation

Definition 2.55 (basis projection)**.** Given a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ and $\ell \in [k]$, the projection orthogonally away from $[\mathbf{b}_1, \ldots, \mathbf{b}_{\ell-1}]$ is denoted by

$$\pi_{\mathbf{B},\ell} = \pi^\perp_{\mathrm{span}(\mathbf{b}_1,\ldots,\mathbf{b}_{\ell-1})} \ ,$$

where we may omit $\mathbf{B}$ if it is clear from context. In particular, we have $\pi_1(\mathbf{x}) = \mathbf{x}$.

Definition 2.56 (Gram–Schmidt)**.** Given a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$, the ith *Gram–Schmidt vector* is $\mathbf{b}_i^* = \pi_i(\mathbf{b}_i)$ for $i \in [k]$. The *Gram–Schmidt basis* is given by $\mathbf{B}^* = [\mathbf{b}_1^*, \ldots, \mathbf{b}_k^*]$, and the *Gram–Schmidt coefficients* μ_{ij} where $i, j \in [k]$ are given by

$$\mu_{ij} = \langle \mathbf{b}_i^*, \mathbf{b}_j \rangle / \|\mathbf{b}_i^*\|^2 \ .$$

Note $\mathbf{B}^*$ is an orthogonal basis. Moreover, note $\mu_{ij} = 0$ when $j < i$ and $\mu_{ii} = 1$. In other words, writing $\mathbf{M} = (\mu_{ij})_{i,j=1}^k$ shows that $\mathbf{M}$ is

unitriangular. We have $\mathbf{b}_i = \mu_{1i}\mathbf{b}_1^* + \cdots + \mu_{i-1,i}\mathbf{b}_{i-1}^* + \mathbf{b}_i^*$, or in matrix form,

$$\mathbf{B} = \mathbf{B}^* \cdot \mathbf{M} \ .$$

By the above, we get $\det \Lambda = |\det \mathbf{B}| = |\det(\mathbf{B}^*)| = \prod_{i=1}^{n} \|\mathbf{b}_i^*\|$.

2

The *Gram–Schmidt orthogonalisation (GSO)* of $\mathbf{B}$ is the pair $(\mathbf{d}, \mathbf{M})$ where $\mathbf{d} \in \mathbb{R}^k$ satisfies $\mathbf{d}_i = \|\mathbf{b}_i^*\|$ for $i \in [k]$. An approximation of the GSO can be computed, assuming no numerical issues arise, by setting $\mathbf{b}_1^* = \mathbf{b}_1$, and then iteratively computing $\mathbf{b}_j^* = \mathbf{b}_j - \sum_{i<j} \mu_{ij}\mathbf{b}_i^*$ for $j = 2, \ldots, n$ using floating-point numbers with certain precision.

For a basis $\mathbf{B} \in \mathbb{R}^{n\times k}$ of a lattice Λ and $1 \le i \le j \le k$, we may consider the *projected sublattice* $\mathbf{B}_{[i,j]} = [\pi_i(\mathbf{b}_i), \ldots, \pi_i(\mathbf{b}_j)]$. Note $\mathbf{B}_{[i,k]}$ is a basis for the lattice $\pi_i(\Lambda)$.

Lemma 2.57. *Given a basis $\mathbf{B} \in \mathbb{R}^{n\times k}$ of Λ, we have*

$$\lambda_1(\Lambda) \ge \min\{\|\mathbf{b}_1^*\|, \ldots, \|\mathbf{b}_k^*\|\} \ .$$

Proof. Suppose $\mathbf{v} \in \Lambda$ is nonzero, and write $\mathbf{v} = \sum_{i=1}^{k} c_i\mathbf{b}_i$ with $c_i \in \mathbb{Z}$. Let $i \in [k]$ be maximal such that $c_i \neq 0$ holds, and note such i exists as $\mathbf{v}$ is nonzero. Using Equation (2.7) and $\pi_i(\mathbf{v}) = c_i \cdot \pi_i(\mathbf{b}_i)$, we obtain

$$\|\mathbf{v}\| \ge \|\pi_i(\mathbf{v})\| = |c_i| \cdot \|\mathbf{b}_i^*\| \ge \|\mathbf{b}_i^*\| \ . \qquad \square$$

QR decomposition

The QR decomposition [Hig02, Chapter 19] is related to Gram–Schmidt orthogonalisation.

Definition 2.58 (QR decomposition)**.** The *QR decomposition of a basis* $\mathbf{B} \in \mathbb{R}^{n\times k}$ is the pair $(\mathbf{Q}, \mathbf{R})$ with $\mathbf{Q} \in \mathbb{R}^{n\times k}$ and $\mathbf{R} \in \mathbb{R}^{k\times k}$ such that $\mathbf{R}$ is an upper triangular matrix with all diagonal entries in $\mathbb{R}_{>0}$, $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}_k(\mathbb{R})$ and $\mathbf{B} = \mathbf{Q} \cdot \mathbf{R}$. Here $\mathbf{R}$ denotes the *R-factor* of $\mathbf{B}$.

There are four remarks to make.

(1) A QR decomposition is unique per basis.
(2) The matrix $\mathbf{Q}$ can be extended by some $\mathbf{Q}' \in \mathbb{R}^{n\times(n-k)}$ to an orthogonal transformation $[\mathbf{Q} \mid \mathbf{Q}'] \in \mathrm{O}_n(\mathbb{R})$.
(3) Software libraries such as LAPACK and NumPy may return an R-factor, in which some diagonal entries are negative.

(4) The QR decomposition $(\mathbf{Q}, \mathbf{R})$ is related to the GSO $(\mathbf{d}, \mathbf{M})$ by

$$\mathbf{Q} = \left[\frac{\mathbf{b}_1^*}{\|\mathbf{b}_1^*\|} \quad \cdots \quad \frac{\mathbf{b}_k^*}{\|\mathbf{b}_k^*\|} \right], \quad \text{and} \quad \mathbf{R} = \operatorname{diag}(\mathbf{d}) \cdot \mathbf{M} \ ,$$

$$\text{so } \mathbf{R}_{ii} = \|\mathbf{b}_i^*\| \text{ and } \mathbf{R}_{ij} = \langle \mathbf{b}_i^*, \mathbf{b}_j \rangle / \|\mathbf{b}_i^*\| = \mathbf{d}_i \mu_{ij} \text{ for } i, j \in [k].$$

The QR decomposition and GSO can both be computed using at most $\mathcal{O}(nk^2)$ floating-point operations. The QR decomposition, or only the R-factor, can be computed using Householder reflections [Hig02, Chapter 19]. Householder reflections are more numerically stable than the GSO, which is useful when computing floating-point approximations of inner products $\langle \mathbf{b}_i^*, \mathbf{b}_j \rangle$, like will be done in Section 2.5.

Proposition 2.59. *Let $\Lambda \subseteq \mathbb{R}^n$ be a rank-k lattice. Then, there exists a full-rank lattice $\Lambda' \subseteq \mathbb{R}^k$ such that Λ is isomorphic to*

$$\left\{ (v_1, \ldots, v_n)^\mathsf{T} \in \mathbb{R}^n \;\middle|\; (v_1, \ldots, v_k)^\mathsf{T} \in \Lambda', \text{ and } v_{k+1} = \cdots = v_n = 0 \right\} \ .$$

Moreover, given a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ of Λ, then its R-factor is a basis for Λ' satisfying the above.

Proof. The second remark above shows there exists $\mathbf{Q}' \in \mathrm{O}_n(\mathbb{R})$ such that $\mathbf{B} = \mathbf{Q}' \cdot \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$ holds. This is equivalent to ${\mathbf{Q}'}^\mathsf{T} \mathbf{B} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$ and the claim follows. □

2.2.5 Lattice isomorphism

Recalling paragraph § Orthogonal transformations (p. 33), if Λ, Λ' are generated by $\mathbf{B}, \mathbf{B}'$ respectively, then they are isomorphic if there exists an orthogonal transformation $\mathbf{O} \in \mathrm{O}_n(\mathbb{R})$ and a unimodular matrix $\mathbf{U} \in \mathrm{GL}_n(\mathbb{Z})$ such that $\mathbf{O} \cdot \mathbf{B} \cdot \mathbf{U} = \mathbf{B}'$. We can remove the real valued orthogonal transformation by moving to quadratic forms.

Definition 2.60 (quadratic form)**.** A *quadratic form* is a positive definite real symmetric matrix $\mathbf{Q} \in \mathcal{S}_n^{>0}(\mathbb{R})$. Given two quadratic forms $\mathbf{Q}, \mathbf{Q}' \in \mathcal{S}_n^{>0}(\mathbb{R})$, an *isomorphism from $\mathbf{Q}$ to $\mathbf{Q}'$* is a unimodular $\mathbf{U} \in \mathrm{GL}_n(\mathbb{Z})$ such that $\mathbf{U}^\mathsf{T} \cdot \mathbf{Q} \cdot \mathbf{U} = \mathbf{Q}'$ holds. We say $\mathbf{Q}$ and $\mathbf{Q}'$ are *isomorphic* if there exists an isomorphism from $\mathbf{Q}$ to $\mathbf{Q}'$.

For any lattice basis $\mathbf{B}$ the Gram matrix $\mathbf{B}^\mathsf{T}\mathbf{B}$, consisting of all pairwise inner products, is a quadratic form. Conversely, the *Cholesky decomposition* of a quadratic form $\mathbf{Q}$ is the unique basis $\mathbf{R}$ that is an upper triangular matrix with strictly positive diagonal such that $\mathbf{R}^\mathsf{T}\mathbf{R} = \mathbf{Q}$ holds. Note in this section, $\mathbf{Q}$ denotes a quadratic form and is not related to QR decomposition.

We have that two bases generate isomorphic lattices if and only if their Gram matrices are isomorphic. This allows us to restate Problem 2.27.

Problem 2.61 (LIP, restated)**.** Given upon input two isomorphic quadratic forms $\mathbf{Q}, \mathbf{Q}' \in \mathcal{S}_n^{>0}(\mathbb{R})$, output any isomorphism from $\mathbf{Q}$ to $\mathbf{Q}'$.

Analogously to Definition 2.28, we define the *orthogonal group* of a quadratic form $\mathbf{Q}$ as follows.

Definition 2.62. Let $\mathbf{Q} \in \mathcal{S}_n^{>0}(\mathbb{R})$ be a quadratic form. An *automorphism of* $\mathbf{Q}$ is an isomorphism from $\mathbf{Q}$ to $\mathbf{Q}$. The *group of automorphisms of* $\mathbf{Q}$ is denoted by $\mathrm{O}(\mathbf{Q})$ and consists of all automorphisms of $\mathbf{Q}$.

ℤLIP is a problem that asks, given a lattice isomorphic to $\mathbb{Z}^n$, to output any isomorphism. Using quadratic forms, we get the following elegant problem description.

Problem 2.63 (ℤLIP)**.** Given upon input a quadratic form $\mathbf{Q} \in \mathrm{GL}_n(\mathbb{Z})$ for which there exists $\mathbf{B} \in \mathrm{GL}_n(\mathbb{Z})$ such that $\mathbf{Q} = \mathbf{B}^\mathsf{T}\mathbf{B}$ holds, output any such $\mathbf{B}$.

If one knows such a $\mathbf{B}$, it becomes easy to compute $\mathrm{O}(\mathbf{Q})$ since we know $\mathrm{O}(\mathbb{Z}^n)$. Without $\mathbf{B}$, we may compute the trivial automorphisms $\pm\mathbf{I}_n$. It is believed to be difficult in general, however, to construct any other automorphism of $\mathbf{Q}$ without solving ℤLIP.

Definition 2.64. The *inner product with respect to a quadratic form* $\mathbf{Q} \in \mathcal{S}_n^{>0}(\mathbb{R})$ is given by,

$$\langle \cdot, \cdot \rangle_\mathbf{Q} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x}^\mathsf{T} \cdot \mathbf{Q} \cdot \mathbf{y} \ .$$

The *norm with respect to* $\mathbf{Q} \in \mathcal{S}_n^{>0}(\mathbb{R})$ is defined as $\|\mathbf{x}\|_\mathbf{Q} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_\mathbf{Q}}$.

Note that for a basis $\mathbf{B}$ and vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have

$$\langle \mathbf{Bx}, \mathbf{By} \rangle = \mathbf{x}^\mathsf{T}\mathbf{B}^\mathsf{T}\mathbf{By} = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{B}^\mathsf{T}\mathbf{B}} \ ,$$

and thus the geometry of $\Lambda(\mathbf{B})$ is fully described by $\mathbf{Q} = \mathbf{B}^\mathsf{T}\mathbf{B}$. Moving from (bases of) lattices to quadratic forms can be viewed as forgetting about the specific embedding of the lattice in $\mathbb{R}^n$, while maintaining all geometric information. Throughout this thesis we will talk about lattices and quadratic forms interchangeably.

2.2.6 Module lattice and Hermitian form

A number field $\mathbb{K}$ is an algebraic extension of $\mathbb{Q}$ of finite degree $n = [\mathbb{K} : \mathbb{Q}]$. We write $\mathcal{O}_\mathbb{K}$ for the ring of integers of a general number field. In this work, we consider the cyclotomic field $\mathbb{K} = \mathbb{Q}(\zeta_{2n}) = \mathbb{Q}\left(e^{-2\pi i/2n}\right) \cong \mathbb{Q}[X]/(X^n+1)$ where $n \geq 2$ is a power of two. This is a CM field and has conductor $2n$. Many of the facts below are not true for general number fields.

The ring of integers $R = \mathcal{O}_\mathbb{K} \cong \mathbb{Z}[X]/(X^n+1)$, or any ideal of $\mathbb{K}$, obtains a rank-n lattice structure by considering its image under the so-called *canonical embedding*, or Minkowski embedding:

$$\sigma\colon \mathbb{K} \to \mathbb{C}^n,\ x \mapsto (\sigma_1(x), \ldots, \sigma_n(x)) \ .$$

Here, $\sigma_1, \sigma_2, \ldots, \sigma_n$ are the n canonical embeddings of $\mathbb{K}$ into $\mathbb{C}$, ordered such that $\sigma_{i+n/2} = \overline{\sigma_i}$ for $i \in [n/2]$ (for $n \geq 2$ there are no real embeddings). The subset $\{(x_1, \ldots, x_n) \in \mathbb{C}^n\colon \forall i \in [n/2], x_{i+n/2} = \overline{x_i}\} \subset \mathbb{C}^n$ is isomorphic as an inner product space to $\mathbb{R}^n$ [LPR10, Sec. 2.1]. We implicitly use this isomorphism and write $\sigma\colon \mathbb{K} \to \mathbb{R}^n$. We also have the *coefficient embedding*,

$$\textsc{Vec}\colon \mathbb{K} \to \mathbb{Q}^n, \quad a_0 + a_1 X + \cdots + a_{n-1}X^{n-1} \mapsto \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} \ ,$$

which is an additive group isomorphism. Moreover, $\textsc{Vec}(R) = \mathbb{Z}^n$.

The algebraic norm and trace are given by $\mathrm{N}(x) = \prod_{i=1}^n \sigma_i(x)$ and $\mathrm{Tr}(x) = \sum_{i=1}^n \sigma_i(x)$ for $x \in \mathbb{K}$, respectively. Since the σ_i are ring homomorphisms the algebraic norm is multiplicative and the trace is additive. If $x \in R$ then $\mathrm{N}(x), \mathrm{Tr}(x) \in \mathbb{Z}$.

The embeddings enable us to view $\mathbb{K}$ as an inner product space over $\mathbb{Q}$ by defining $\langle \cdot, \cdot \rangle : \mathbb{K} \times \mathbb{K} \to \mathbb{Q}$ as

$$\langle f, g \rangle = \frac{1}{n} \cdot \sum_{i=1}^{n} \overline{\sigma_i(f)} \cdot \sigma_i(g). \tag{2.8}$$

2

We renormalise by $\frac{1}{n}$ as there is an isometry, up to a scaling factor of n, from the canonical embedding to the coefficient embedding, i.e., we have $\langle f, g \rangle = \langle \textsc{Vec}(f), \textsc{Vec}(g) \rangle$ with the right hand inner product over $\mathbb{R}^n$. This gives a geometric norm on $\mathbb{K}$ as $\| \cdot \| : \mathbb{K} \to \mathbb{Q},\ f \mapsto \sqrt{\langle f, f \rangle}$, which agrees with the Euclidean norm of $\textsc{Vec}(f)$. In particular, given any $f = f_0 + f_1 X + \cdots + f_{n-1} X^{n-1} \in \mathbb{K}$, note $\langle 1, f \rangle = f_0$.

Because $\mathbb{K}$ is a CM field, it comes with a unique involutive automorphism $(\,\cdot\,)^\star : \mathbb{K} \to \mathbb{K}$ that acts as complex conjugation on its embeddings, i.e., we have $\sigma_i(x^\star) = \overline{\sigma_i(x)}$ for all $x \in \mathbb{K}$ and $i \in [n]$. We call $(\,\cdot\,)^\star$ *complex conjugation*, and we say $f \in \mathbb{K}$ is *self-adjoint* if $f^\star = f$. For all $f = f_0 + f_1 X + \cdots + f_{n-1} X^{n-1} \in \mathbb{K}$, we have

$$f^\star = f_0 - f_{n-1} X - \cdots - f_1 X^{n-1} \ .$$

Moreover, Equation (2.8) becomes,

$$\langle f, g \rangle = \frac{1}{n} \sum_{i=1}^{n} \sigma_i(f^\star g) = \frac{1}{n} \operatorname{Tr}(f^\star g) \ .$$

Module lattice

For any $\ell \in \mathbb{Z}_{\geq 1}$, we define $\mathbb{K}^\ell = \underbrace{\mathbb{K} \oplus \cdots \oplus \mathbb{K}}_{\ell \text{ times}}$, and similarly for the R-module R^ℓ of rank ℓ. Extend $\textsc{Vec} \colon \mathbb{K}^\ell \to \mathbb{Q}^{n\ell}$ in the natural way. We extend the inner product and norm to vectors $\mathbf{f}, \mathbf{g} \in \mathbb{K}^\ell$ by

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{i=1}^{\ell} \langle f_i, g_i \rangle \ , \qquad \text{and} \qquad \|\mathbf{f}\| = \sqrt{\langle \mathbf{f}, \mathbf{f} \rangle} \ .$$

We write $\textsc{Rot}(f) = \big(\textsc{Vec}(f), \textsc{Vec}(Xf), \ldots, \textsc{Vec}(X^{n-1} f)\big) \in \mathbb{Q}^{n \times n}$ for $f \in \mathbb{K}$, which is a basis for the lattice $\sigma(f)$ given by the (possibly fractional) ideal (f). Note that $\textsc{Rot}$ is a ring homomorphism, satisfying

$\textsc{Vec}(xy) = \textsc{Rot}(x)\,\textsc{Vec}(y)$ and $\textsc{Rot}(xy) = \textsc{Rot}(x)\,\textsc{Rot}(y)$ for $x, y \in \mathbb{K}$. We extend Rot to matrices $\mathbf{B} \in \mathbb{K}^{k\times\ell}$ in the natural way

$$\textsc{Rot}(\mathbf{B}) = \begin{pmatrix} \textsc{Rot}(\mathbf{B}_{11}) & \cdots & \textsc{Rot}(\mathbf{B}_{1\ell}) \\ \vdots & \ddots & \vdots \\ \textsc{Rot}(\mathbf{B}_{k1}) & \cdots & \textsc{Rot}(\mathbf{B}_{k\ell}) \end{pmatrix}.$$

We now define a module lattice. Since R is the ring of integers of a number field, it is a Dedekind domain and the notion of rank is well-defined for R-modules.

Definition 2.65. Let $M \subset \mathbb{K}^k$ be an R-module of rank $\ell \leq k$. Then, we have the map,

$$\boldsymbol{\sigma} = (\sigma, \ldots, \sigma)\colon \mathbb{K}^k \to \mathbb{R}^{nk}, \qquad (x_1, \ldots, x_k) \mapsto (\sigma(x_1), \ldots, \sigma(x_k)) \ ,$$

and we say the image $\boldsymbol{\sigma}(M)$ is a *module lattice*.

Any module lattice $\boldsymbol{\sigma}(M)$ is a rank-$n\ell$ lattice in $\mathbb{R}^{nk}$. We may refer to 'the module lattice M' to mean $\boldsymbol{\sigma}(M)$. If $\mathbf{B} \in \mathbb{K}^{k\times\ell}$ is a basis for an R-module M then $\textsc{Rot}(\mathbf{B}) \in \mathbb{Q}^{nk\times n\ell}$ is a basis for the module lattice $\boldsymbol{\sigma}(M)$. A matrix $\mathbf{B}$ is an R-basis for the module lattice R^ℓ whenever $\mathbf{B} \in \mathrm{GL}_\ell(R)$.

Definition 2.66. Given $k, \ell \in \mathbb{Z}_{\geq 1}$, the *Hermitian adjoint* of a matrix $\mathbf{B} \in \mathbb{K}^{k\times\ell}$ is denoted by $\mathbf{B}^\star$ and is obtained by applying complex conjugation coefficientwise to $\mathbf{B}^\mathsf{T}$. The Hermitian adjoint of a vector $\mathbf{f} \in \mathbb{K}^\ell$, is the row vector $\mathbf{f}^\star$, by interpreting the column vector as an $\ell \times 1$ matrix.

Hermitian form

Analogously to defining an R-basis, we may consider a structured variant of a quadratic form: the Hermitian form.

Definition 2.67. For $\ell \geq 1$, the *set of Hermitian forms* $\mathcal{H}_\ell^{>0}(\mathbb{K})$ consists of all $\mathbf{Q} \in \mathbb{K}^{\ell\times\ell}$ such that $\mathbf{Q}^\star = \mathbf{Q}$ and $\mathrm{Tr}(\mathbf{v}^\star\mathbf{Q}\mathbf{v}) > 0$ for all $\mathbf{v} \in \mathbb{K}^\ell \setminus \{\mathbf{0}\}$.

An *isomorphism* of Hermitian forms $\mathbf{Q}$ and $\mathbf{Q}'$ is a matrix $\mathbf{U} \in \mathrm{GL}_\ell(R)$ such that $\mathbf{U}^\star\mathbf{Q}\mathbf{U} = \mathbf{Q}'$ holds, and we analogously write $\mathrm{O}(\mathbf{Q})$ for the *group of automorphisms* of $\mathbf{Q}$.

Equivalently, $\mathbf{Q}$ is a Hermitian form whenever $\textsc{Rot}(\mathbf{Q})$ is a quadratic form. Given a basis $\mathbf{B} \in \mathbb{K}^{\ell \times \ell}$, its Gram matrix $\mathbf{B}^\star\mathbf{B}$ is a Hermitian form.

Note that $\textsc{Rot}$ maps $\mathrm{O}(\mathbf{Q})$ into $\mathrm{O}(\textsc{Rot}(\mathbf{Q}))$. The automorphism group of $\mathbf{I}_2(R)$ is equal to

$$\mathrm{O}(\mathbf{I}_2(R)) = \left\{ \begin{pmatrix} X^a & 0 \\ 0 & X^b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^c \;\middle|\; 0 \le a, b < n,\ c \in \{0,1\} \right\} ,$$

and has size $8n^2$. This is very little compared to the sheer number of automorphisms of $\textsc{Vec}(R^2) = \mathbb{Z}^{2n}$, which is $2^{2n}(2n)!$ by Example 2.30. Similar to Definition 2.64, a Hermitian form induces an inner product and norm.

Definition 2.68. The *inner product with respect to a Hermitian form* $\mathbf{Q} \in \mathcal{H}_\ell^{>0}(\mathbb{K})$ is given by,

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbf{Q}} = \frac{1}{n} \cdot \mathrm{Tr}(\mathbf{f}^\star \mathbf{Q} \mathbf{g}) \ , \qquad \text{and} \qquad \|\mathbf{f}\|_{\mathbf{Q}} = \sqrt{\langle \mathbf{f}, \mathbf{f} \rangle_{\mathbf{Q}}} \ ,$$

for $\mathbf{f}, \mathbf{g} \in \mathbb{K}^\ell$.

Once again, observe that for any basis $\mathbf{B}$, we have $\langle \mathbf{Bf}, \mathbf{Bg} \rangle = \langle \mathbf{f}, \mathbf{g} \rangle_{\mathbf{B}^\star\mathbf{B}}$ and $\|\mathbf{Bf}\| = \|\mathbf{f}\|_{\mathbf{B}^\star\mathbf{B}}$.

The following problem is the module analogue of ZLIP.

Problem 2.69 (smLIP)**.** Given upon input a matrix $\mathbf{Q} \in \mathrm{GL}_\ell(R)$ for which there exist $\mathbf{B} \in \mathrm{GL}_\ell(R)$ such that $\mathbf{Q} = \mathbf{B}^\star\mathbf{B}$, output any such $\mathbf{B}$.

Note that smLIP is not harder than ZLIP in dimension $n\ell$.

2.3 Duality

There are two ways to define the dual of a lattice. The first one is inherited from groups, while the second one is geometric and specific to lattices. In the context of Fourier transforms, it is useful to consider both definitions, and relate them. Note that characters are only taken for full-rank lattices.

Definition 2.70 (group-theoretic dual)**.** For a locally compact group G (e.g. a finite group or a torus $\mathbb{R}^n/\Lambda$ of a full-rank lattice Λ), the *group of characters on* G, is denoted by

$$\widehat{G} = \left\{ \chi \colon G \to \mathcal{S}^1 \;\middle|\; \chi \text{ continuous group homomorphism} \right\} .$$

Note that when G is a finite abelian group, any homomorphism $\chi\colon G \to \mathcal{S}^1$ is continuous.

Definition 2.71 (dual lattice)**.** The *dual of a lattice* $\Lambda \subset \mathbb{R}^n$ is denoted by $\Lambda^\vee$ and consists of all vectors $\mathbf{x} \in \operatorname{span}(\Lambda)$ for which $\langle \mathbf{x}, \Lambda \rangle \subseteq \mathbb{Z}$ holds.

We refer to Λ as the *primal lattice* and $\Lambda^\vee$ as the *dual lattice*.

Example 2.72. We have $(\mathbb{Z}^n)^\vee = \mathbb{Z}^n$ and $\mathcal{L}_q(\mathbf{A})^\vee = \frac{1}{q}\,\mathcal{L}_q^\perp(\mathbf{A})$. Moreover, for any lattice Λ and nonzero $\alpha \in \mathbb{R}$, $(\alpha\Lambda)^\vee = \Lambda^\vee/\alpha$.

The following lemma shows that we may interchange these two notions of duality, i.e., any dual vector defines a character and vice versa.

Lemma 2.73. *The map from* $\Lambda^\vee$ *to* $\widehat{\mathbb{R}^n/\Lambda}$ *that sends a dual vector* $\mathbf{w}$ *to the character*

$$\chi_{\mathbf{w}}\colon \mathbf{t} \mapsto \exp(2\pi i \cdot \langle \mathbf{t}, \mathbf{w} \rangle), \tag{2.9}$$

is a group isomorphism.

Proof. We prove the map $\mathbf{w} \mapsto \chi_{\mathbf{w}}$ is a well-defined group homomorphism, injective and surjective.

Well-definedness: It follows directly from the definition of $\Lambda^\vee$ that $\chi_{\mathbf{w}}$ is a character given $\mathbf{w} \in \Lambda^\vee$. It is also easy to see $\mathbf{w} \mapsto \chi_{\mathbf{w}}$ is a group homomorphism.

Injectivity: Suppose $\chi_{\mathbf{w}}(\mathbf{t}) = 1$ holds for all $\mathbf{t} \in \mathbb{R}^n$. In particular, for $\mathbf{t} = c\mathbf{w}$ with $c \in \mathbb{R}_{>0}$, we have $\chi_{\mathbf{w}}(c\mathbf{w}) = 1$, which implies $\langle \mathbf{w}, c\mathbf{w} \rangle \in \mathbb{Z}$. Assume $\mathbf{w} \neq \mathbf{0}$ so in particular $\|\mathbf{w}\| > 0$. Then, for $c < 1/\|\mathbf{w}\|^2$, we get a contradiction as $\langle \mathbf{w}, c\mathbf{w} \rangle = c\|\mathbf{w}\|^2$ must be simultaneously a (strictly) positive integer and smaller than 1. We conclude $\mathbf{w} = \mathbf{0}$.

Surjectivity: Let $\chi \in \widehat{\mathbb{R}^n/\Lambda}$ be given. By continuity, there is an open ball $U \subseteq \mathbb{R}^n$ centred at $\mathbf{0}$ such that $\chi(U + \Lambda) \subseteq \{z \in \mathcal{S}^1 | \operatorname{Re}(z) > 0\}$ holds. On this ball, we can find a linear map $\varphi\colon U \to (-\frac{1}{4}, \frac{1}{4})$ such that $\chi(\mathbf{x} + \Lambda) = \exp(2\pi i \varphi(\mathbf{x}))$ holds for all $\mathbf{x} \in U$ as there is exactly one value possible for $\varphi(\mathbf{x})$. The character χ is completely determined by its values on U, i.e., for any $\mathbf{x} \in \mathbb{R}^n$ there exists $m \in \mathbb{Z}_{\geq 1}$ such that $\mathbf{x}/m \in U$ so $\chi(\mathbf{x}) = \chi(\mathbf{x}/m)^m$. We may now extend φ to a linear function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ that satisfies $\chi(\mathbf{x} + \Lambda) = \exp(2\pi i \varphi(\mathbf{x}))$ for all $\mathbf{x} \in \mathbb{R}^n$. Now there is a $\mathbf{w} \in \mathbb{R}^n$ such that $\varphi = \langle \cdot, \mathbf{w} \rangle$, as any linear map is of this form. Note because $\chi(\Lambda) = 1$ holds, we have $\langle \mathbf{w}, \Lambda \rangle \subset \mathbb{Z}$ so $\mathbf{w} \in \Lambda^\vee$. Thus, $\chi = \chi_{\mathbf{w}}$. □

For a sublattice $\Lambda' \subset \Lambda$, the dual of the finite group Λ/Λ' has a natural connection to the dual lattices of Λ' and Λ by the following lemma.

Lemma 2.74. *For two full-rank lattices $\Lambda_1 \subset \Lambda_2 \subset \mathbb{R}^n$, there is a canonical group isomorphism of abelian groups,*

$$\left(\widehat{\mathbb{R}^n/\Lambda_1}\right)\Big/\left(\widehat{\mathbb{R}^n/\Lambda_2}\right) \to \widehat{\Lambda_2/\Lambda_1},$$

given by restriction of a character $\chi\colon \mathbb{R}^n/\Lambda_1 \to \mathcal{S}^1$ (modulo $\widehat{\mathbb{R}^n/\Lambda_2}$) to Λ_2/Λ_1.

Proof. We prove restriction is a well-defined group homomorphism, injective and surjective.

Well-definedness: Any Λ_1-periodic character χ can be multiplied by a Λ_2-periodic character in $\widehat{\mathbb{R}^n/\Lambda_2}$ as the function stays the same on Λ_2/Λ_1.

Injectivity: Any character $\chi \in \widehat{\mathbb{R}^n/\Lambda_1}$ that is 1 on Λ_2/Λ_1 is coming from a function $\mathbb{R}^n \to \mathcal{S}^1$ that is Λ_2-periodic, i.e., a character from $\widehat{\mathbb{R}^n/\Lambda_2}$.

Surjectivity: Left hand side has size

$$\left|\widehat{\mathbb{R}^n/\Lambda_1}\Big/\widehat{\mathbb{R}^n/\Lambda_2}\right| = |\Lambda_1^\vee/\Lambda_2^\vee| = |\Lambda_2/\Lambda_1|,$$

and right hand side has size $|\widehat{\Lambda_2/\Lambda_1}| = |\Lambda_2/\Lambda_1|$, where we use that a finite abelian group G is isomorphic to its dual. Since there is an injection from one set to another set of the same finite cardinality, it must be surjective as well. □

One can compute a Λ-periodic function by evaluating its Fourier transform at dual lattice points.

Theorem 2.75 (Poisson summation formula, [SW71, Chapter VII])**.** *Given a full-rank lattice Λ and a function $f\colon \mathbb{R}^n \to \mathbb{R}$ satisfying certain decaying conditions, for all $\mathbf{t} \in \mathbb{R}^n$ one has*

$$\sum_{\mathbf{v}\in\Lambda} f(\mathbf{v}+\mathbf{t}) = \frac{1}{\det(\Lambda)} \sum_{\mathbf{w}\in\Lambda^\vee} e^{2\pi i\langle \mathbf{t},\mathbf{w}\rangle} \widehat{f}(\mathbf{w}).$$

2.3.1 Dual basis

Given a basis $\mathbf{B}$ of a primal lattice Λ, one can compute a basis for the dual lattice $\Lambda^\vee$.

Definition 2.76 (dual basis)**.** Given a basis $\mathbf{B}$ of the primal lattice Λ, the *dual basis associated to* $\mathbf{B}$ is

$$\mathbf{B}^\vee = [\mathbf{b}_1^\vee, \dots, \mathbf{b}_k^\vee] = \mathbf{B} \cdot \left(\mathbf{B}^\mathsf{T} \cdot \mathbf{B}\right)^{-1} ,$$

and ${}^\vee\mathbf{B} = [\mathbf{b}_k^\vee, \dots, \mathbf{b}_1^\vee]$ is used to denote the *reversed dual basis*, which contains the basis vectors of $\mathbf{B}^\vee$ in reversed ordering.

There are three remarks to make here.

(1) The map $\mathbf{B} \mapsto \mathbf{B}^\vee$ is an involution, i.e., $\mathbf{B} = (\mathbf{B}^\vee)^\vee$. Similarly, we have $\mathbf{B} = {}^\vee({}^\vee\mathbf{B})$;
(2) The dual basis $\mathbf{D} = [\mathbf{b}_1^\vee, \dots, \mathbf{b}_k^\vee] = \mathbf{B}^\vee$ is the unique basis with vectors in $\mathrm{span}(\mathbf{B})$ that satisfies $\left\langle \mathbf{b}_i, \mathbf{b}_j^\vee \right\rangle = [\![i = j]\!]$, or equivalently, $\mathbf{D}^\mathsf{T}\mathbf{B} = \mathbf{I}_k(\mathbb{Z})$. Hence, $\mathbf{B}^\vee = \mathbf{B}^{-\mathsf{T}}$ if $\mathbf{B}$ is full-rank.
(3) Any point $\mathbf{x} = \sum_{i=1}^k c_i \mathbf{b}_i$ with $c_i \in \mathbb{R}$ satisfies $c_i = \langle \mathbf{x}, \mathbf{b}_i^\vee \rangle$, because $\mathbf{D}^\mathsf{T}\mathbf{x} = (\mathbf{D}^\mathsf{T}\mathbf{B})\mathbf{c} = \mathbf{c}$. Hence, a point $\mathbf{x} \in \mathrm{span}(\mathbf{B})$ satisfies $\mathbf{x} \in \mathrm{span}(\mathbf{b}_2, \dots, \mathbf{b}_k)$ if and only if $\langle \mathbf{x}, \mathbf{b}_1^\vee \rangle = 0$. Generalizing this further, leads to following proposition.

Proposition 2.77 ([Mic14, Thm. 5])**.** *Let $\mathbf{B} \in \mathbb{R}^{n \times k}$ be a basis and $1 \le \ell \le r \le k$. Then, the projected sublattice $({}^\vee\mathbf{B})_{[k+1-r,k+1-\ell]}$ is the reversed dual basis associated to the projected sublattice $\mathbf{B}_{[\ell,r]}$.*

Informally, sectioning in the primal basis corresponds to projecting in the reversed dual basis.

Proof. Iteratively and using (1), we may reduce to the simplest case $\ell = 2$ and $r = k$. The above shows $({}^\vee\mathbf{B})_{[1,k-1]}$ is the reversed dual of $\mathbf{B}_{[2,k]}$. □

This provides three corollaries: (1) $\det(\Lambda^\vee) = 1/\det(\Lambda)$, (2) for all $i \in [k]$ we have $\|\mathbf{b}_i^\vee\| = 1/\|\mathbf{b}_i^*\|$, and (3) $\mathbf{b}_k^*/\|\mathbf{b}_k^*\|^2 = \mathbf{b}_k^\vee$ is a dual vector.

2.3.2 Discrete Fourier transform

For any set S let $\mathbb{C}^S$ be the group of sequences $(x_s)_{s\in S}$ having complex coefficients x_s, where the group operation is given by pointwise addition. Given a finite group G, the *discrete Fourier transform* (DFT) of a sequence $(x_g)_{g\in G} \subset \mathbb{C}$ is the $\mathbb{C}$-linear map

$$\begin{aligned} \mathrm{DFT}_G\colon \qquad & \mathbb{C}^G \to \mathbb{C}^{\widehat{G}}, \\ & (x_g)_{g\in G} \mapsto \left(\sum_{g\in G} x_g \cdot \overline{\chi(g)}\right)_{\chi\in\widehat{G}}. \end{aligned} \tag{2.10}$$

The m-dimensional *fast Fourier transform* (FFT) is an algorithm that, upon input a group G, given as $n_1, \dots, n_m \in \mathbb{Z}_{\geq 2}$ such that $G \cong \bigoplus_{j=1}^m (\mathbb{Z}/n_j\mathbb{Z})$, and $(x_g)_{g\in G}$, outputs $\mathrm{DFT}_G\Big((x_g)_{g\in G}\Big)$ in time $O(|G|\log|G|)$. There are various FFTs known for any finite group G (even when an n_i is a large prime) [Rad68, DV90]. When the group G is not cyclic, the algorithm is often referred to as a multidimensional FFT. When $G \cong (\mathbb{Z}/2\mathbb{Z})^k$, the algorithm is a *Walsh–Hadamard transform* (WHT), which is more efficient in practice. For a finite group G, the inverse of DFT_G is given by

$$\mathrm{DFT}_G^{-1}\Big((y_\chi)_{\chi\in\widehat{G}}\Big) = \frac{1}{|G|}\cdot\left(\sum_{\chi\in\widehat{G}} y_\chi\chi(g)\right)_{g\in G}.$$

Identifying an element $g \in G$ with the evaluation map $\mathrm{ev}_g\colon \chi \mapsto \chi(g)$ gives the canonical isomorphism $G \cong \widehat{\left(\widehat{G}\right)}$, so an inverse DFT is basically a DFT up to some reordering.

2.4 Discrete Gaussian sampling

Given a lattice Λ and parameter $\sigma \in \mathbb{R}_{>0}$, we may consider a discrete analogue of the continuous Gaussian.

Definition 2.78. For any lattice $\Lambda \subset \mathbb{R}^n$ and vector $\mathbf{c} \in \mathbb{R}^n$, we denote the *discrete Gaussian distribution* on $\Lambda + \mathbf{c}$ with parameter σ by $\mathcal{D}_{\Lambda+\mathbf{c},\sigma}$, which assigns a probability

$$\frac{1}{\sum_{y\in\Lambda+\mathbf{c}} \rho_\sigma(\mathbf{y})} \cdot \rho_\sigma(\mathbf{x}),$$

to a point $\mathbf{x} \in \Lambda + \mathbf{c}$, and probability zero to all points in $\mathbb{R}^n \setminus (\Lambda + \mathbf{c})$. The discrete Gaussian distribution $\mathcal{D}_{\Lambda,\sigma}$ is shorthand for $\mathcal{D}_{\Lambda+\mathbf{0},\sigma}$.

Note that the discrete Gaussian is usually defined [MR07] with respect to a width $s = \sqrt{2\pi}\sigma$, while we mimic Falcon in this regard [PFH+22]. Moreover, σ is generally not the standard deviation of a discrete Gaussian, though it will be close for large enough σ, see Section 2.4.1. A discrete Gaussian has a similar tail bound to a continuous Gaussian.

Lemma 2.79 ([Ban93, Lem. 1.5(ii)])**.** *For any lattice $\Lambda \subset \mathbb{R}^n$, point $\mathbf{c} \in \mathbb{R}^n$ and $\tau \geq 1$, we have*

$$\Pr_{\mathbf{x} \sim \mathcal{D}_{\Lambda+\mathbf{c},\sigma}} [\|\mathbf{x}\| > \tau\sigma\sqrt{n}] \leq 2\frac{\rho_\sigma(\Lambda)}{\rho_\sigma(\Lambda + \mathbf{c})} \cdot \tau^n e^{-\frac{n}{2}(\tau^2-1)}.$$

We may also define a discrete Gaussian distribution with respect to a quadratic form instead of a lattice.

Definition 2.80. *Gaussian mass with respect to* $\mathbf{Q} \in \mathcal{S}_n^{>0}(\mathbb{R})$ is given by

$$\rho_{\mathbf{Q},\sigma} \colon \mathbb{R}^n \to \mathbb{R}, \mathbf{x} \mapsto \exp\left(-\|\mathbf{x}\|_{\mathbf{Q}}^2/2\sigma^2\right) .$$

For any $\mathbf{c} \in \mathbb{R}^n$, the *discrete Gaussian distribution on* $\mathbb{Z}^n + \mathbf{c}$ *with respect to* $\mathbf{Q}$ *and parameter* σ is denoted by $\mathcal{D}_{\mathbf{Q},\mathbb{Z}^n+\mathbf{c},\sigma}$, which assigns a probability

$$\frac{1}{\sum_{\mathbf{y} \in \mathbb{Z}^n + \mathbf{c}} \rho_{\mathbf{Q},\sigma}(\mathbf{y})} \rho_{\mathbf{Q},\sigma}(\mathbf{x}) \quad ,$$

to a point $\mathbf{x} \in \mathbb{Z}^n + \mathbf{c}$, and zero otherwise. The distribution $\mathcal{D}_{\mathbf{Q},\sigma}$ is shorthand for $\mathcal{D}_{\mathbf{Q},\mathbb{Z}^n+\mathbf{0},\sigma}$.

If $\mathbf{Q} = \mathbf{B}^\mathsf{T} \cdot \mathbf{B}$, note we have

$$\mathbf{B} \cdot \mathcal{D}_{\mathbf{Q},\mathbb{Z}^n+\mathbf{c},\sigma}(\mathbf{x}) = \mathcal{D}_{\Lambda(\mathbf{B})+\mathbf{B}\cdot\mathbf{c},\sigma}(\mathbf{B} \cdot \mathbf{x}) \quad ,$$

as $\rho_{\mathbf{Q},\sigma}(\mathbf{x}) = \rho_\sigma(\mathbf{B} \cdot \mathbf{x})$.

When σ is large enough compared to the maximum length of a Gram–Schmidt basis vector, denoted by $\|\mathbf{B}_{\mathbf{Q}}^*\|$, we can efficiently sample a discrete Gaussian.

Lemma 2.81 ([BLP+13, Lem. 2.3], adapted)**.** *There is a probabilistic polynomial-time algorithm that on input a quadratic form $\mathbf{Q} \in \mathcal{S}_n^{>0}(\mathbb{R})$, $\mathbf{c} \in \mathbb{R}^n$ and parameter $\sigma \geq \|\mathbf{B}_{\mathbf{Q}}^*\| \cdot (1/\pi) \cdot \sqrt{\log(2n+4)/2}$ outputs a sample according to $\mathcal{D}_{\mathbf{Q},\mathbb{Z}^n+\mathbf{c},\sigma}$.*

Additionally, we extend the discrete Gaussian to Hermitian forms.

Definition 2.82. Given a Hermitian form $\mathbf{Q} \in \mathcal{H}_\ell^{>0}(\mathbb{K})$, the discrete Gaussian distribution with respect to $\mathbf{Q}$ is given by

$$\mathcal{D}_{\mathbf{Q},\sigma}(\mathbf{x}) = \mathcal{D}_{\mathrm{Rot}(\mathbf{Q}),\sigma}(\mathrm{Vec}(\mathbf{x})) \ ,$$

for any $\mathbf{x} \in \mathbb{K}^\ell$.

Due to our choice of $\mathbb{K}$ and the definition of our $\mathbf{Q}$-norm, this is equivalent to the natural definition that follows from $\rho_{\mathbf{Q},\sigma}$ when forgetting the module structure.

2.4.1 Smoothing parameter

Definition 2.83. For $\varepsilon > 0$ we define the *smoothing parameter* $\eta_\varepsilon(\Lambda)$ as the smallest $\sigma \in \mathbb{R}_{>0}$ satisfying $\rho_{1/(2\pi\sigma)}(\Lambda^\vee \setminus \{\mathbf{0}\}) \leq \varepsilon$.

Note that η_ε is usually defined with respect to a width $s = \sqrt{2\pi}\sigma$. Here its value is a factor $\sqrt{2\pi}$ smaller than usual [MR07, GPV08]. If $\sigma \geq \eta_\varepsilon(\Lambda)$ then $\mathcal{D}_{\Lambda+\mathbf{c},\sigma}$ exhibits several useful properties. For example, σ is close to the standard deviation of $\mathcal{D}_{\Lambda+\mathbf{c},\sigma}$, with the closeness parametrised by ε, see [MR07, Lem. 4.3], and cosets have similar weights. We may say σ is 'above smoothing' to refer to $\sigma \geq \eta_\varepsilon(\Lambda)$ for some implicit appropriate ε.

Lemma 2.84 ([MR07, Proof of Lem. 4.4])**.** *For any lattice $\Lambda \subset \mathbb{R}^n$, point $\mathbf{c} \in \mathbb{R}^n$, and $\varepsilon \in (0,1)$, $\sigma \geq \eta_\varepsilon(\Lambda)$, we have*

$$(1-\varepsilon) \cdot \frac{\left(\sigma\sqrt{2\pi}\right)^n}{\det(\Lambda)} \leq \rho_\sigma(\Lambda + \mathbf{c}) \leq (1+\varepsilon) \cdot \frac{\left(\sigma\sqrt{2\pi}\right)^n}{\det(\Lambda)},$$

To use above lemma for quadratic forms, note that for any basis $\mathbf{B}$ of Λ, we have $\rho_\sigma(\Lambda + \mathbf{c}) = \rho_{\mathbf{B}^\intercal\mathbf{B},\sigma}(\mathbb{Z}^n + \mathbf{B}^{-1}\mathbf{c})$.

2.5 Lattice basis reduction

There are an infinite number of bases for one lattice. It is hard to find a basis with short basis vectors, and this process is called lattice basis reduction. On the other hand, given a basis $\mathbf{B}$, one can easily

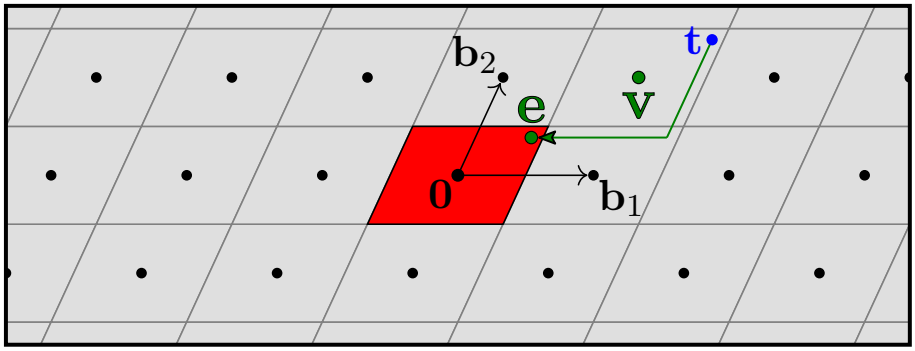


Figure 2.3: Fundamental domain of $\mathbf{B}$

obtain the basis $\mathbf{BU}$ with (generally) longer vectors, by generating some $\mathbf{U} \in \mathrm{GL}_k(\mathbb{Z})$ arbitrarily.

We discuss reductions of a vector with respect to a basis, a basis itself and a rank-2 lattices. Then we discuss a polynomial-time basis reduction algorithm [LLL82], and an exponential-time algorithm that provides stronger basis reduction [SE94].

2.5.1 Babai's reduction algorithms

Here, we will look at the Rounding-Off algorithm and Babai's Nearest-Plane algorithm [Bab86], which both solve an approximation variant of CVP.

Definition 2.85 (approx-CVP$_\gamma$)**.** Given an approximation factor $\gamma \geq 1$, a lattice Λ and a vector $\mathbf{t} \in \mathbb{R}^n$, find a vector $\mathbf{v} \in \Lambda$ such that $\|\mathbf{v} - \mathbf{t}\| \leq \gamma \|\mathbf{w} - \mathbf{t}\|$ holds for all $\mathbf{w} \in \Lambda$.

Intuitively, this definition asks given a target $\mathbf{t}$ having lattice point $\mathbf{v}_0$ closest by, to find any nearby lattice point that is at most a factor γ further away from $\mathbf{t}$.

Rounding-Off

Given a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ and a vector $\mathbf{t} = \sum_{i=1}^{k} c_i \mathbf{b}_i$ with $c_i \in \mathbb{R}$, rounding each c_i to the nearest integer, i.e., $n_i = \lceil c_i \rfloor$, provides a lattice point $\mathbf{v} = \sum_{i=1}^{k} n_i \mathbf{b}_i \in \Lambda$ that is somewhat near to $\mathbf{t}$.

Algorithm 2.2. RO$(\mathbf{B}, \mathbf{t})$: Rounding-Off algorithm

Input: basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ and target $\mathbf{t} \in \mathbb{R}^n$
Output: $(\mathbf{v}, \mathbf{e}) \in \Lambda \times \mathbb{R}^n$ such that $\mathbf{t} = \mathbf{v} + \mathbf{e}$ and $\pi_{\mathrm{span}(\mathbf{B})}(\mathbf{e}) \in \mathcal{P}(\mathbf{B})$

1: $\mathbf{c} \leftarrow \left\lceil (\mathbf{B}^\mathsf{T}\mathbf{B})^{-1} \cdot \mathbf{B}^\mathsf{T} \cdot \mathbf{t} \right\rfloor$ ▷ $\mathbf{c} \in \mathbb{Z}^k$
2: $\mathbf{v} \leftarrow \mathbf{B}\mathbf{c}$
3: **return** $(\mathbf{v}, \mathbf{t} - \mathbf{v})$

We will refer to this procedure, which can be found in Algorithm 2.2, the *Rounding-Off algorithm (RO)*. Note, this procedure may be considered "straightforward" [Bab86].

Remark 2.86. We assume that the rounding function,

$$\lceil \cdot \rfloor : \mathbb{R} \to \mathbb{Z}, x \mapsto \lceil x \rfloor \ ,$$

satisfies $x - \lceil x \rfloor \in (-1/2, 1/2]$ for all $x \in \mathbb{R}$, or equivalently, given $n \in \mathbb{Z}$, we have $\lceil x \rfloor = n$ whenever $x \in (n - \frac{1}{2}, n + \frac{1}{2}]$. Programming languages, however, round differently. For example, C's `lrint` function, and Python round a floating-point number $x \in \mathbb{Z} + \frac{1}{2}$ towards an even integer, whereas C's `round` function rounds away from zero.

Definition 2.87 (fundamental domain)**.** Given a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$, its *fundamental domain* is given by

$$\mathcal{P}(\mathbf{B}) = \left\{ \mathbf{x} \in \mathrm{span}(\mathbf{B}) \;\middle|\; \exists \mathbf{y} \in \left(-\frac{1}{2}, \frac{1}{2}\right]^k : \mathbf{x} = \mathbf{B} \cdot \mathbf{y} \right\} \ .$$

The output vector $\mathbf{e}$ of the pair $(\mathbf{v}, \mathbf{e})$ lies in the fundamental domain of $\mathbf{B}$, which can be seen in Figure 2.3.

Nearest-Plane

Instead of determining all $c_1, \ldots, c_k$ in one go as done in Algorithm 2.2, we may as well choose to reduce $\mathbf{t}$ using an integer multiple of $\mathbf{b}_n$ that minimises $\langle \mathbf{t} - x\mathbf{b}_n, \mathbf{b}_n^* \rangle$, because any subsequent additions of $\mathbf{b}_{n-1}, \ldots, \mathbf{b}_1$ keep the inner product with $\mathbf{b}_n^*$ the same by definition of Gram–Schmidt orthogonalisation.

We will call this procedure, which can be found in Algorithm 2.3, *Babai's Nearest-Plane algorithm (NP)* [Bab86], although it was known before Babai's work [LLL82, Len83].

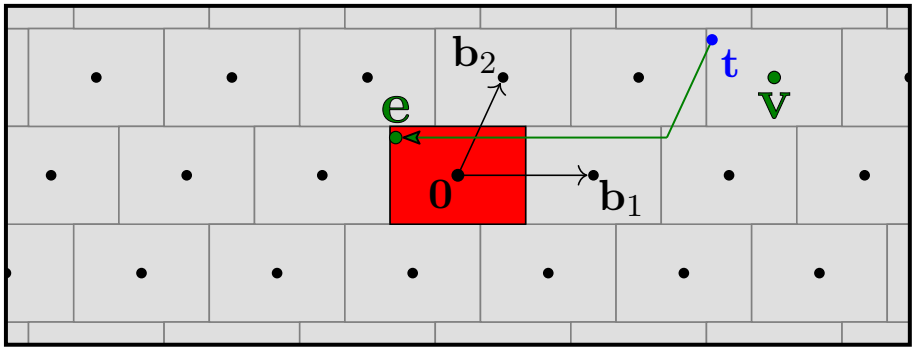


Figure 2.4: Babai's domain of $\mathbf{B}$

Algorithm 2.3. NP$(\mathbf{B}, \mathbf{t})$: Babai's Nearest-Plane algorithm

Input: basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ and target $\mathbf{t} \in \mathbb{R}^n$
Output: $(\mathbf{v}, \mathbf{e}) \in \Lambda \times \mathbb{R}^n$ such that $\mathbf{t} = \mathbf{v} + \mathbf{e}$ and $\pi_{\mathrm{span}(\mathbf{B})}(\mathbf{e}) \in \mathcal{P}(\mathbf{B}^*)$

$(\mathbf{v}, \mathbf{e}) \leftarrow (\mathbf{0}, \mathbf{t})$
for $i = k, k-1, \ldots, 1$ **do** ▷ Invariant: $\mathbf{t} = \mathbf{v} + \mathbf{e}$
  $c_i \leftarrow \lceil \langle \mathbf{b}_i^*, \mathbf{e} \rangle / \|\mathbf{b}_i^*\|^2 \rfloor$
  $\mathbf{v} \leftarrow \mathbf{v} + c_i \mathbf{b}_i$
  $\mathbf{e} \leftarrow \mathbf{e} - c_i \mathbf{b}_i$
return $(\mathbf{v}, \mathbf{e})$

Definition 2.88 (Babai's domain)**.** *Babai's domain* with respect to a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ is given by $\mathcal{P}(\mathbf{B}^*)$.

The output vector $\mathbf{e}$ of the pair $(\mathbf{v}, \mathbf{e})$ lies in Babai's domain, which can be seen in Figure 2.4.

Proposition 2.89. *Given a basis* $\mathbf{B} \in \mathbb{R}^{n \times k}$ *we have*

$$\max \left\{ \|\mathbf{e}\|^2 \;\middle|\; \mathbf{e} \in \mathcal{P}(\mathbf{B}^*) \right\} = \frac{1}{4} \sum_{i=1}^{n} \|\mathbf{b}_i^*\|^2 \ .$$

On the other hand, there is no such simple bound on the norm of elements in $\mathcal{P}(\mathbf{B})$. Namely, for every $L \in \mathbb{R}$ and unit-volume lattice in dimension 2 and higher, there is a basis such that $\mathcal{P}(\mathbf{B})$ has a vector of norm $\geq L$.

2.5.2 Size reduction

Here we discuss the reduction of a basis that makes the basis vectors nearly orthogonal using Rounding-Off or Babai's Nearest-Plane. We may reduce the length of a basis vector $\mathbf{b}_i$ using Babai's Nearest-Plane algorithm with respect to $[\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}]$. This reduces the size of the μ_{ij} coefficients and effectively makes pairs of basis vectors "nearly orthogonal", while $\mathbf{B}^*$ is kept the same.

Definition 2.90 (size-reduced)**.** A basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ is *size-reduced* if for all $1 \leq i < j \leq k$ we have $|\mu_{ij}| \leq \frac{1}{2}$.

Algorithm 2.4 transforms the basis $\mathbf{B}$ into a size-reduced basis. Moreover, the implicit basis transformation matrix $\mathbf{U}$ is unitriangular. Thus, a new $\mathbf{b}'_i$ is the sum of $\mathbf{b}_i$ and a $\mathbb{Z}$-linear combination of $\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}$.

Algorithm 2.4. SizeReduce($\mathbf{B}$): size reduction of a basis

Input: basis $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_k] \in \mathbb{R}^{n \times k}$
Output: size-reduced basis $\mathbf{B}' = \mathbf{B}\mathbf{U}$ for some $\mathbf{U} \in \mathrm{GL}_k(\mathbb{Z})$

1: $\mathbf{b}'_1 \leftarrow \mathbf{b}_1$
2: **for** $i = 2, 3, \ldots, k$ **do**
3: $\quad (\_, \mathbf{b}'_i) \leftarrow \mathrm{NP}([\mathbf{b}'_1, \ldots, \mathbf{b}'_{i-1}], \mathbf{b}_i)$
4: **return** $\mathbf{B}' = [\mathbf{b}'_1, \ldots, \mathbf{b}'_k]$

Proposition 2.91. *Given a size-reduced basis* $\mathbf{B} \in \mathbb{R}^{n \times k}$*, for all* $\mathbf{e} \in \mathcal{P}(\mathbf{B})$ *we have*

$$\|\mathbf{e}\|^2 \leq \frac{k^2}{4} \sum_{i=1}^{k} \|\mathbf{b}_i^*\|^2 \ .$$

Proof. Let us write $\mathbf{b}_j = \mathbf{b}_j^* + \sum_{i<j} \mu_{ij} \mathbf{b}_i^*$ with $|\mu_{ji}| \leq \frac{1}{2}$ due to size-reduction, and $\mathbf{e} = \sum_{i=1}^{k} c_i \mathbf{b}_i$ with $c_i \in (-1/2, 1/2]$ by definition of $\mathcal{P}(\mathbf{B})$. Then,

$$\|\mathbf{e}\|^2 = \sum_{i=1}^{k} \|\mathbf{b}_i^*\|^2 \cdot \left(c_i + \sum_{j=i+1}^{k} c_j \mu_{ij} \right)^2 \ .$$

By the triangle inequality, for all $i \in [k]$ we have

$$\left| c_i + \sum_{j=i+1}^{k} c_j \mu_{ij} \right| \leq \frac{1}{2} + \frac{k-i}{4} \leq \frac{k}{2} \ ,$$

and the claim follows. □

Comparing Proposition 2.89 and Proposition 2.91, Nearest-Plane reduces vectors to a norm that may be a factor k smaller than that of Rounding-Off.

Proposition 2.92. *A size-reduced basis* $\mathbf{B} \in \mathbb{R}^{n\times k}$ *satisfies*

$$\forall i \in [k]\colon \qquad \|\mathbf{b}_i^*\|^2 \le \|\mathbf{b}_i\|^2 \le \|\mathbf{b}_i^*\|^2 + \frac{1}{4}\sum_{j=1}^{i-1}\|\mathbf{b}_j^*\|^2 \ .$$

Proof. The lower bound on $\|\mathbf{b}_i\|$ follows from Equation (2.7), and the upper bound follows from Proposition 2.89 using basis $\mathbf{B}_{[1,i-1]}$ and target $\mathbf{e} = \mathbf{b}_i - \mathbf{b}_i^*$. □

Thus, we need to reduce the Gram–Schmidt norms to obtain a basis with short basis vectors.

The condition number of a basis is a way to measure tolerance of numerical errors in algorithms utilising floating-points. Moreover, it enables one to bound the unimodular transformation to make a basis size-reduced [RH23, App. B.2].

Definition 2.93 (matrix norms)**.** Let $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_k] \in \mathbb{R}^{k\times k}$ be a full-rank matrix. Its *ℓ^∞-norm and Frobenius norm* are given by, respectively,

$$\|\mathbf{A}\|_\infty = \max_{i,j} |\mathbf{A}_{ij}|, \quad \text{and} \quad \|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{k}\|\mathbf{a}_i\|^2} = \sqrt{\sum_{i,j=1}^{k}\mathbf{A}_{ij}^2} \ .$$

Its *condition number* is denoted by $\kappa(\mathbf{A}) = \|\mathbf{A}\|_{\mathrm{F}} \cdot \|\mathbf{A}^{-1}\|_{\mathrm{F}}$. We say $\mathbf{A}$ is *well-conditioned* if $\kappa(\mathbf{A})$ is small, and ill-conditioned if it is large.

Note $\|\mathbf{A}\|_\infty \le \|\mathbf{A}\|_{\mathrm{F}} \le k \cdot \|\mathbf{A}\|_\infty$ holds for any matrix $\mathbf{A}$ so these norms are comparable in size. As the ℓ^2-norm of a vector is invariant under orthogonal transformations, the Frobenius norm is invariant as well, i.e., for all $\mathbf{O} \in \mathrm{O}_k(\mathbb{R})$ we have $\|\mathbf{OA}\|_{\mathrm{F}} = \|\mathbf{A}\|_{\mathrm{F}}$. Therefore, given a QR decomposition of a full-rank basis $\mathbf{B} = \mathbf{QR}$, we have $\kappa(\mathbf{B}) = \kappa(\mathbf{R})$.

Remark 2.94. Although a size-reduced basis has a small Frobenius norm, its condition number may be exponentially large in k. Namely, given

a size-reduced unitriangular $\mathbf{R} \in \mathbb{R}^{k\times k}$, its inverse has entries satisfying $|(\mathbf{R}^{-1})_{ij}| \leq \frac{1}{2}(3/2)^{k-2}$. This follows from an upper bound on the adjugant matrix. For example, for $k = 4$ we have

2

$$\left(\mathbf{R}^{-1}\right)_{14} = (-1)\cdot \det \begin{pmatrix} \mathbf{R}_{12} & \mathbf{R}_{13} & \mathbf{R}_{14} \\ 1 & \mathbf{R}_{23} & \mathbf{R}_{24} \\ 0 & 1 & \mathbf{R}_{34} \end{pmatrix} .$$

Developing this determinant and taking absolute values yields

$$\left|\left(\mathbf{R}^{-1}\right)_{14}\right| \leq |\mathbf{R}_{12}| \cdot \left|\det \begin{pmatrix} \mathbf{R}_{23} & \mathbf{R}_{24} \\ 1 & \mathbf{R}_{34} \end{pmatrix}\right| + 1 \cdot \left|\det \begin{pmatrix} \mathbf{R}_{13} & \mathbf{R}_{14} \\ 1 & \mathbf{R}_{34} \end{pmatrix}\right| .$$

The first determinant satisfies

$$\left|\det \begin{pmatrix} \mathbf{R}_{23} & \mathbf{R}_{24} \\ 1 & \mathbf{R}_{34} \end{pmatrix}\right| \leq |\mathbf{R}_{23}\mathbf{R}_{34}| + |\mathbf{R}_{24}| \leq \frac{1}{4} + \frac{1}{2} = \frac{3}{4} ,$$

and the second determinant has the same bound. Hence,

$$\left|\left(\mathbf{R}^{-1}\right)_{14}\right| \leq |\mathbf{R}_{12}| \cdot \frac{3}{4} + \frac{3}{4} \leq \left(\frac{1}{2} + 1\right) \cdot \frac{3}{4} = \frac{1}{2}\left(\frac{3}{2}\right)^2 .$$

The upper bound is attained for the following unitriangular matrix:

$$\mathbf{R} = \begin{pmatrix} 1 & -\frac{1}{2} & \cdots & -\frac{1}{2} \\ & 1 & \ddots & \vdots \\ & & \ddots & -\frac{1}{2} \\ & & & 1 \end{pmatrix} .$$

Its inverse is given by the unitriangular matrix having $(\mathbf{R}^{-1})_{ij} = \frac{1}{2}(3/2)^{j-i-1}$ for $1 \leq i < j \leq k$.

Seysen's reduction

Seysen investigated a way to simultaneously reduce the size of the basis and of its dual [Sey93]. Specifically, in a proof [Sey93, Prop. 5], a variant of size-reduction is defined, and an algorithm is provided to reduce any

basis into such one. In this section, we will explain this algorithm and show how to phrase Seysen's reduction method [Sey93, Prop. 5] as a variant of size-reduction using block matrices and Rounding-Off.

Recently, Seysen's reduction has regained notable attention [KEF21, RH23, DPS25] for use inside lattice reduction. First, we will give the definition and the algorithm that transform a basis to a Seysen-reduced basis. Then, we will show that Seysen-reduction makes a basis more so-called 'well-conditioned' than size-reduction, This allows basis reduction to succeed using lower precision floating-point numbers.

Definition 2.95 (Seysen-reduced)**.** Let $\mathbf{B} \in \mathbb{R}^{n \times k}$ be a basis. We say $\mathbf{B}$ is *Seysen-reduced* if either $k = 1$, or parsing its R-factor as $\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}$ with $\mathbf{R}_{11} \in \mathbb{R}^{\lfloor k/2 \rfloor \times \lfloor k/2 \rfloor}$, satisfies these three properties: (1) $\mathbf{R}_{11}$ is Seysen-reduced; (2) $\mathbf{R}_{22}$ is Seysen-reduced, and (3) $\mathbf{R}_{12} \subset \mathcal{P}(\mathbf{R}_{11})$, i.e., each column of $\mathbf{R}_{12}$ is in $\mathcal{P}(\mathbf{R}_{11})$.

In addition, a basis $\mathbf{B}$ is *dual-Seysen-reduced* if either $k = 1$, or its R-factor $\mathbf{R}$, now parsed with $\mathbf{R}_{11} \in \mathbb{R}^{\lceil k/2 \rceil \times \lceil k/2 \rceil}$, i.e., larger for odd k, satisfies: (1) $\mathbf{R}_{11}$ is dual-Seysen-reduced; (2) $\mathbf{R}_{22}$ is dual-Seysen-reduced, and (3) $\mathbf{R}_{12} \subset -\mathcal{P}(\mathbf{R}_{11})$.

The reader should be warned that our definition of Seysen-reduction originates from the proof of Seysen [Sey93, Prop. 5], and is not related to what Seysen calls S-reduced nor S_2-reduced [Sey93, Sec. 5].

To provide a simpler description of our Seysen-reduction algorithm, let us extend Algorithm 2.2 to multiple targets, i.e., given $\ell \in \mathbb{Z}_{\geq 1}$ we write

$$([\mathbf{v}_1, \ldots, \mathbf{v}_\ell], [\mathbf{e}_1, \ldots, \mathbf{e}_\ell]) \leftarrow \mathrm{RO}(\mathbf{B},\ [\mathbf{t}_1, \ldots, \mathbf{t}_\ell])\ ,$$

where $(\mathbf{v}_i, \mathbf{e}_i) \leftarrow \mathrm{RO}(\mathbf{B},\ \mathbf{t}_i)$ holds for $i \in [\ell]$.

Algorithm 2.5 shows how to reduce an upper triangular matrix $\mathbf{R}$. The general case can be reduced to this case using QR decomposition.

If Line 6 of Algorithm 2.5 would call Babai's Nearest-Plane algorithm instead of Rounding-Off, then the output basis would be size-reduced, as $\mathbf{R}_{12} \subset \mathcal{P}(\mathbf{R}_{11}^*)$ implies $|\mu_{ij}| \leq \frac{1}{2}$ for all $i \leq k/2 < j \leq k$.

If a basis is Seysen-reduced, then its dual basis is dual-Seysen-reduced.

Lemma 2.96. *If a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ is Seysen-reduced, then the reversed dual basis ${}^{\vee}\mathbf{R}$ is dual-Seysen-reduced.*

Algorithm 2.5. SeysenReduce($\mathbf{R}$): Seysen-reduction of a basis

Input: upper triangular basis $\mathbf{R} \in \mathbb{R}^{k\times k}$ of Λ
Output: Seysen-reduced basis $\mathbf{R}' = \mathbf{RU}$ of Λ and $\mathbf{U} \in \mathrm{GL}_k(\mathbb{Z})$

1: **if** $\mathbf{R} \in \mathbb{R}^{1\times 1}$ **then**
2: **return** $\left(\mathbf{R}, \begin{bmatrix}1\end{bmatrix}\right)$
3: parse $\begin{bmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{0} & \mathbf{R}_{22}\end{bmatrix} = \mathbf{R}$ ▷ Blocks have half the size
4: $(\mathbf{R}'_{11}, \mathbf{U}_{11}) \leftarrow$ SeysenReduce($\mathbf{R}_{11}$)
5: $(\mathbf{R}'_{22}, \mathbf{U}_{22}) \leftarrow$ SeysenReduce($\mathbf{R}_{22}$)
6: $(\mathbf{V}, \mathbf{R}'_{12}) \leftarrow \mathrm{RO}(\mathbf{R}'_{11},\ \mathbf{R}_{12}\cdot\mathbf{U}_{22})$
7: $\mathbf{U}_{12} \leftarrow -\mathbf{R}_{11}^{-1}\cdot\mathbf{V}$ ▷ $\mathbf{R}'_{12} = \mathbf{R}_{12}\mathbf{U}_{22} - \mathbf{V}$
8: **return** $\left(\begin{bmatrix}\mathbf{R}'_{11} & \mathbf{R}'_{12}\\ \mathbf{0} & \mathbf{R}'_{22}\end{bmatrix}, \begin{bmatrix}\mathbf{U}_{11} & \mathbf{U}_{12}\\ \mathbf{0} & \mathbf{U}_{22}\end{bmatrix}\right)$

Proof. Without loss of generality, we will prove this lemma on an upper triangular R-factor $\mathbf{R} \in \mathbb{R}^{k\times k}$, and use induction on k. The case $k = 1$ is trivially satisfied.

Let us consider $k > 1$, an upper triangular basis $\mathbf{R} = \begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{0} & \mathbf{R}_{22}\end{pmatrix}$ and its lower triangular associated dual basis,

$$\mathbf{R}^\vee = \begin{pmatrix}\mathbf{S}_{11} & \mathbf{0}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix} = \mathbf{R}^{-\mathsf{T}} = \begin{pmatrix}\mathbf{R}_{11}^{-\mathsf{T}} & \mathbf{0}\\ -\mathbf{R}_{22}^{-\mathsf{T}}\mathbf{R}_{12}^{\mathsf{T}}\mathbf{R}_{11}^{-\mathsf{T}} & \mathbf{R}_{22}^{-\mathsf{T}}\end{pmatrix},$$

which is easily checked to be inverse transpose of $\mathbf{R}$. It follows by induction that the reversed submatrices ${}^\vee\mathbf{R}_{11}$ and ${}^\vee\mathbf{R}_{22}$ are Seysen-reduced. Because $\left((\mathbf{R}^\vee)_{k+1-i,k+1-j}\right)_{i,j=1}^{k}$ is the R-factor of ${}^\vee\mathbf{R}$, what remains to show is $\mathbf{S}_{21} \subset -\mathcal{P}(\mathbf{S}_{22})$. Since $\mathbf{R}$ is Seysen-reduced we know $\mathbf{R}_{12} \subset \mathcal{P}(\mathbf{R}_{11})$, or equivalently $\mathbf{R}_{11}^{-\mathsf{T}}\mathbf{R}_{12} \in (-1/2, 1/2]^{k\times k}$. By transposing and negating, we get $-\mathbf{S}_{22}^{-\mathsf{T}}\mathbf{S}_{21} = \mathbf{R}_{12}^{\mathsf{T}}\mathbf{R}_{11}^{-\mathsf{T}} \in (-1/2, 1/2]^{k\times k}$, which proves $\mathbf{S}_{21} \subset -\mathcal{P}(\mathbf{S}_{22})$. □

Lemma 2.97. *A Seysen-reduced upper triangular basis $\mathbf{R} \in \mathbb{R}^{k\times k}$ satisfies*

$$\lg(\|\mathbf{R}\|_\infty) \le \max\left(0, \frac{\lg(k/2)\lg(k/4)}{2}\right) + \max_{i=1,\dots,k}\lg(|\mathbf{R}_{ii}|)\ .$$

Proof. We follow Seysen's original proof using recursion [Sey93, Prop. 5]. By scaling $\mathbf{R}$, we assume without loss of generality that $\mathbf{R}$ is unitriangular.

For $k \leq 7$, one may check $\|\mathbf{R}\|_\infty \leq 1$ holds by close inspection. For example, for $k = 5$ the first row of $\mathbf{R}$ is given by $(1, r_2, r_3, r_4, r_5)$, with $|r_2| \leq \frac{1}{2} \cdot 1$, and $r_i \in \mathcal{P}([1, r_2])$ implies $|r_i| \leq \frac{1}{2}(1 + \frac{1}{2}) = 3/4$ for $i = 3, 4, 5$.

Now consider $\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{pmatrix}$ with $k \geq 8$. By induction, we have

$$\lg(\|\mathbf{R}_{11}\|_\infty) \leq \frac{\lg(\lfloor k/2 \rfloor/2)\lg(\lfloor k/2 \rfloor/4)}{2} \leq \frac{\lg(k/4)\lg(k/8)}{2} \quad , \quad \text{and}$$
$$\lg(\|\mathbf{R}_{22}\|_\infty) \leq \frac{\lg(\lceil k/2 \rceil/2)\lg(\lceil k/2 \rceil/4)}{2} \leq \frac{\lg(k/2)\lg(k/4)}{2} \quad .$$

We get $\|\mathbf{R}_{12}\|_\infty \leq \frac{1}{2}\lfloor k/2 \rfloor \cdot \|\mathbf{R}_{11}\|_\infty$ as each column of $\mathbf{R}_{12}$ is equal to a sum $\sum_{i=1}^{\lfloor k/2 \rfloor} c_i \cdot (\mathbf{R}_{11})_i$ with some $c_i \in (-1/2, 1/2]$. Thus,

$$\lg(\|\mathbf{R}_{12}\|_\infty) \leq \lg(k/4) + \frac{\lg(k/4)\lg(k/8)}{2} = \frac{\lg(k/2)\lg(k/4)}{2} \quad . \qquad \square$$

A dual-Seysen-reduced basis has slightly larger entries than a Seysen-reduced basis. Still, both grow as $\exp(\mathcal{O}(\log^2(k)))$ as $k \to \infty$.

Lemma 2.98. *A dual-Seysen-reduced upper triangular basis $\mathbf{R} \in \mathbb{R}^{k \times k}$ satisfies*

$$\lg(\|\mathbf{R}\|_\infty) \leq \frac{\lg(2k)\lg(k)}{2} + \max_{i=1,\ldots,k} \lg(|\mathbf{R}_{ii}|) \quad .$$

Proof. The proof proceeds similarly. For an R-factor $\mathbf{R} \in \mathbb{R}^{k \times k}$ with $k \geq 2$, using $\lceil k/2 \rceil \leq k$ we may bound

$$\lg(\|\mathbf{R}_{12}\|) \leq \lg\left(\frac{\lceil k/2 \rceil}{2}\right) + \lg(\|\mathbf{R}_{11}\|_\infty) \quad .$$

Denoting the maximum bound for a $k \times k$ R-factor by a_k, writing out the recursion

$$\begin{aligned} \lg(a_k) &\leq \lg\left(\frac{\lceil k/2 \rceil}{2}\right) + \lg(a_{\lceil k/2 \rceil}) \\ &\leq \lg\left(\frac{\lceil k/2 \rceil}{2}\right) + \lg\left(\frac{\lceil k/4 \rceil}{2}\right) + \lg(a_{\lceil k/4 \rceil}) \\ &\leq \cdots \leq \lg(a_2) + \sum_{i=0}^{\lceil \lg(k) - 2 \rceil} \lg\left(\frac{\lceil k/2^{i+1} \rceil}{2}\right) \quad , \end{aligned}$$

where we collected contributions of $\lg(\lceil \ell/2 \rceil/2)$ into a sum from $\ell = k, \lceil k/2 \rceil$, and so on up to the last term with $\ell = \lceil k/2^{\lceil \lg(k)-2 \rceil} \rceil$ ($\in \{3,4\}$). As $a_2 = 1$, we have

$$\begin{aligned}\lg(a_k) &\leq \sum_{i=1}^{\lceil \lg(k)-1 \rceil} \lg\left(\frac{\lceil k/2^i \rceil}{2}\right) \leq \sum_{i=1}^{\lceil \lg(k)-1 \rceil} \lg\left(\frac{k}{2^i}\right) \\ &\leq \lg^2(k) - (1+2+\cdots+\lfloor \lg(k) \rfloor) \leq \frac{\lceil \lg(k)-1 \rceil \cdot \lceil \lg(k) \rceil}{2} \\ &\leq \frac{\lg(k)(2\lg(k) - \lceil \lg(k) \rceil)}{2} \leq \frac{\lg(2k)\lg(k)}{2} \quad . \end{aligned}$$

□

Lemmata 2.96, 2.97 and 2.98 now lead to the following conclusion.

Corollary 2.99. *Given a Seysen-reduced basis $\mathbf{R} \in \mathbb{R}^{k \times k}$, its condition number is bounded by*

$$\kappa(\mathbf{R}) \leq k^2 2^{\ln^2(k)} \cdot \frac{\max_{i \in [k]} |\mathbf{R}_{ii}|}{\min_{i \in [k]} |\mathbf{R}_{ii}|} \quad .$$

2.5.3 Lagrange reduction

A basis for a rank-2 lattice Λ can be fully reduced in a polynomial number of operations.

Definition 2.100 (Lagrange-reduced)**.** A basis $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^n$ of a lattice Λ is *Lagrange-reduced* if $\|\mathbf{b}_1\| = \lambda_1(\Lambda)$ and $|\langle \mathbf{b}_1, \mathbf{b}_2 \rangle| \leq \frac{1}{2}\|\mathbf{b}_1\|^2$ hold.

Algorithm 2.6 finds a Lagrange-reduced basis in a polynomial number of operations. We remark that $(\_, \mathbf{b}_2) \leftarrow \mathrm{RO}([\mathbf{b}_1], \mathbf{b}_2)$ is equivalent to what is written on Lines 1 and 4. The same goes for NP as $\mathcal{P}([\mathbf{b}_1]) = \mathcal{P}([\mathbf{b}_1^*])$. Figure 2.5 visualises a Lagrange-reduced basis.

Lemma 2.101. *On input a basis $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^n$ of a lattice Λ, Algorithm 2.6 executes the* **while** *loop at most* $\max(1, \lg(\|\mathbf{b}_1\|^2/\det(\Lambda)))$ *times and then outputs a Lagrange-reduced basis for Λ.*

Proof. First, Algorithm 2.6, if it terminates, outputs a basis for Λ as any update of the basis $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2]$ is of the form $\mathbf{B} \leftarrow \mathbf{BU}$ with $\mathbf{U} \in \mathrm{GL}_2(\mathbb{Z})$. Specifically, Line 3 uses $\mathbf{U}_{\text{swap}} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, while Lines 1 and 4 use $\mathbf{U}_c = \left(\begin{smallmatrix} 1 & -c \\ 0 & 1 \end{smallmatrix}\right)$ when updating $\mathbf{b}_2 \leftarrow \mathbf{b}_2 - c\mathbf{b}_1$ for some $c \in \mathbb{Z}$.

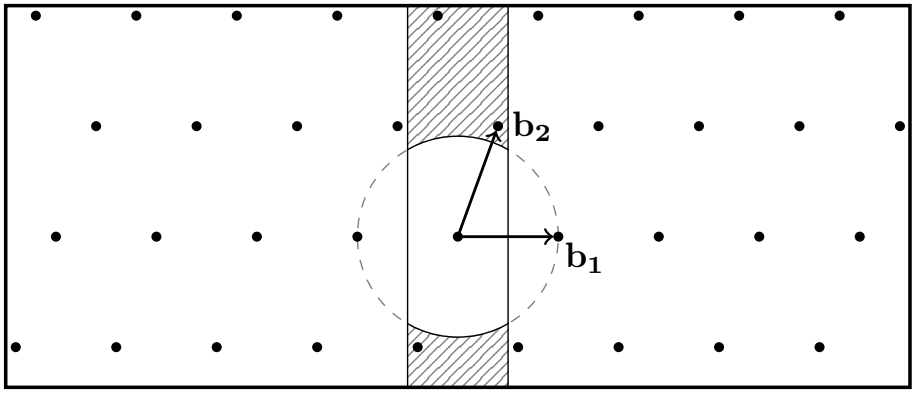


Figure 2.5: Possible final state of Algorithm 2.6

Algorithm 2.6. LagrangeReduce$(\mathbf{b}_1, \mathbf{b}_2)$: Lagrange basis reduction

Input: basis $[\mathbf{b}_1, \mathbf{b}_2]$ of Λ
Output: Lagrange-reduced basis $[\mathbf{b}_1, \mathbf{b}_2]$ of Λ

1: $\mathbf{b}_2 \leftarrow \mathbf{b}_2 - \lceil \langle \mathbf{b}_1, \mathbf{b}_2 \rangle / \|\mathbf{b}_1\|^2 \rfloor \cdot \mathbf{b}_1$
2: **while** $\|\mathbf{b}_2\| < \|\mathbf{b}_1\|$ **do**
3: swap $\mathbf{b}_1 \leftrightarrows \mathbf{b}_2$
4: $\mathbf{b}_2 \leftarrow \mathbf{b}_2 - \lceil \langle \mathbf{b}_1, \mathbf{b}_2 \rangle / \|\mathbf{b}_1\|^2 \rfloor \cdot \mathbf{b}_1$
5: **return** $\mathbf{b}_1, \mathbf{b}_2$

Directly after executing Line 1 or Line 4, Rounding-Off guarantees $\pi_{\mathrm{span}(\mathbf{b}_1)}(\mathbf{b}_2) \in \mathcal{P}([\mathbf{b}_1])$ holds, so we have $|\langle \mathbf{b}_1, \mathbf{b}_2 \rangle| \leq \frac{1}{2}\|\mathbf{b}_1\|^2$. The output thus satisfies this as well.

Finally, we bound the number of loop executions and prove that $\|\mathbf{b}_1\| = \lambda_1(\Lambda)$ holds on termination. Note $\|\mathbf{b}_1\|$ strictly decreases in each iteration, and we have the following loop invariant:

$$\det(\Lambda) = \det[\mathbf{b}_1, \mathbf{b}_2] = \|\mathbf{b}_1\| \cdot \|\mathbf{b}_2^*\| \ .$$

Assume after executing Line 3, we have $\|\mathbf{b}_1\|^2 \geq 2 \cdot \det(\Lambda)$. Using the loop invariant and this assumption, $\|\mathbf{b}_2^*\| = \det(\Lambda)/\|\mathbf{b}_1\| \leq \|\mathbf{b}_1\|/2$. Line 4 leaves $\mathbf{b}_2^*$ untouched, while the reduction now guarantees

$$\|\mathbf{b}_2\|^2 \leq \left\| \frac{\mathbf{b}_1}{2} \right\|^2 + \|\mathbf{b}_2^*\|^2 \leq \frac{\|\mathbf{b}_1\|^2}{4} + \frac{\|\mathbf{b}_1\|^2}{4} = \frac{1}{2}\|\mathbf{b}_1\|^2 \ .$$

Thus, one iteration significantly reduces the squared norm of $\mathbf{b}_1$, and the assumption $\|\mathbf{b}_1\|^2 \geq 2 \cdot \det(\Lambda)$ cannot hold for many iterations. Namely, on input $(\mathbf{b}_1^\circ, \mathbf{b}_2^\circ)$, if after k iterations $\|\mathbf{b}_1\|^2 \geq 2 \cdot \det(\Lambda)$ still holds, then we must have $\|\mathbf{b}_1\|^2 \leq 2^{-k}\|\mathbf{b}_1^\circ\|^2$. To prevent a contradiction here, we conclude $\|\mathbf{b}_1\|^2 < 2 \cdot \det(\Lambda)$ holds at the start ($k = 0$) or after $k \leq \lg(\|\mathbf{b}_1^\circ\|^2/(2 \cdot \det \Lambda))$ iterations.

We can conclude $\|\mathbf{b}_1\|^2 < 2 \cdot \det(\Lambda)$ holds after k iterations. Let $\mathbf{v} = \alpha\mathbf{b}_1 + \beta\mathbf{b}_2$ be a shortest nonzero vector of Λ with $\alpha, \beta \in \mathbb{Z}$. If $\|\mathbf{v}\| = \|\mathbf{b}_1\|$, the algorithm terminates with $\|\mathbf{b}_1\| = \lambda_1(\Lambda)$ as there are no nonzero lattice vectors of strictly smaller norm, and the lemma is proved.

Otherwise, $\|\mathbf{v}\| < \|\mathbf{b}_1\|$, and we have $\beta \neq 0$ as $\|\alpha\mathbf{b}_1\| \geq \|\mathbf{b}_1\|$ for nonzero α. The proof of Lemma 2.57 shows $\|\mathbf{v}\| \geq |\beta| \cdot \|\mathbf{b}_2^*\| > |\beta|\sqrt{\det(\Lambda)/2}$, where we use $\|\mathbf{b}_2^*\| = \det(\Lambda)/\|\mathbf{b}_1\| > \sqrt{\det(\Lambda)/2}$ in the second inequality. Because we also have $\|\mathbf{v}\| \leq \|\mathbf{b}_1\| < \sqrt{2\det(\Lambda)}$ and $\beta \in \mathbb{Z}$, we conclude $\beta \in \{-1, 1\}$. Let us assume $\beta = 1$ without loss of generality as we may negate $\mathbf{v}$. Now, $\mathbf{v} = \alpha\mathbf{b}_1 + \mathbf{b}_2$ has a squared norm

$$\begin{aligned}\|\mathbf{v}\|^2 &= \alpha^2\|\mathbf{b}_1\|^2 + \|\mathbf{b}_2\|^2 - 2\alpha \cdot \langle \mathbf{b}_1, \mathbf{b}_2 \rangle \\ &\geq (\alpha^2 - |\alpha|) \cdot \|\mathbf{b}_1\|^2 + \|\mathbf{b}_2\|^2 \geq \|\mathbf{b}_2\|^2 \ ,\end{aligned}$$

using $|\langle \mathbf{b}_1, \mathbf{b}_2 \rangle| \leq \frac{1}{2}\|\mathbf{b}_1\|^2$ and $\alpha^2 \geq |\alpha|$ for $\alpha \in \mathbb{Z}$. As $\mathbf{v}$ is a shortest vector of Λ, we conclude $\|\mathbf{b}_2\| = \|\mathbf{v}\| < \|\mathbf{b}_1\|$. Thus, the algorithm continues with iteration $k + 1$. During this iteration, the swap on Line 3 executes $\mathbf{b}_1 \leftarrow \mathbf{b}_2$, making $\mathbf{b}_1$ a shortest vector as $\|\mathbf{b}_1\| = \|\mathbf{v}\| = \lambda_1(\Lambda)$. Thus, the algorithm will return the shortest vector after at most $k + 1$ iterations. □

This Lemma shows that the algorithm can reduce an integer basis $\mathbf{B} \in \mathbb{Z}^{2\times 2}$ in at most $2\lg\|\mathbf{b}_1\|$ iterations, and each iteration performs an arithmetic operation on $\mathcal{O}(\log(\|\mathbf{B}\|_\infty))$ bit numbers, so the overall running time is $\mathcal{O}(\log^3(\|\mathbf{B}\|_\infty))$. More efficient algorithms [Sch91, Yap92] can achieve a runtime of $\mathcal{O}(L \log^{2+\varepsilon} L)$ where $L = \log(\|\mathbf{B}\|_\infty)$ and $\varepsilon > 0$, using fast integer arithmetic [SS71].

2.5.4 Lenstra–Lenstra–Lovász

Lenstra, Lenstra and Lovász provided a polynomial-time basis reduction algorithm [LLL82] that finds a short lattice point of norm at most $2^{k-1}\lambda_1(\Lambda)$ in a rank-k lattice.

LLL reduction has numerous applications such as factoring integer polynomials [LLL82, vH02, Klü10], attacking knapsack-based public key cryptosystems [LO83], and it was used to disprove Mertens' conjecture [OtR85]. A complete book has been written on LLL reduction [NV10]. Moreover, LLL is a fundamental tool in lattice-based cryptography and is used by stronger basis reduction algorithms.

First, we define what an LLL-reduced basis is. Then, we give an algorithm that reduces a basis to such basis using a combination of size-reduction and Lagrange reduction.

Definition 2.102 (LLL-reduced)**.** Let $\mathbf{B} \in \mathbb{R}^{n\times k}$ be a basis and $\delta \in (1/4, 1]$. Then, $\mathbf{B}$ is *δ-LLL-reduced* if it is size-reduced and it satisfies *Lovász' condition*, i.e., for all $i \in [k-1]$, we have

$$\delta\|\pi_i(\mathbf{b}_i)\|^2 \leq \|\pi_i(\mathbf{b}_{i+1})\|^2.$$

A 1-LLL-reduced basis of rank two is Lagrange-reduced. An LLL-reduced basis has Gram–Schmidt norms that do not decrease by too much, i.e., writing $\pi_i(\mathbf{b}_{i+1}) = \mu_{i,i+1}\mathbf{b}_i^* + \mathbf{b}_{i+1}^*$, we have

$$\|\mathbf{b}_{i+1}^*\|^2 = \|\pi_i(\mathbf{b}_{i+1})\|^2 - \mu_{i,i+1}^2\|\mathbf{b}_i^*\|^2 \geq \left(\delta - \frac{1}{4}\right)\|\pi_i(\mathbf{b}_i)\|^2 \ .$$

This has the following corollary. We note that size-reduction is a stronger requirement than necessary, and we only use Lovász' condition and $|\mu_{i,i+1}| \leq \frac{1}{2}$, which is also implied by Seysen's reduction.

Corollary 2.103. *Let $\delta \in (1/4, 1]$ and $\gamma = 1/\sqrt{\delta - \frac{1}{4}} \geq \sqrt{4/3}$. A δ-LLL-reduced basis $\mathbf{B} \in \mathbb{R}^{n\times k}$ of a lattice Λ satisfies:*

$$\forall i \in [k]\colon \quad \|\mathbf{b}_i^*\| \leq \gamma^{k-i} \cdot \lambda_i(\Lambda) \ , \tag{2.11}$$

$$\|\mathbf{b}_1\| \leq \gamma^{(k-1)/2} \cdot \det(\Lambda)^{1/k} \ . \tag{2.12}$$

Proof. The paragraph above shows $\|\mathbf{b}_j^*\| \leq \gamma\|\mathbf{b}_{j+1}^*\|$ for $j \in [k-1]$, so using induction we get $\|\mathbf{b}_i^*\| \leq \gamma^{j-i} \cdot \|\mathbf{b}_j^*\|$ for all $1 \leq i \leq j \leq k$.

Equation (2.11)**:** Let $\mathbf{v}_1, \ldots, \mathbf{v}_i \in \Lambda$ be a basis for a sublattice of Λ with $\|\mathbf{v}_j\| \leq \lambda_i(\Lambda)$ for all $j \in [i]$. Since the $\mathbf{v}_1, \ldots, \mathbf{v}_i$ generate a rank-i lattice, there exists $j \in [i]$ such that $\pi_i(\mathbf{v}_j) \neq \mathbf{0}$. By Lemma 2.57 we get

$$\lambda_i(\Lambda) \geq \|\mathbf{v}_j\| \geq \|\pi_i(\mathbf{v}_j)\| \geq \min_{i\leq j\leq k}\|\mathbf{b}_j^*\| \geq \|\mathbf{b}_i^*\|/\gamma^{k-i+1} \ .$$

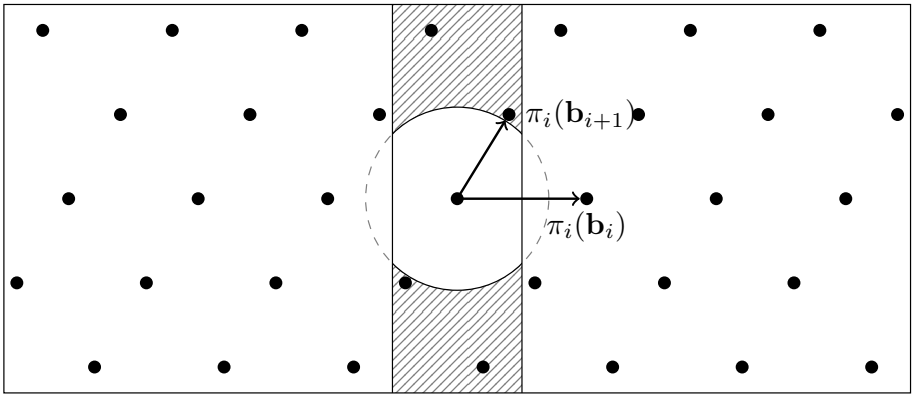


Figure 2.6: Situation in which the **if** condition on Line 4 is satisfied, cf. Figure 2.5. The radius of the circle is $\|\mathbf{b}_i^*\|\sqrt{\delta}$ where $\delta = \frac{1}{2}$.

Equation (2.12)**:** For all $i \in [k]$ we have $\|\mathbf{b}_1\| \leq \gamma^{i-1}\|\mathbf{b}_i^*\|$ so

$$\|\mathbf{b}_1\|^k \leq \prod_{i=1}^{k}\left(\gamma^{i-1}\|\mathbf{b}_i^*\|\right) = \gamma^{k(k-1)/2}\det(\Lambda) \ . \qquad \square$$

In particular, taking $\delta = 1/4$ as in the original LLL paper, yields $\gamma = 2$ and Equation (2.11) becomes $\|\mathbf{b}_1\| \leq 2^{k-1}\lambda_1(\Lambda)$.

A naïve approach to make a basis LLL-reduced is to perform Lagrange reduction on every rank-2 projected sublattice that is not Lagrange-reduced. However, such an algorithm may not terminate in polynomial time as Lagrange reduction may provide too small improvements, causing iteration over a super-polynomial number of lattice points of norm at most $\max_i\|\mathbf{b}_i\|$. The ingenious idea to run in polynomial time, is to *only* perform a Lagrange reduction if $\|\mathbf{b}_i^*\|$ can be improved by some fixed factor $\delta < 1$, which is depicted in Figure 2.6.

Algorithm 2.7 computes an LLL-reduced basis. The algorithm requires $\delta < 1$ to reduce an integer basis in polynomial time. Note that a basis can be size-reduced by Algorithm 2.4, while LLL requires many iterations before a basis satisfies Lovász' condition.

Theorem 2.104 ([LLL82, Prop. (1.26)])**.** *Let $\mathbf{B} \in \mathbb{Z}^{n\times k}$ be an integer basis, $\delta \in (1/4, 1)$ fixed and let $L \in \mathbb{R}$ be an upper bound on all basis vector norms, i.e., for all $i \in [k]$ we have $\|\mathbf{b}_i\| \leq L$. On input an integer basis $\mathbf{B} \in \mathbb{Z}^{n\times k}$ and fixed $\delta \in (1/4, 1)$, Algorithm 2.7 provides an LLL-reduced basis, by using at most $\mathcal{O}(nk^3\log(L))$ arithmetic operations on integers, and these integers have a binary length $\mathcal{O}(k\log L)$.*

Algorithm 2.7. LLL$(\mathbf{B}, \delta)$: LLL basis reduction

Input: basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ of Λ and $\delta \in (1/4, 1)$
Output: LLL-reduced basis $\mathbf{B}$ of Λ

1: $i \leftarrow 1$
2: **while** $i < k$ **do** ▷ invariant: $[\mathbf{b}_1, \ldots, \mathbf{b}_i]$ is LLL-reduced
3: $(\_, \mathbf{b}_{i+1}) \leftarrow \mathrm{NP}([\mathbf{b}_1, \ldots, \mathbf{b}_i], \mathbf{b}_{i+1})$ ▷ reduce $\mathbf{b}_{i+1}$
4: **if** $\delta \|\pi_i(\mathbf{b}_i)\|^2 \leq \|\pi_i(\mathbf{b}_{i+1})\|^2$ **then** ▷ check Lovász' condition
5: $i \leftarrow i + 1$
6: **else**
7: swap $\mathbf{b}_i \leftrightarrows \mathbf{b}_{i+1}$
8: $i \leftarrow \max\{i - 1, 1\}$
9: **return** $\mathbf{B}$

Note, using naïve integer multiplication, this leads to $\mathcal{O}(nk^5 \log^3(L))$ bit operations. This theorem can be proven using the following potential function.

Definition 2.105. The *potential of a basis* $\mathbf{B} \in \mathbb{R}^{n \times k}$ is

$$\mathrm{Pot}(\mathbf{B}) = \prod_{i=1}^{k} \det(\mathbf{B}_{[1,i]}^{\mathsf{T}} \cdot \mathbf{B}_{[1,i]}) = \|\mathbf{b}_1^*\|^{2k} \cdot \|\mathbf{b}_2^*\|^{2k-2} \cdot \cdots \cdot \|\mathbf{b}_k^*\|^2 \ .$$

Proof of Theorem 2.104. The **while** loop on Line 2 has the invariant that is mentioned, and because it holds for $i = k$ on termination, the output is LLL-reduced.

Technically, one would need to show the GSO can be efficiently computed for size-reduction, and for computing $\pi_i(\mathbf{b}_i)$ and $\pi_i(\mathbf{b}_{i+1})$. As the input basis is integral, the squared norm of each Gram–Schmidt vector is rational, and thus can be handled by exact integer arithmetic. Such technical details can be found in the original paper [LLL82]. Instead, we will derive an upper bound on the number of swaps using the potential. Specifically, we will show the potential (1) is bounded from above, (2) is bounded from below by 1 and (3) decreases by a factor δ for each swap, from which the bound on the number of arithmetic operations can be derived.

Initially, we have $\|\mathbf{b}_i^*\| \leq \|\mathbf{b}_i\| \leq L$ so we may bound the potential by $\mathrm{Pot}(\mathbf{B}) \leq L^{k(k+1)}$, which shows (1).

Suppose at some point in the execution, the algorithm works with a basis $\mathbf{C}$. Such a basis has to be an integer basis $\mathbf{C} \in \mathbb{Z}^{n\times k}$ so the Gram matrix $\mathbf{C}_{[1,i]}^{\mathsf{T}} \cdot \mathbf{C}_{[1,i]} \in \mathbb{Z}^{i\times i}$ is an integer. This shows that each factor of $\mathrm{Pot}(\mathbf{C})$ is a positive integer, and thus $\mathrm{Pot}(\mathbf{C}) \in \mathbb{Z}_{\geq 1}$, which shows (2).

Note that size-reduction is a basis transformation on each $\mathbf{B}_{[1,i]}$ so it
2 does not affect $\mathrm{Pot}(\mathbf{B})$. Now suppose the basis is updated from $\mathbf{B}$ to $\mathbf{C}$ by swapping vectors $\mathbf{b}_i$ and $\mathbf{b}_{i+1}$. This means the check on Line 4 failed, so we have $\|\pi_i(\mathbf{b}_{i+1})\|^2 < \delta\|\mathbf{b}_i^*\|^2$, and therefore

$$\begin{aligned}\det(\mathbf{C}_{[1,i]}^{\mathsf{T}}\mathbf{C}_{[1,i]}) = \prod_{j=1}^{i}\|\mathbf{c}_j^*\|^2 &= \|\pi_i(\mathbf{b}_{i+1})\|^2 \cdot \prod_{j=1}^{i-1}\|\mathbf{b}_j^*\|^2 \\ &< \delta\prod_{j=1}^{i}\|\pi_j(\mathbf{b}_j)\|^2 = \delta\det(\mathbf{B}_{[1,i]}^{\mathsf{T}}\mathbf{B}_{[1,i]}) \ .\end{aligned}$$

As $\mathbf{C}_{[1,j]}$ is unchanged for $j < i$, and transformed by some $\mathrm{GL}_k(\mathbb{Z})$ for $j \geq i+1$, going from $\mathbf{B}$ to $\mathbf{C}$, the potential only changes in the jth factor. That is, $\mathrm{Pot}(\mathbf{C}) < \delta\,\mathrm{Pot}(\mathbf{B})$, which shows (3).

Hence, Algorithm 2.7 performs at most $k(k+1)\cdot\log_{1/\delta}(L)$ swaps. □

Algorithm 2.7 can be altered to take as input any generating set $\mathbf{B} \in \mathbb{Z}^{n\times k}$ such that the output is an LLL-reduced basis for the same lattice [Poh87], by detecting linear relations during LLL reduction.

LLL-reducedness is a self-dual property, in the following sense.

Proposition 2.106. *Given a δ-LLL-reduced basis $\mathbf{B} \in \mathbb{R}^{n\times k}$, size-reducing the reverse dual ${}^{\vee}\mathbf{B}$ yields a δ-LLL-reduced basis.*

Proof. The Lovász' condition is a *local* property, i.e., it is a statement on the projected sublattices $\mathbf{B}_{[i,i+1]}$ of rank two, so by Proposition 2.77 the claim follows from the case $k = 2$.

Without loss of generality consider a δ-LLL-reduced basis and its reverse dual in dimension 2, given by

$$\mathbf{B} = \begin{pmatrix} 1 & \mu_{12} \\ 0 & a \end{pmatrix} \quad , \quad \text{and} \quad {}^{\vee}\mathbf{B} = [\mathbf{d}_1, \mathbf{d}_2] = \begin{pmatrix} 0 & 1 \\ \frac{1}{a} & -\frac{\mu_{12}}{a} \end{pmatrix} \quad ,$$

respectively. Because $\mathbf{B}$ is δ-LLL-reduced, we have $|\mu_{12}| \leq \frac{1}{2}$ and $\delta \leq \mu_{12}^2 + a^2$. Hence, we have $\delta\|\mathbf{d}_1\|^2 = \delta/a^2 \leq \mu_{12}^2/a^2 + 1 = \|\mathbf{d}_2\|^2$. □

Floating-point variants

Although Theorem 2.104 shows that LLL reduction runs in polynomial time, it may be too slow in practice to LLL-reduce an integer basis of high dimension using exact integer arithmetic. In this case, one may speed up the algorithm using floating-point numbers to compute the Gram–Schmidt vectors and coefficients [SE94]. However, floating-point arithmetic may lead to numerical stability issues and the algorithm may not terminate or may not output an LLL-reduced basis anymore. Nguyen and Stehlé [NS09] developed a floating-point variant of LLL reduction requiring

$$\mathcal{O}(nk^4(k + \log L) \log L)$$

bit operations using naïve multiplication. This algorithm was the first with a runtime depending quadratically on $\log L$. Moreover, this floating-point variant has been implemented in the fplll software, and is the most widely used lattice reduction software nowadays.

This algorithm computes floating-point approximations of Gram–Schmidt coefficients using the Cholesky decomposition of the Gram matrix. Alternatively, there is an L^2 variant [MSV09] that uses Householder reflections to compute the QR decomposition. These Householder reflections were assumed to be numerically more stable [KS01b, Sch06] as given by a perturbation analysis [CSV12]. This 'H-LLL' algorithm has been implemented in fplll thereafter. Stehlé [Ste10] provides more background on this topic.

There is still active research [KEF21, RH23, DPS25] in designing faster LLL reduction algorithms using a combination of segments [KS01a, KS01b, NS16], iterative compression, Seysen's reduction and linear algebra libraries.

2.5.5 Block Korkine–Zolotarev

While the first basis vector in Corollary 2.103 has a norm that is exponential factor larger than that of a shortest vector, one may obtain smaller factors by spending more time on basis reduction. Algorithm 2.8 contains pseudocode for this so-called *block Korkine–Zolotarev* (BKZ) algorithm [Sch87, SE94]. This algorithm takes a block size β as input that allows to control the runtime and the approximation factor. Its runtime is, depending on the implementation of the SVP oracle, at least exponential in the block size β.

Algorithm 2.8. BKZ($\mathbf{B}, \beta$): BKZ basis reduction

Input: basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ of Λ and block size $2 \leq \beta \leq k$
Output: β-BKZ-reduced basis $\mathbf{B}$ of Λ

1: $\mathbf{B} \leftarrow \mathrm{LLL}(\mathbf{B},\ 0.99)$
2: **repeat**
3: **for** $i = 1, 2, \ldots, k-1$ **do** ▷ one tour
4: $j \leftarrow \min(k, i+\beta-1)$
5: $\mathbf{v} \leftarrow \mathrm{SVP}(\mathbf{B}_{[i,j]})$
6: $(\mathbf{w}, \_) \leftarrow \mathrm{NP}\left(\mathbf{B}_{[1,j]},\ \mathbf{v}\right)$ ▷ lift $\mathbf{v}$ to $\mathbf{w} \in \Lambda$
7: **if** $\mathbf{w} \neq \pm\mathbf{b}_i$ **then**
8: $\mathbf{B}' \leftarrow [\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, \mathbf{w}, \mathbf{b}_i, \ldots, \mathbf{b}_k]$ ▷ insert $\mathbf{w}$
9: $\mathbf{B} \leftarrow \mathrm{LLL}(\mathbf{B}',\ 0.99)_{[1,k]}$ ▷ remove linear dependence
10: **until** $\mathbf{B}$ has not changed
11: **return** $\mathbf{B}$

Line 5 calls an oracle SVP that, on input $\mathbf{C}$, is able to find a shortest vector in the lattice $\mathcal{L}(\mathbf{C})$ of rank at most β. This oracle can be realised using lattice point enumeration [GNR10], or lattice sieving [AKS01]. The runtime grows exponentially as a function of the block size β, even when require the oracle to output a lattice point of a random lattice having norm $1.05\,\lambda_1(\Lambda)$, for example.

Line 8 inserts the vector $\mathbf{w}$ into the basis $\mathbf{B}$ at index i, causing $\mathbf{B}'$ to be rank-deficient as $\mathbf{w}$ is a linear combination of $\mathbf{b}_1, \ldots \mathbf{b}_j$. Line 9 turns $\mathbf{B}$ back into a basis by eliminating the linear dependency using Pohst's modification of LLL [Poh87]. That is, column $k+1$ of matrix $\mathbf{B}'$ is the zero vector after LLL-reduction.

Algorithm 2.8 performs many tours because of Line 10. Yet, the first tours improve the basis the most. Hence, one limits the number of tours in practice and the norm of the first basis vector can still be bounded heuristically [HPS11].

Lemma 2.107 ([LN25])**.** *Suppose Algorithm 2.8 receives as input a basis $\mathbf{B}' \in \mathbb{R}^{n \times k}$ and a block size β. After $\Theta(k^2 \log(k)/\beta^2)$ executions of the loop on Line 2, the first basis vector $\mathbf{b}_1$ of the reduced basis $\mathbf{B}$ has norm,*

$$\|\mathbf{b}_1\| \leq \gamma_\beta^{\frac{k-1}{2(\beta-1)} + \frac{\beta(\beta-2)}{2k(\beta-1)}} \cdot \det(\Lambda)^{1/k} \quad ,$$

where γ_β is defined in Definition 2.17.

Proposition 2.92 shows one needs short Gram–Schmidt vectors to have a good basis. Therefore, we use the basis profile to determine the reduction quality.

Definition 2.108 (basis profile and quality)**.** The *profile* of a basis $\mathbf{B} \in \mathbb{R}^{n\times k}$ is the sequence $(\lg\|\mathbf{b}_i^*\|)_{i=1}^k$. We consider three quality measures:

(1) the *kth root Hermite factor* (RHF) given by

$$\delta(\mathbf{B}) = \frac{\|\mathbf{b}_1\|^{1/k}}{(\det\Lambda)^{1/k^2}} \ ,$$

not to be confused with the parameter of Algorithm 2.7;

(2) the *slope* of the line given by simple linear regression on the graph of the basis profile, i.e.,

$$\mathrm{sl}(\mathbf{B}) = \frac{\sum_{i=1}^k \lg\|\mathbf{b}_i^*\| \left(i - \frac{k+1}{2}\right)}{\sum_{i=1}^k \left(i - \frac{k+1}{2}\right)^2} \ , \text{ and}$$

(3) the *potential* as in Definition 2.105.

For random lattices, the slope is usually a negative number as Gram–Schmidt norms generally decrease. Intuitively, a good basis has a small potential, which corresponds to a slope that is close to zero or positive.

Proposition 2.109. *Given a basis $\mathbf{B} \in \mathbb{R}^{n\times k}$, we have*

$$\lg \mathrm{Pot}(\mathbf{B}) + \frac{(k+1)k(k-1)}{6}\mathrm{sl}(\mathbf{B}) = (k+1)\lg(\det\Lambda) \ .$$

Proof. Let us denote $S_k = \sum_{i=1}^k \left(i - \frac{k+1}{2}\right)^2$ for the denominator in the definition of $\mathrm{sl}(\mathbf{B})$. It is easy to verify $S_k = \frac{(k+1)k(k-1)}{12}$ because S_k is a cubic polynomial in k that equals the given identity when plugging in $k = 0, 1, 2$ and $k = 3$. Therefore,

$$2S_k \cdot \mathrm{sl}(\mathbf{B}) = \sum_{i=1}^k 2i\lg\|\mathbf{b}_i^*\| - (k+1)\cdot\lg(\det\Lambda) \qquad \text{and}$$

$$\lg\mathrm{Pot}(\mathbf{B}) = 2(k+1)\cdot\lg(\det\Lambda) - \sum_{i=1}^k 2i\lg\|\mathbf{b}_i^*\| \ .$$

As a consequence we get $\lg\mathrm{Pot}(\mathbf{B}) + 2S_k\cdot\mathrm{sl}(\mathbf{B}) = (k+1)\cdot\lg(\det\Lambda)$. □

An LLL-reduced basis using LLL-parameter $\delta = 0.99$ has $\delta(\mathbf{B}) < 1.08$ as can be computed from Equation (2.12). For a random lattice, the expected RHF of an LLL-reduced basis is $\delta(\mathbf{B}) \approx 1.02$ [NS06].

Using the bound on γ_β and $\beta(\beta-2)/k(\beta-1) < 1$, Lemma 2.107 proves a BKZ-reduced basis has an RHF bounded by

$$\delta(\mathbf{B}) \leq (2\,\mathrm{GH}(\beta))^{\frac{1}{\beta-1}+\frac{1}{k}} \quad . \tag{2.13}$$

For random lattices, heuristically a BKZ-reduced basis has a smaller RHF than the worst-case bound [GN08], similar to the RHF after LLL-reduction.

Heuristic 2.110 (BKZ behaviour)**.** *Let β be fixed. On input block size β and some basis $\mathbf{B}' \in \mathbb{R}^{n\times n}$ of a full-rank random lattice Λ, Algorithm 2.8 outputs a basis $\mathbf{B}$ having RHF $\delta(\mathbf{B}) \sim \mathrm{GH}(\beta)^{1/(\beta-1)}$ as $n \to \infty$.*

See Chen's thesis for a detailed justification [Che13, Sec. 4.1].

Heuristic Justification. In short, the SVP oracle guarantees that $\mathbf{b}_i^*$ is a shortest vector of the lattice generated by the basis $\mathbf{B}_{[i,i+\beta-1]}$ for any $1 \leq i \leq n-\beta$. All these rank-β lattices are random, so we may use Heuristic 2.41 to obtain

$$\log\|\mathbf{b}_i^*\| = \log \mathrm{GH}(\beta) + \frac{1}{\beta}\sum_{j=i}^{i+\beta-1} \log\|\mathbf{b}_j^*\| \quad .$$

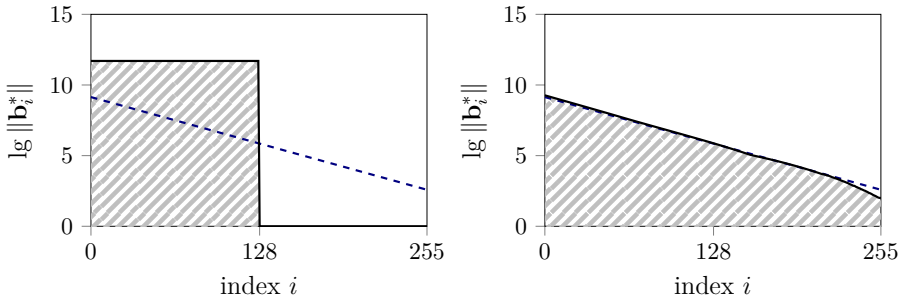
Writing $a_i = \log\|\mathbf{b}_i^*\|$ and $c = \frac{\beta}{\beta-1}\log \mathrm{GH}(\beta)$, there is a recurrence relation,

$$a_i = c + \frac{a_{i+1} + \cdots + a_{i+\beta-1}}{\beta-1} \quad .$$

A solution to this recurrence is given by $a_{i+1} = a_i - 2c/\beta$ for $i \geq 1$. Approximating the last β Gram–Schmidt norms by the solution of the recurrence yields

$$\begin{aligned}\log\left(\frac{\|\mathbf{b}_1\|^{1/n}}{\det(\Lambda)^{1/n^2}}\right) &= \frac{a_1}{n} - \frac{1}{n^2}\sum_{i=1}^{n} a_i = \frac{2c}{\beta n^2}(1+2+\cdots+n) \\ &= \frac{c(n+1)}{\beta n} \xrightarrow{(\text{as } n\to\infty)} \frac{c}{\beta} = \frac{\log(\mathrm{GH}(\beta))}{\beta-1} \quad . \qquad \triangle\end{aligned}$$

After BKZ-reduction, one may approximate the basis profile by a line [Sch03], see Figure 2.7 for motivation.



(a) Input (b) Output

Figure 2.7: Example of the basis profile before (a) and after (b) BKZ reduction using a q-ary lattice of dimension 256. The dashed blue line shows the GSA prediction.

Heuristic 2.111 (geometric series assumption, GSA)**.** *On input a block size β and a basis for a random lattice, Algorithm 2.8 outputs a basis* $\mathbf{B}$ *having Gram–Schmidt norms that form a geometric series. Equivalently, the basis profile lies on a line.*

It is known that the GSA does not accurately predict the last β Gram–Schmidt norms [AD21], which is also clear from Figure 2.7. Additionally, some of the first Gram–Schmidt norms are smaller than the GSA prediction when β is small [BSW18].

BKZ 2.0 [CN11] is a variant of BKZ consisting of four improvements:

(1) Early aborts [HPS11], which significantly improves over plain BKZ as the number of calls to the SVP oracle grows exponentially [GN08].
(2) Sound pruning [GNR10], i.e., use a pruned lattice point enumeration tree, introducing a probability that the SVP oracle is successful. If the SVP oracle was unsuccessful, we retry using a rerandomised local basis.
(3) Preprocessing of local bases, i.e., a local basis is reduced by BKZ using a smaller block size $\beta' < \beta$ instead of LLL.
(4) Taking a shorter enumeration radius, i.e., the SVP oracle on Line 5 is asked to find a vector of an approximate length according to Heuristic 2.41 that is shorter than $\|\mathbf{b}_i^*\|$.

BKZ 2.0 is implemented in fplll. Moreover, there exist models for predicting the Gram–Schmidt norms when using BKZ 2.0 to reduce a

basis for a random lattice [CN11, BSW18, DDGR20, PV21].

HKZ reduction

Algorithm 2.8 is called HKZ-reduction for $\beta = k$.

2

Definition 2.112 (HKZ-reduced)**.** A basis $\mathbf{B} \in \mathbb{R}^{n\times k}$ of a lattice Λ is *Hermite–Korkine–Zolotarev-reduced* (HKZ-reduced) if it is size-reduced, and

$$\forall i \in [k]\colon \|\mathbf{b}_i^*\| = \lambda_1(\pi_i(\Lambda)) \ . \tag{2.14}$$

This notion of reduction is named after Hermite [Her1850] and Korkine–Zolotarev [KZ1873] for their work on size-reduction and Equation (2.14), respectively.

HKZ-reduction of a lattice is computationally expensive as the oracle is asked to solve SVP in dimension k. Note that the basis is already HKZ-reduced when Algorithm 2.8 reaches Line 10. Thus, for $\beta = k$, after performing one tour in Algorithm 2.8, the basis is HKZ-reduced and the algorithm can return the basis.

Note that BKZ-reduction using block size β causes the projected sublattice $\mathbf{B}_{[n-\beta+1,n]}$ to be HKZ-reduced.

2.6 Lattice sieving

Lattice sieves provide a way to efficiently produce a list of many short lattice vectors [NV08, MV10, BDGL16]. Although there are many variations, e.g. [BLS16], one may think of a sieve as initially generating a list L of random linear combinations of some basis vectors defining a lattice $\Lambda \subset \mathbb{R}^n$, and then iteratively sieving, i.e., finding reductions that replace $\mathbf{v} \in L$ by a shorter $\mathbf{v} - \mathbf{w}$ for some $\mathbf{w} \in L$.

Throughout this thesis, we assume a lattice sieve will never return both $\mathbf{w}$ and $-\mathbf{w}$, since this optimisation is used in most implementations.

Definition 2.113 (lattice sieve)**.** A lattice sieve is a probabilistic algorithm that has the following input and output. Input consists of a basis $\mathbf{B} \in \mathbb{R}^{n\times k}$ of a random lattice Λ, a saturation radius $r_{\text{sat}} \in \mathbb{R}_{\geq 1}$ and saturation ratio $f_{\text{sat}} \in (0, 1]$. Output is a list of $N = \left\lceil \frac{1}{2} f_{\text{sat}} r_{\text{sat}}^n \right\rfloor$ lattice points, each of norm at most $r_d = r_{\text{sat}}\,\mathrm{GH}(k) \cdot \det(\Lambda)^{1/k}$. We use $\textsc{Sieve}(\Lambda, r_{\text{sat}}, f_{\text{sat}})$ to denote such a lattice sieve.

In order to analyse an algorithm that uses lattice sieving, we make a heuristic assumption regarding the distribution of the lattice vectors that a lattice sieve algorithm outputs.

Heuristic 2.114. *Let $\Lambda \subset \mathbb{R}^n$ be a full-rank unit-volume lattice and let $r_{\text{sat}} \in \mathbb{R}_{\geq 1}$, $f_{\text{sat}} \in (0, 1]$. The output* $\text{Sieve}(\Lambda, r_{\text{sat}}, f_{\text{sat}})$ *is a list of N i.i.d. samples from $\mathrm{U}(r_d \mathcal{B}^n)$ where $r_d = r_{\text{sat}}\,\mathrm{GH}(n)$.*

This heuristic makes three simplifying assumptions on the output of a lattice sieve.

(1) The output list is assumed to consist of N i.i.d. samples from the set of possible outputs, i.e., lattice vectors of norm at most r_d. Thus, the simplified list may have duplicates. Yet, there is a moderate difference with the actual output of a sieve, as sampling also produces many distinct values. Namely, using linearity of expectation one can show that sampling n times uniformly from $[n]$ yields in expectation $n(1 - 1/e) \approx 0.63n$ distinct values as $n \to \infty$.
(2) Moreover, the output samples are assumed to follow the uniform distribution on $\Lambda \cap (r_d \mathcal{B}^n)$.
(3) Last, we forget the lattice structure Λ in the output list. Heuristic 2.42 predicts that the norm distribution of lattice vectors is similar to that of points uniformly from a ball. When sieving a random lattice, this assumption seems fair, because we do not expect the lattice to be distorted in any particular direction.

We remark that this heuristic does not assume all vectors are of the same length r_d, cf. Guo–Johansson [GJ21]. Instead, the heuristic predicts there may be some shorter vectors, albeit with smaller probability. Still, most of the vectors are close to the boundary of the ball.

By default, a lattice sieve uses a saturation radius of $r_{\text{sat}} = \sqrt{4/3}$, because in that case enough vectors are initially generated to reduce vectors in the sieve in each step [NV08]. Although in theory, one sometimes assumes a lattice sieve finds *all* lattice vectors inside a ball of radius r_d, i.e., $f_{\text{sat}} = 1$, in practice a much lower saturation ratio is chosen for efficiency. For instance, the G6K software [ADH+19] uses a saturation ratio of 0.5 by default, and even lower saturation ratios of 0.375 were used in sieves to break certain SVP challenges with GPUs [DSvW21].

One can also sieve on a lattice that is not full-rank. Sieve algorithms require both time and memory exponential in the rank of Λ in general [BDGL16].

2.7 Signature scheme

A *signature scheme* is a triple of probabilistic polynomial-time algorithms $\Pi = (\textsc{KeyGen}, \textsc{Sign}, \textsc{Verify})$ such that Verify is deterministic.

- On input 1^n, KeyGen outputs a public and private key $(\mathsf{pk}, \mathsf{sk})$. We assume n can be determined from either key.
- On input sk and a message m from a message space that may depend on pk, Sign outputs a signature sig.
- On input pk, a message m and a signature sig, Verify outputs a bit $b \in \{0, 1\}$. We say sig is a valid signature on m if and only if $b = 1$.

In our practical cryptanalysis of Section 5.3 we discuss two types of forgery an adversary may produce, strong and weak.

- A *strong forgery* is a valid signature on a message for which an adversary does not know a signature. If an adversary cannot produce a strong forgery, we call the signature scheme *weakly unforgeable*, or EUF-CMA secure.
- A *weak forgery* is a valid signature on a message for which an adversary already knows a valid signature. Moreover, to be a weak forgery on a message m, this signature must be different from all the valid signatures on m known to the adversary. If an adversary can produce neither weak nor strong forgery, we call the signature scheme *strongly unforgeable*, or SUF-CMA secure.