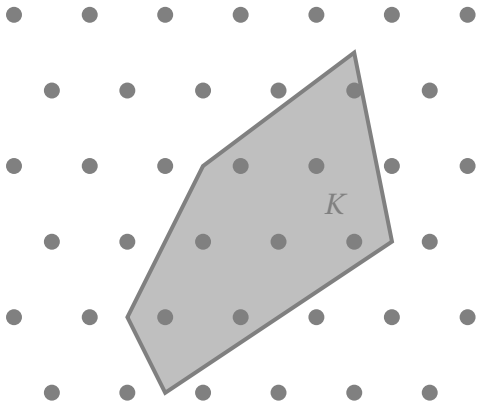
Lattices

CSE 599S — Winter 2023

Thomas Rothvoss



UNIVERSITY of WASHINGTON

Last changes: July 8, 2025

Contents

Chapter 1

Introduction to Lattices

In this chapter, we introduce the concept of *lattices*. Lattices are fundamentally important in discrete geometry, cryptography, discrete optimization and computer science as a whole. For this introductory chapter, we follow to some extend the short, but very readable material in Chapter 2 of the text book "*Lectures on Discrete Geometry*" by Jiri Matousek [Mat02] and to some extend the excellent lecture notes by Oded Regev[1] as well as the ones by Chris Peikert[2]. The author is grateful to Victor Reis for carefully checking the manuscript and providing useful feedback.

1.1 Basics of Lattices

Lattices are integral combinations of linearly independent vectors. Formally, a lattice is a set

$$\left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i \mid \lambda_1, \ldots, \lambda_k \in \mathbb{Z} \right\}$$

where $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k \in \mathbb{R}^n$ are linearly independent vectors. An alternative definition is to say that a lattice is a *discrete subgroup* of $\mathbb{R}^n$, where "discrete" means that there is an $\varepsilon > 0$ so that all points in the lattice have distance at least ε from each other. We denote the number $\operatorname{rank}(\Lambda) := k$ as the *rank* of the lattice. In other words, $\operatorname{rank}(\Lambda) = \dim(\operatorname{span}(\Lambda))$.

If $k = n$, then the lattice has *full rank*. As every lattice is just a full rank lattice when restricted to the subspace $\operatorname{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k\}$, most of the time we will consider full-rank lattices. In fact, we will drop the term "full-rank" and assume it implicitly from now on if not announced otherwise.

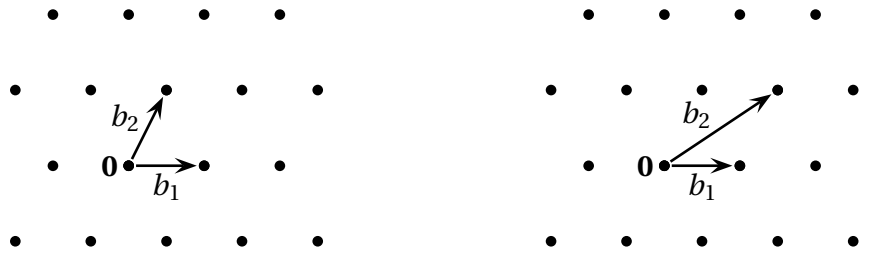
[1] See http://www.cims.nyu.edu/~regev/teaching/lattices_fall_2009

[2] see http://www.cc.gatech.edu/~cpeikert/lic13

For the sake of brevity, let $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ be the matrix that has the basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ as columns. Then we abbreviate the lattice as

$$\Lambda(\boldsymbol{B}) = \Big\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{b}_i \mid \lambda_1, \ldots, \lambda_n \in \mathbb{Z} \Big\}$$

The matrix $\boldsymbol{B}$ itself is also called a *basis* of the lattice $\Lambda(\boldsymbol{B})$. Note that a lattice has more than one basis. For example adding any integral multiple of $\boldsymbol{b}_i$ to $\boldsymbol{b}_j$ for $j \neq i$ will preserve the set of points that can be generated. Below we see an example of two different bases for the same underlying lattice:



General notation and definitions. Let us fix some notation for later: We denote $B_2^n := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 \leq 1\}$ as the *Euclidean ball* and more generally $B_p^n := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_p \leq 1\}$ is the ball for the norm $\|\boldsymbol{x}\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $1 \leq p < \infty$ with $B_\infty^n := [-1,1]^n$ is the *cube*. For a measurable set $A \subseteq \mathbb{R}^n$ we write $\mathrm{Vol}_n(A) := \int_A 1 d\boldsymbol{x}$ as the *n-dimensional volume*. Frequently we will need an estimate on the volume of B_2^n. A loose but convenient estimate is as follows:

Lemma 1.1. *For all $n \geq 1$ one has $2^n \leq Vol_n(\sqrt{n}B_2^n) \leq (2e)^n$.*

Proof. On the one hand $\sqrt{n}B_2^n$ contains the cube B_∞^n and so $\mathrm{Vol}_n(\sqrt{n}B_2^n) \geq \mathrm{Vol}_n(B_\infty^n) = 2^n$. On the other hand $\sqrt{n}B_2^n$ is contained in B_1^n and so $\mathrm{Vol}_n(\sqrt{n}B_2^n) \leq \mathrm{Vol}_n(nB_1^n) = \frac{(2n)^n}{n!} \leq (2e)^n$ as $n! \geq (\frac{n}{e})^n$ for all $n \geq 2$ (and the claim is true anyway for $n = 1$). □

For a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ we write $\det_n(\boldsymbol{A})$ as the *determinant*. We use the index n to indicate the format of the matrix.

Recall that a set $K \subseteq \mathbb{R}^n$ is called *convex* if for all $\boldsymbol{x}, \boldsymbol{y} \in K$ and $0 \leq \lambda \leq 1$ one also has $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in K$. A set K is *(centrally) symmetric* if $\boldsymbol{x} \in K$ if and only if $-\boldsymbol{x} \in K$. In particular for a convex symmetric set K, we can define a norm that is called the *Minkowski norm* $\|\boldsymbol{x}\|_K := \min\{\lambda \geq 0 : \boldsymbol{x} \in \lambda K\}$. In other words, $\|\boldsymbol{x}\|_K$ gives the scaling factor that one needs until the scaled copy of K includes $\boldsymbol{x}$. For example the Euclidean norm $\|\cdot\|_2$ is the norm $\|\cdot\|_K$ for $K := B_2^n$. For a vector $\boldsymbol{y} \in \mathbb{R}^n$, we define $\boldsymbol{y} + K := \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in K\}$ as the *translate* of K by $\boldsymbol{y}$. For sets $A, B \subseteq \mathbb{R}^n$ we write $A + B = \{\boldsymbol{a} + \boldsymbol{b} \mid \boldsymbol{a} \in A,\ \boldsymbol{b} \in B\}$ as the *Minkowski sum*. Throughout this manuscript

we will denote vectors and matrices in bold. For example the matrix $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ has columns $\boldsymbol{B}^1,\ldots,\boldsymbol{B}^n$ and rows $\boldsymbol{B}_1,\ldots,\boldsymbol{B}_m$ but entries $B_{ij} \in \mathbb{R}$.

1.1.1 Unimodular matrices

We want to spend a bit of time characterizing the different bases for a lattice.

Definition 1.2. An $n\times n$ matrix $\boldsymbol{U}$ is called *unimodular*, if $\boldsymbol{U} \in \mathbb{Z}^{n\times n}$ and $\det_n(\boldsymbol{U}) \in \{\pm 1\}$.

Lemma 1.3. *If $\boldsymbol{U}$ is unimodular, then $\boldsymbol{U}^{-1}$ is unimodular.*

Proof. We have $\det_n(\boldsymbol{U}^{-1}) = \frac{1}{\det_n(\boldsymbol{U})} \in \{-1,1\}$. So, it remains to argue that $\boldsymbol{U}^{-1}$ has only integral entries. Set $\boldsymbol{U}^{ij} \in \mathbb{Z}^{n\times n}$ as the matrix where the ith column is replaced by the jth unit vector $\boldsymbol{e}_j$[3]. Then $\det_n(\boldsymbol{U}^{ij}) \in \mathbb{Z}$ and by *Cramer's rule*

$$U^{-1}_{ij} = \frac{\det_n(\boldsymbol{U}^{ij})}{\det_n(\boldsymbol{U})} \in \mathbb{Z}.$$

□

We will now see that two matrices span the same lattice if and only if they differ by a unimodular matrix:

Lemma 1.4. *Let $\boldsymbol{B}_1,\boldsymbol{B}_2 \in \mathbb{R}^{n\times n}$ non-singular. Then $\Lambda(\boldsymbol{B}_1) = \Lambda(\boldsymbol{B}_2)$ if and only if there is a unimodular matrix $\boldsymbol{U}$ with $\boldsymbol{B}_2 = \boldsymbol{B}_1\boldsymbol{U}$.*

Proof. First, suppose that $\boldsymbol{B}_2 = \boldsymbol{B}_1\boldsymbol{U}$ with $\boldsymbol{U}$ being unimodular. The important observation is that the map $f : \mathbb{Z}^n \to \mathbb{Z}^n$ with $f(\boldsymbol{x}) := \boldsymbol{U}\boldsymbol{x}$ is a *bijection* on the integer lattice as $\boldsymbol{U}\boldsymbol{x} \in \mathbb{Z}^n$ for $\boldsymbol{x} \in \mathbb{Z}^n$ and any vector $\boldsymbol{y} \in \mathbb{Z}^n$ is hit by $\boldsymbol{U}(\boldsymbol{U}^{-1}\boldsymbol{y}) = \boldsymbol{y}$. Then

$$\Lambda(\boldsymbol{B}_2) = \{\boldsymbol{B}_2\boldsymbol{\lambda} \mid \boldsymbol{\lambda} \in \mathbb{Z}^n\} = \{\boldsymbol{B}_1\boldsymbol{U}\boldsymbol{\lambda} \mid \boldsymbol{\lambda} \in \mathbb{Z}^n\} = \Lambda(\boldsymbol{B}_1).$$

Now, let us go the other way around and assume that $\Lambda(\boldsymbol{B}_1) = \Lambda(\boldsymbol{B}_2)$. Then any column of $\boldsymbol{B}_1$ is an integral combination of columns in $\boldsymbol{B}_2$ and vice versa. We can use those integral coefficients to fill matrices $\boldsymbol{U},\boldsymbol{V} \in \mathbb{Z}^{n\times n}$ so that $\boldsymbol{B}_2 = \boldsymbol{B}_1\boldsymbol{U}$ and $\boldsymbol{B}_1 = \boldsymbol{B}_2\boldsymbol{V}$. Then

$$\det_n(\boldsymbol{B}_1) = \det_n(\boldsymbol{B}_2\boldsymbol{V}) = \det_n(\boldsymbol{B}_1\boldsymbol{U}\boldsymbol{V}) = \det_n(\boldsymbol{B}_1)\cdot\det_n(\boldsymbol{U})\cdot\det_n(\boldsymbol{V})$$

As $\det_n(\boldsymbol{U}),\det_n(\boldsymbol{V}) \in \mathbb{Z}$, we must have $\det_n(\boldsymbol{U}) \in \{-1,1\}$ (and in fact it is not hard to argue that $\boldsymbol{V} = \boldsymbol{U}^{-1}$). □

[3] Sounds like we have accidentally switched row and column indices — but it was on purpose.

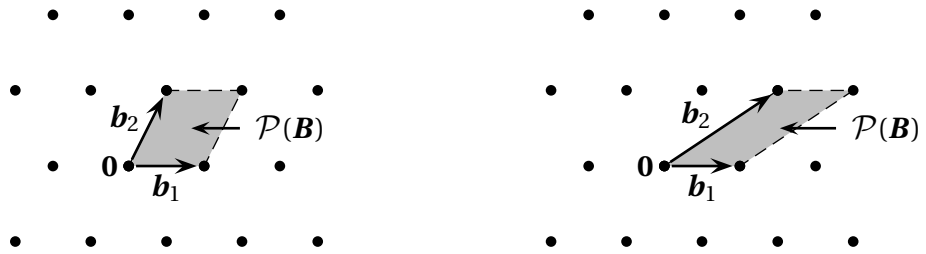
Note that the unimodular matrix $\boldsymbol{U}$ can be found in polynomial time using row reduction / Gauss elimination. Hence, for given matrices $\boldsymbol{B}_1, \boldsymbol{B}_2$ one can test in polynomial time whether they generate the same lattice.

1.1.2 The fundamental parallelepiped

The *fundamental parallelepiped* of the lattice $\Lambda(\boldsymbol{B})$ is the polytope

$$\mathcal{P}(\boldsymbol{B}) := \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{b}_i \mid 0 \le \lambda_i < 1 \ \forall i \in [n] \right\}$$

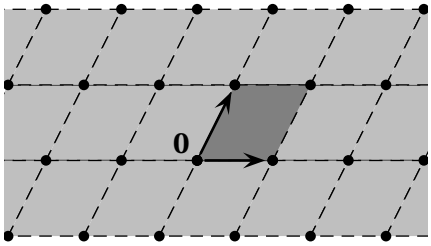
We see that this definition actually depends on the basis:



Let us make some observation: Since $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ is a basis of $\mathbb{R}^n$, we know that for every $\boldsymbol{x} \in \mathbb{R}^n$ there is a unique coefficient vector $\boldsymbol{\lambda} \in \mathbb{R}^n$ so that $\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i \boldsymbol{b}_i$. That means $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = \underbrace{\sum_{i=1}^{n} \lfloor \lambda_i \rfloor \boldsymbol{b}_i}_{\in \Lambda(\boldsymbol{B})} + \underbrace{\sum_{i=1}^{n} (\lambda_i - \lfloor \lambda_i \rfloor) \boldsymbol{b}_i}_{\in \mathcal{P}(\boldsymbol{B})}.$$

In other words, the *translates* of the parallelepiped placed at lattices points exactly partition the $\mathbb{R}^n$. We call this a *tiling* of $\mathbb{R}^n$. The tiling of the space with parallelepipeds can be visualized as follows:



Note that actually we can rewrite the fundamental parallelepiped as

$$\mathcal{P}(\boldsymbol{B}) = \{\boldsymbol{B}\boldsymbol{x} : \boldsymbol{x} \in [0,1)^n\},$$

that means it is the image of the hypercube $[0,1)^n$ under the linear map given by the matrix $\boldsymbol{B}$. Then by the *transformation formula*

$$\mathrm{Vol}_n(\mathcal{P}(\boldsymbol{B})) = \underbrace{\mathrm{Vol}_n([0,1)^n)}_{=1} \cdot |\det_n(\boldsymbol{B})|$$

While the fundamental parallelepiped itself does depend on the choice of the basis — its volume does not!

Lemma 1.5. *Let $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ and let $\Lambda := \Lambda(\boldsymbol{B})$ be the generated lattice. Then the* determinant of the lattice $\det(\Lambda) := |\det_n(\boldsymbol{B})|$ *is independent of the chosen basis. Moreover,* $\det(\Lambda) = \mathit{Vol}_n(\mathcal{P}(\boldsymbol{B}))$.

Proof. Clear, because for different basis $\boldsymbol{B}, \boldsymbol{B}' \in \mathbb{R}^n$ of the same lattice, there is a unimodular transformation $\boldsymbol{U} \in \mathbb{Z}^{n\times n}$ with $\boldsymbol{B}' = \boldsymbol{B}\boldsymbol{U}$. Then $|\det_n(\boldsymbol{B}')| = |\det_n(\boldsymbol{B})| \cdot |\det_n(\boldsymbol{U})| = |\det_n(\boldsymbol{B})|$. □

Note that the quantity $\frac{1}{\mathrm{Vol}_n(\mathcal{P}(\boldsymbol{B}))}$ gives the *density* of the lattice. For example the number of lattice points in a ball of large radius $R > 0$ is

$$|\Lambda(\boldsymbol{B}) \cap R \cdot B_2^n| \approx \frac{\mathrm{Vol}_n(R \cdot B_2^n)}{\mathrm{Vol}_n(\mathcal{P}(\boldsymbol{B}))}$$

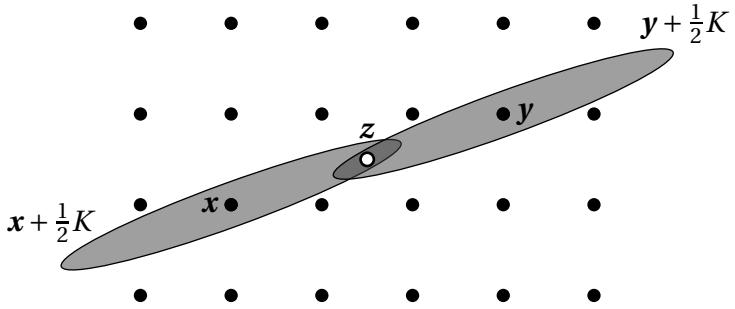
The "≈" should be understood that the ratio of both sides goes to 1 as $R \to \infty$. This is another more geometric argument why the volume of the fundamental parallelepiped cannot depend on the basis.

1.2 Minkowski's Theorem

We now come to Minkowski's Theorem which says that every large enough symmetric convex set must contain a non-zero lattice point.

Theorem 1.6 (Minkowski's First Theorem (1889))**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $K \subseteq \mathbb{R}^n$ be a bounded symmetric convex set with $\mathit{Vol}_n(K) > 2^n \det(\Lambda)$. Then $K \cap (\Lambda \setminus \{\mathbf{0}\}) \neq \emptyset$.*

Proof. First, by assumption we have $\mathrm{Vol}_n(\frac{1}{2}K) > \det(\Lambda)$. Next, place copies of $\frac{1}{2}K$ at every lattice point in Λ.



Then on average, points in $\mathbb{R}^n$ are covered more than once. So there must be two different lattice points $\boldsymbol{x}, \boldsymbol{y} \in \Lambda$ so that the translates $\boldsymbol{x}+\frac{1}{2}K$ and $\boldsymbol{y}+\frac{1}{2}K$ overlap. Let $\boldsymbol{z} \in (\boldsymbol{x}+\frac{1}{2}K) \cap (\boldsymbol{y}+\frac{1}{2}K)$ be a point in that intersection. Then

$$\|\boldsymbol{x}-\boldsymbol{y}\|_K \leq \underbrace{\|\boldsymbol{x}-\boldsymbol{z}\|_K}_{\leq 1/2} + \underbrace{\|\boldsymbol{y}-\boldsymbol{z}\|_K}_{\leq 1/2} \leq 1$$

Hence $(\boldsymbol{x}-\boldsymbol{y}) \in \Lambda/\{\boldsymbol{0}\}$ is the lattice point that we are looking for. □

Admittedly, this argument was a bit informal as we talked about the average density of an infinite covering. But one can make the argument nicely finite. Let $D>0$ be so that $K \subseteq D \cdot \boldsymbol{B}[-1,1]^n = D \cdot (\mathcal{P}(\boldsymbol{B}) - \mathcal{P}(\boldsymbol{B}))$. For $R \in \mathbb{N}$, $R \cdot \mathcal{P}(\boldsymbol{B}) + K$ fully contains at least R^n translates of K placed at lattice points. Hence

$$\begin{aligned} \frac{\mathrm{Vol}_n(\text{translates in } R \cdot \mathcal{P}(\boldsymbol{B}) + K)}{\mathrm{Vol}_n(R \cdot \mathcal{P}(\boldsymbol{B}) + K)} &\geq \frac{R^n \cdot \mathrm{Vol}_n(\frac{1}{2}K)}{\mathrm{Vol}_n(R \cdot \mathcal{P}(\boldsymbol{B}) + D \cdot (\mathcal{P}(\boldsymbol{B}) - \mathcal{P}(\boldsymbol{B})))} \\ &= \Big(\frac{R}{R+2D}\Big)^n \underbrace{\frac{\mathrm{Vol}_n(\frac{1}{2}K)}{\det(\Lambda)}}_{>1} > 1 \end{aligned}$$

if R is large enough and some point in $R \cdot \mathcal{P}(\boldsymbol{B}) + K$ must be covered twice.

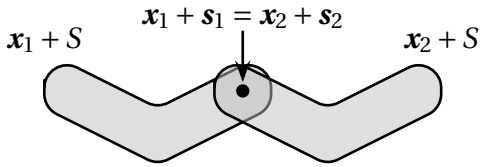
It is not difficult to give a quantitative version of Minkowski's Theorem; we will discuss the proof in the Exercises:

Theorem 1.7. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $K \subseteq \mathbb{R}^n$ be a bounded symmetric convex set. Then $|K \cap \Lambda| \geq \frac{Vol_n(K)}{2^n \det(\Lambda)}$.*

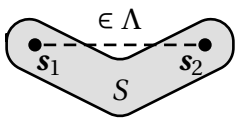
There is another theorem that is closely related to the one of Minkowski. Suppose that S is an arbitrary set; then S could be large without containing a lattice point. But it still has to contain *differences* that are lattice points.

Theorem 1.8 (Blichfeldt)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $S \subseteq \mathbb{R}^n$ be a measurable set with $Vol_n(S) > \det(\Lambda)$. Then there are $\boldsymbol{s}_1, \boldsymbol{s}_2 \in S$ with $\boldsymbol{s}_1 - \boldsymbol{s}_2 \in \Lambda$.*

Proof. Place copies of $\boldsymbol{x}+S=\{\boldsymbol{x}+\boldsymbol{s} \mid \boldsymbol{s}\in S\}$ for all $\boldsymbol{x}\in\Lambda$. Again one can argue that there will be different points $\boldsymbol{x}_1,\boldsymbol{x}_2\in\Lambda$ with $(\boldsymbol{x}_1+S)\cap(\boldsymbol{x}_2+S)\neq\emptyset$ (similar to above; we are skipping the details).



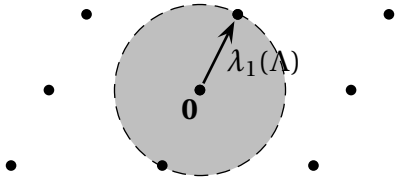
Let $\boldsymbol{s}_1,\boldsymbol{s}_2\in S$ be the points with $\boldsymbol{x}_1+\boldsymbol{s}_1=\boldsymbol{x}_2+\boldsymbol{s}_2$. Rearranging gives $\boldsymbol{s}_1-\boldsymbol{s}_2=\boldsymbol{x}_2-\boldsymbol{x}_1\in\Lambda$. This gives the claim.



□

1.2.1 Minkowski's Theorem and the Shortest Vector

A particularly interesting vector in a lattice is the *Shortest Vector* with respect to the $\|\cdot\|_2$-norm. Let us abbreviate $\lambda_1(\Lambda):=\min\{\|\boldsymbol{x}\|_2 \mid \boldsymbol{x}\in\Lambda/\{\boldsymbol{0}\}\}$ as its length.



In fact, finding the shortest vector (or its length) is an **NP**-hard problem[4]. However, one can get some estimates on it:

Theorem 1.9. *Any lattice $\Lambda\subseteq\mathbb{R}^n$ one has $\lambda_1(\Lambda)\leq\sqrt{n}\cdot\det(\Lambda)^{1/n}$.*

Proof. First, for $r:=\det(\Lambda)^{1/n}$, the hypercube $[-r,r]^n$ has a volume of

$$\mathrm{Vol}_n([-r,r]^n)=(2r)^n\geq 2^n\det(\Lambda).$$

Hence by Theorem 1.6 there is a point $\boldsymbol{x}\in\Lambda\setminus\{\boldsymbol{0}\}$ with $\|\boldsymbol{x}\|_\infty\leq\det(\Lambda)^{1/n}$. Of course $\|\boldsymbol{x}\|_2\leq\sqrt{n}\cdot\|\boldsymbol{x}\|_\infty$, which implies the claim. □

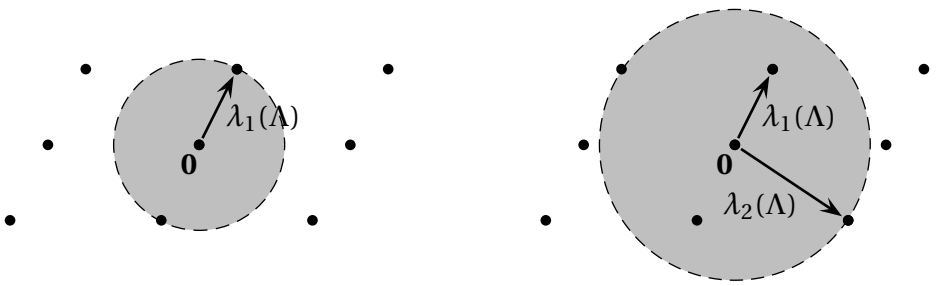
[4] To be precise, finding the shortest vector in the $\|\cdot\|_\infty$ is **NP**-hard [vEB81]. For the Euclidean norm this is only known under randomized reductions [Ajt98].

Note that the scaling in the claim actually makes sense: if we scale the lattice Λ by a factor $s > 0$, then the length of the shortest vector also scales with s, while $\det(\Lambda)$ changes by a factor of s^n. The analysis can be improved to $c\sqrt{n}\det(\Lambda)^{1/n}$ for a constant $c < 1$ — however, this is best possible. There are lattices for every n with determinant 1 and shortest vector of length $\Omega(\sqrt{n})$.

Rather than only studying $\lambda_1(\Lambda)$ we are also interested in the "ith shortest vector" in a lattice. More precisely we call

$$\lambda_i(\Lambda) := \min\{r \geq 0 \mid \dim(\mathrm{span}(rB_2^n \cap \Lambda)) \geq i\}$$

the *ith successive minimum*. That means one has i many linearly independent vectors of length at most $\lambda_i(\Lambda)$ and $0 < \lambda_1(\Lambda) \leq \lambda_2(\Lambda) \leq \ldots \leq \lambda_n(\Lambda)$.



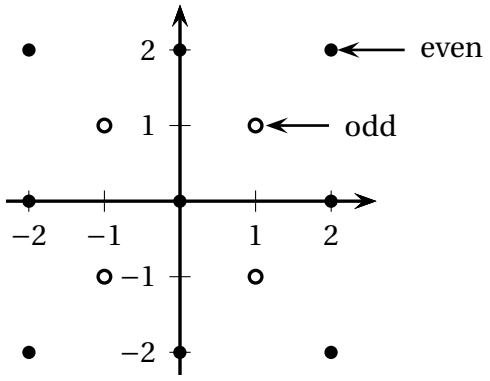
1.2.2 More on Successive Minima

Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \Lambda$ be the linearly independent vectors attaining the successive minima, that means the vectors are linearly independent and $\|\boldsymbol{v}_i\|_2 = \lambda_i(\Lambda)$. Clearly, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are a basis of $\mathbb{R}^n$, so naturally one might ask whether they are also a basis of the lattice Λ? We want to discuss an illustrative example showing that this is not the case.

Let us call an integer vector $\boldsymbol{x} \in \mathbb{Z}^n$ *even* if all its coordinates $x_1, \ldots, x_n$ are even. Similarly, let us call the vector *odd* when all coordinates are odd. Consider the set

$$\Lambda := \{\boldsymbol{x} \in \mathbb{Z}^n \mid (\boldsymbol{x} \text{ is even}) \text{ or } (\boldsymbol{x} \text{ is odd})\}.$$

Note that this is indeed a lattice because the sum of two odd vectors is even; the sum of an odd and an even vector is odd; and so on. For $n = 2$ the lattice looks as follows:



Now consider the higher-dimensional case with $n \geq 5$. Then every odd vector $\boldsymbol{x} \in \Lambda$ has length $\|\boldsymbol{x}\|_2 \geq \sqrt{n}$. We have $\lambda_i(\Lambda) = 2$ for all $i \in \{1,\ldots,n\}$ since the vectors $2\boldsymbol{e}_1,\ldots,2\boldsymbol{e}_n$ are linearly independent even integer vectors and the lattice does not contain any shorter vector (the only options would be $(1,0,\ldots,0)$ and $(1,1,0,\ldots,0)$ but these are not in the lattice). On the other hand, every odd vector has length at least $\sqrt{n}$. The lattice does include odd vectors like $\mathbf{1} := (1,\ldots,1)$ and these vectors are not integer combinations of even vectors. Hence any basis for Λ must include an odd vector of length at least $\sqrt{n}$. In particular, the vectors attaining the successive minima do not form a basis of the lattice.

1.2.3 Dirichlet's Theorem

We will now see another elegant application of Minkowski's First Theorem. Suppose we have a vector $\boldsymbol{\alpha} \in [0,1]^n$ of *real* numbers and we want to approximate the vector as well as possible with a vector of *rational* numbers so that the common denominator is at most a parameter Q. One can find some obvious applications in computer science, where one simply cannot work with real numbers but has to use rational approximations all the time. Then the most obvious choice would be

$$\left(\frac{\lceil \alpha_1 Q \rfloor}{Q},\ldots,\frac{\lceil \alpha_n Q \rfloor}{Q}\right)$$

where $\lceil\cdot\rfloor$ rounds up or down to the nearest integer. One can easily see that the rounding error in every component is upper bounded by $\frac{1}{2Q}$. Is this best possible? Well, the task was to have a common denominator that is *at most* Q. So, we are allowed to pick any denominator in $\{1,\ldots,Q\}$, but we haven't made use of that freedom.

Theorem 1.10 (Dirichlet)**.** *For any $\boldsymbol{\alpha} \in]0,1]^n$ and $Q \in \mathbb{N}$, there are numbers $p_1,\ldots,p_n \in$*

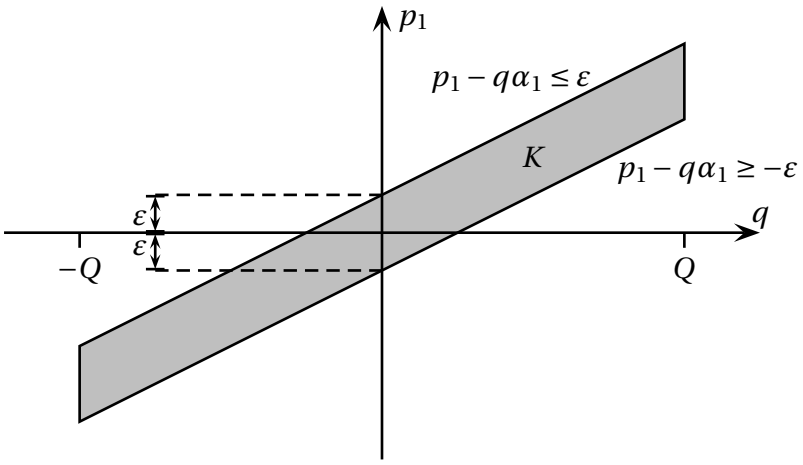
$\mathbb{Z}_{\geq 0}$ *and* $q \in \{1,\dots,Q\}$ *so that*

$$\max_{i=1,\dots,n} \left| \frac{p_i}{q} - \alpha_i \right| \leq \frac{1}{Q^{1/n} q}$$

Proof. If we abbreviate $\varepsilon := \frac{1}{Q^{1/n}}$ and multiply the inequality in the claim by q, then we get the constraint $|p_i - q \cdot \alpha_i| \leq \varepsilon$ for all $i = 1,\dots,n$. Note that this is a linear constraint, hence we can reduce our problem to finding an integer point in the polytope

$$K := \left\{ (p_1,\dots,p_n,q) \in \mathbb{R}^{n+1} \mid |p_i - q \cdot \alpha_i| \leq \varepsilon \; \forall i = 1,\dots,n; \; |q| \leq Q \right\}$$

Note that on purpose, we admitted negative numbers to make the set K symmetric. For example for $n = 1$, one obtains the following picture:



One should think about K as a thin, but long "slab" along the line defined by $\boldsymbol{p} - q \cdot \boldsymbol{\alpha} = \mathbf{0}$. Geometrically speaking, the set K is a *parallelepiped* and his volume is equal to the volume of the box with length $2Q$ in one direction and 2ε in n directions. Hence

$$\mathrm{Vol}_n(K) = 2Q \cdot (2Q^{-1/n})^n = 2^{n+1}.$$

Now we can apply Minkowski's theorem and we obtain $(p_1,\dots,p_n,q) \in (K \cap \mathbb{Z}^{n+1}) \setminus \{\mathbf{0}\}$. For symmetry reasons, we can assume that $q \geq 0$. Note that it is impossible that $q = 0$, because otherwise $|p_i| \leq Q^{-1/n} < 1$ which implies that $p_1 = \ldots = p_n = 0$ and we would get a contradiction. Hence $q \in \{1,\dots,Q\}$ and we have the desired approximation. □

1.3 The Gram Schmidt orthogonalization

We have already mentioned that it is **NP**-hard to find the shortest vector in a lattice. Our goal is to be at least able to find an *approximate* shortest vector. That means for a lattice $\Lambda(\boldsymbol{B})$ we want to find a vector $\boldsymbol{x} \in \Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}$ in polynomial time that has length $\|\boldsymbol{x}\|_2 \leq \alpha \cdot \lambda_1(\Lambda(\boldsymbol{B}))$. Here, $\alpha := \alpha(n) \geq 1$ is the so-called *approximation factor* that we would like to be as small as possible. Before we come to that algorithm, we need a useful procedure.

The *Gram-Schmidt orthogonalisation* takes linearly independent vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n \in \mathbb{R}^n$ as input and it computes an orthogonal basis $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$ so that $\text{span}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k) = \text{span}(\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_k^*)$ for all $k = 1, \ldots, n$. The idea is that we go through the vectors $\boldsymbol{b}_1, \ldots \boldsymbol{b}_n$ in that order and for each i we subtract all components of $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}$ from $\boldsymbol{b}_i$ and call the remainder $\boldsymbol{b}_i^*$. Formally the method is as follows:

Gram-Schmidt orthogonalisation

Input: Vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n \in \mathbb{R}^n$

Output: Orthogonal basis $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$

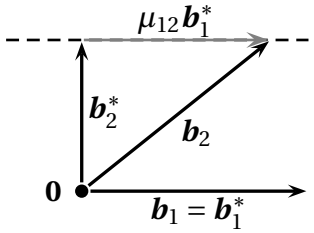
(1) $\boldsymbol{b}_1^* := \boldsymbol{b}_1$

(2) $\boldsymbol{b}_2^* := \boldsymbol{b}_2 - \mu_{1,2}\boldsymbol{b}_1^*$ with $\mu_{1,2} := \frac{\langle \boldsymbol{b}_2, \boldsymbol{b}_1^* \rangle}{\|\boldsymbol{b}_1^*\|_2^2}$

(3) ...

(j) $\boldsymbol{b}_j^* := \boldsymbol{b}_j - \sum_{i<j} \mu_{i,j}\boldsymbol{b}_i^*$ with $\mu_{ij} := \frac{\langle \boldsymbol{b}_j, \boldsymbol{b}_i^* \rangle}{\|\boldsymbol{b}_i^*\|_2^2} \quad \forall\, j = 1, \ldots, n$

Note that $\boldsymbol{b}_i^*$ is the *projection* of $\boldsymbol{b}_i$ on $\text{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}\}^\perp$. For example for $n = 2$ the method can be visualized as follows:



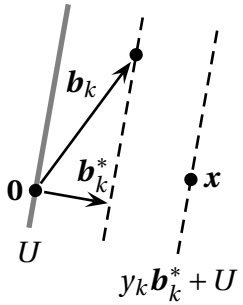
Note that the outcome of the Gram-Schmidt orthogonalization crucially depends on the *order* of the vectors. The next observation is that by *Cavalieri's principle*, the "shifting" does not change the volume of the fundamental parallelepiped. Hence

$$\det(\Lambda(\boldsymbol{B})) = \text{Vol}_n(\mathcal{P}(\boldsymbol{B})) = \prod_{i=1}^{n} \|\boldsymbol{b}_i^*\|_2.$$

The Gram-Schmidt orthogonalization gives us a nice lower bound on the length of a shortest vector.

Theorem 1.11. *Let $\boldsymbol{B}$ be a basis and $\boldsymbol{B}^* = (\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*)$ be its Gram-Schmidt orthogonalization. Then $\lambda_1(\Lambda(\boldsymbol{B})) \geq \min_{i=1,\ldots,n} \|\boldsymbol{b}_i^*\|_2$.*

Proof. Let $\boldsymbol{x} \in \Lambda(\boldsymbol{B})$ be any lattice vector and let $\boldsymbol{x} = \sum_{i=1}^n y_i \boldsymbol{b}_i$ with $y_i \in \mathbb{Z}$ be the linear combination that generates it. Let k be the largest index with $\lambda_k \neq 0$. Define the subspace $U := \operatorname{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k-1}\} = \operatorname{span}\{\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_{k-1}^*\}$.



Then $\boldsymbol{x}$ lies on a translate of that subspace which is $y_k\boldsymbol{b}_k^* + U$. Hence $\|\boldsymbol{x}\|_2 \geq d(\boldsymbol{x}, U) = |y_k| \cdot \|\boldsymbol{b}_k^*\|_2 \geq \|\boldsymbol{b}_k^*\|_2$ where $d(\boldsymbol{x}, U)$ tells the distance of $\boldsymbol{x}$ to U. □

In particular $\boldsymbol{b}_1^* = \boldsymbol{b}_1$ is always lattice vector — $\boldsymbol{b}_2^*, \ldots, \boldsymbol{b}_n^*$ generally not. One idea to find a short lattice vector would be to find a basis (and an ordering on the vectors!) so that $\|\boldsymbol{b}_1^*\|_2 \leq \rho \cdot \|\boldsymbol{b}_i^*\|_2$ for all i and some ρ. Then by the previous lemma $\lambda_1(\Lambda(\boldsymbol{B})) \geq \min_{i=1,\ldots,n} \|\boldsymbol{b}_i^*\|_2 \geq \frac{1}{\rho}\|\boldsymbol{b}_1\|_2$, hence $\boldsymbol{b}_1$ would be a ρ-approximation to the shortest vector. Later, this will be the goal of the LLL-algorithm.

1.4 Minkowski's 2nd Theorem

The Gram-Schmidt orthogonalization will also be helpful in deriving Minkowski's Second Theorem that gives us some control over the successive minima $\lambda_i(\Lambda)$:

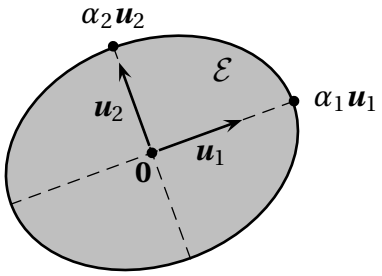
Theorem 1.12 (Minkowski's Second Theorem)**.** *For any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ one has*

$$\Big(\prod_{i=1}^n \lambda_i(\Lambda)\Big)^{1/n} \leq \sqrt{n} \cdot \det(\Lambda)^{1/n}.$$

While Minkowski's First Theorem only tells us that $\lambda_1(\Lambda)$ is at most $\sqrt{n} \cdot \det(\Lambda)^{1/n}$, the Second Theorem gives the stronger statement that even the *geometric average* of $\lambda_1(\Lambda), \ldots, \Lambda_n(\Lambda)$ is bounded by that same quantity. Take any *orthonormal basis* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in \mathbb{R}^n$ and any positive coefficients $\alpha_1, \ldots, \alpha_n > 0$. Then

$$\mathcal{E} = \Big\{\boldsymbol{x} \in \mathbb{R}^n \mid \sum_{i=1}^n \frac{1}{\alpha_i^2} \cdot \langle \boldsymbol{x}, \boldsymbol{u}_i \rangle^2 \leq 1\Big\}$$

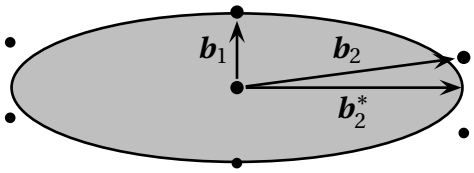
is an *ellipsoid*. Ellipsoids are convex symmetric bodies and their volume is $\mathrm{Vol}_n(\mathcal{E}) = \mathrm{Vol}_n(B_2^n) \cdot \prod_{i=1}^n \alpha_i$.



Now we come to the proof where we follow the exposition of Regev [Reg09a].

Proof. We abbreviate $\lambda_i := \lambda_i(\Lambda)$. Let $\boldsymbol{b}_1, \dots, \boldsymbol{b}_n \in \Lambda \setminus \{\boldsymbol{0}\}$ be the vectors that attain the successive minima, that means $\lambda_i = \|\boldsymbol{b}_i\|_2$ with $\lambda_1 \le \dots \le \lambda_n$. Let $\boldsymbol{b}_1^*, \dots, \boldsymbol{b}_n^*$ be the Gram Schmidt orthogonalization (in that order). We consider the ellipsoid

$$\mathcal{E} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid \sum_{i=1}^n \left(\frac{\langle \boldsymbol{x}, \boldsymbol{b}_i^* \rangle}{\|\boldsymbol{b}_i^*\|_2 \cdot \lambda_i} \right)^2 \le 1 \right\}$$



Let $\mathrm{int}(\mathcal{E}) = \{\boldsymbol{x} \mid \dots < 1\}$ be the interior of that ellipsoid. We claim that $\mathrm{int}(\mathcal{E}) \cap \Lambda = \{\boldsymbol{0}\}$. Take any lattice vector $\boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}$. Let k be maximal so that $\lambda_k \le \|\boldsymbol{x}\|_2$ (i.e. $\|\boldsymbol{x}\|_2 < \lambda_{k+1}$ or $k = n$). Note that this means that $\boldsymbol{x} \in \mathrm{span}(\boldsymbol{b}_1, \dots, \boldsymbol{b}_k) = \mathrm{span}(\boldsymbol{b}_1^*, \dots, \boldsymbol{b}_k^*)$ since otherwise, we could have chosen $\boldsymbol{x}$ instead of $\boldsymbol{b}_{k+1}$ and the value of λ_{k+1} would have been shorter than it is. Now we can bound

$$\sum_{i=1}^n \left(\frac{\langle \boldsymbol{x}, \boldsymbol{b}_i^* \rangle}{\|\boldsymbol{b}_i\|_2 \cdot \lambda_i} \right)^2 \overset{\substack{\langle \boldsymbol{x}, \boldsymbol{b}_i^* \rangle = 0 \ \forall i > k \\ \lambda_1 \le \dots \le \lambda_k}}{\ge} \frac{1}{\lambda_k^2} \underbrace{\sum_{i=1}^k \langle \boldsymbol{x}, \frac{\boldsymbol{b}_i^*}{\|\boldsymbol{b}_i^*\|_2} \rangle^2}_{= \|\boldsymbol{x}\|_2^2} = \frac{\|\boldsymbol{x}\|_2^2}{\lambda_k^2} \ge 1$$

This implies that indeed $\boldsymbol{x} \notin \mathrm{int}(\mathcal{E})$. Since $\mathcal{E}$ does not have a lattice point in its interior, *Minkowski's First Theorem* gives an upper bound on its volume:

$$2^n \cdot \det(\Lambda) \ge \mathrm{Vol}_n(\mathcal{E}) = \mathrm{Vol}_n(B_2^n) \cdot \prod_{i=1}^n \lambda_i \overset{\text{Lem 1.1}}{\ge} \left(\frac{2}{\sqrt{n}} \right)^n \cdot \prod_{i=1}^n \lambda_i$$

Rearranging then gives the claim. ∎

1.5 The LLL-algorithm

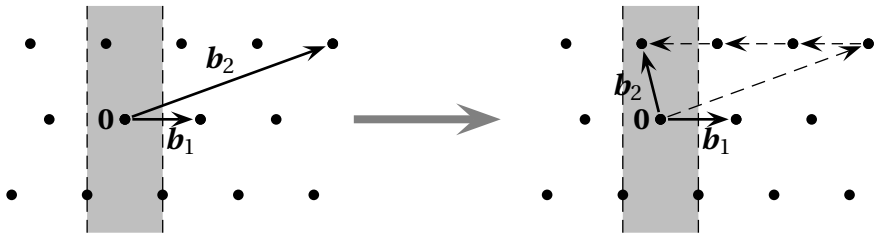
For the presentation of the LLL-algorithm, we are loosely following the exposition of Chris Peikert's excellent lecture notes [Pei13]. The main statement will be:

Theorem 1.13 (Lenstra-Lenstra-Lovász 1982)**.** *Given a regular matrix $\boldsymbol{B} \in \mathbb{Q}^{n\times n}$ one can compute a vector $\boldsymbol{x} \in \Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}$ of length $\|\boldsymbol{x}\|_2 \leq 2^{n/2} \cdot \lambda_1(\Lambda(\boldsymbol{B}))$ in polynomial time.*

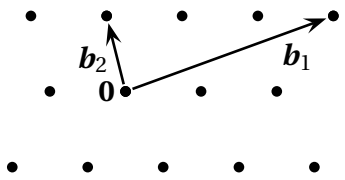
If $\boldsymbol{B}$ has integral entries then the running time is actually of the form $O(n^6 \log^3(n\|\boldsymbol{B}\|_\infty))$. The importance of this algorithm cannot be underestimated. Until now — 40 years after its discovery — the LLL-algorithm is basically the only algorithm that gives any kind of non-trivial guarantee for any lattice problem in polynomial time!

Let us consider a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ (here for $n = 2$) and wonder what kind of operations we could perform make the basis *as orthogonal as possible*, while it still generates the same lattice.

- *Subtracting vectors from each other:* In $n = 2$, if we have a vector $\boldsymbol{b}_2$, we can always subtract multiples of $\boldsymbol{b}_1$ from it so that $|\mu_{1,2}| \leq \frac{1}{2}$. In higher dimensions we will see that we can always achieve that $|\mu_{ij}| \leq \frac{1}{2}$ for all $i < j$.



- *Switching the order:* One the other hand, in $n = 2$ dimensions it might be that $\boldsymbol{b}_1$ is a lot longer than $\boldsymbol{b}_2$ so that we would not make progress in subtracting $\boldsymbol{b}_1$ from $\boldsymbol{b}_2$. But in that case we can swap the order of $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. In higher dimensions it will make sense to swap $\boldsymbol{b}_i$ and $\boldsymbol{b}_{i+1}$ if $\|\boldsymbol{b}_i\|_2 \gg \|\boldsymbol{b}_{i+1}\|_2$.



1.5.1 Coefficient reduction

We want to begin by discussing how to make use of the first procedure, where we subtract integer multiples of vectors in the basis from other basis vectors. Let $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ be a basis and let $\mu_{ij} := \frac{\langle \boldsymbol{b}_j, \boldsymbol{b}_i^* \rangle}{\|\boldsymbol{b}_i^*\|_2^2}$ be the coefficients from the Gram-Schmidt orthogonalisation.

Definition 1.14. We call a basis *Coefficient-reduced* if $|\mu_{ij}| \leq \frac{1}{2}$ for all $1 \leq i < j \leq n$.

Lemma 1.15. *Given any basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ one can compute a coefficient-reduced basis $\tilde{\boldsymbol{B}}$ in polynomial time so that $\Lambda(\tilde{\boldsymbol{B}}) = \Lambda(\boldsymbol{B})$ and the Gram-Schmidt orthogonalizations are identical.*

Proof. Suppose that $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ is a lattice basis and μ_{ij} are the Gram-Schmidt coefficients, that means

$$\boldsymbol{b}_j = \boldsymbol{b}_j^* + \sum_{i=1}^{j-1} \mu_{ij} \cdot \boldsymbol{b}_i^* \quad \forall\, j \in [n]. \qquad (*)$$

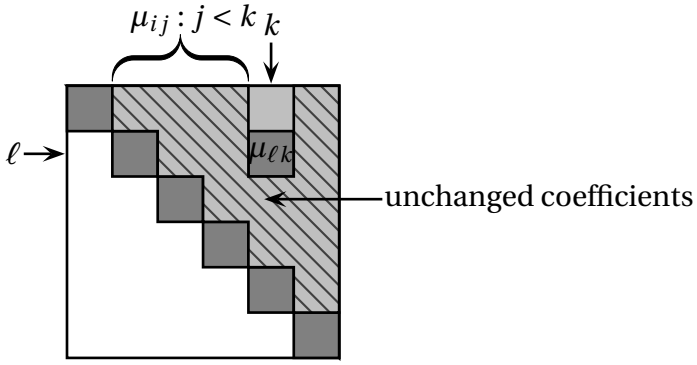
Now fix indices $1 \leq \ell < k \leq n$ and $q \in \mathbb{Z}$ and consider the updated basis $\tilde{\boldsymbol{B}}$ with vectors

$$\tilde{\boldsymbol{b}}_j := \begin{cases} \boldsymbol{b}_j + q \cdot \boldsymbol{b}_\ell & \text{if } j = k \\ \boldsymbol{b}_j & \text{otherwise.} \end{cases}$$

Clearly, $\Lambda(\tilde{\boldsymbol{B}}) = \Lambda(\boldsymbol{B})$. Let $\tilde{\mu}_{ij}$ be the updated Gram-Schmidt coefficients for $\tilde{\boldsymbol{B}}$. In particular $\tilde{\boldsymbol{b}}_1^* = \boldsymbol{b}_1^*, \ldots, \tilde{\boldsymbol{b}}_{k-1}^* = \boldsymbol{b}_{k-1}^*$ and for the coefficients we have $\tilde{\mu}_{ij} = \mu_{ij}$ for all pairs (i, j) with $i < j < k$ since only $\boldsymbol{b}_k$ has changed. Adding up $(*)$ for $j = k$ and $j = \ell$ we obtain

$$\tilde{\boldsymbol{b}}_k = \boldsymbol{b}_k + q \cdot \boldsymbol{b}_\ell = \boldsymbol{b}_k^* + \sum_{i<\ell} \underbrace{\left(\mu_{ik} + q \cdot \mu_{i\ell}\right)}_{=\tilde{\mu}_{ik}} \cdot \boldsymbol{b}_i^* + \underbrace{(\mu_{\ell k} + q)}_{=\tilde{\mu}_{\ell k}} \cdot \boldsymbol{b}_\ell^* + \sum_{i=\ell+1}^{k-1} \underbrace{\mu_{ik}}_{=\tilde{\mu}_{ik}} \boldsymbol{b}_i^*$$

Then we see that $\tilde{\mu}_{ik} = \mu_{ik}$ for all $i > \ell$. Moreover, we can choose $q \in \mathbb{Z}$ so that $|\tilde{\mu}_{\ell,k}| = |\mu_{\ell,k} + q| \leq \frac{1}{2}$. In fact, also $\tilde{\mu}_{ij} = \tilde{\mu}_{ij}$ for all $j > k$, though we would need that property. In the figure below we show which coefficients are guaranteed to not have changed:



Now suppose we denote the above procedure by

$$\tilde{\boldsymbol{B}} := \texttt{update}(\boldsymbol{B}, \ell, k) = \big(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k-1}, \boldsymbol{b}_k - \lceil \mu_{\ell,k} \rfloor \boldsymbol{b}_\ell, \boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_n\big).$$

Then it is clear that we need to go through all index pairs (ℓ, k) in the right order and we can bring all coefficients into the interval $[-\frac{1}{2}, \frac{1}{2}]$:

(1) FOR $k = 1$ TO n DO

  (2) FOR $\ell = k - 1$ DOWNTO 1 DO

    (3) $\boldsymbol{B} := \texttt{update}(\boldsymbol{B}, \ell, k)$

That shows the claim. □

1.5.2 The main procedure

The crucial definition in the LLL-algorithm is the following:

Definition 1.16. Let $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ be a lattice basis and let μ_{ij} be the coefficients from the Gram-Schmidt orthogonalization. The basis is called *LLL-reduced* if the following is satisfied

- *Coefficient reduced:* $|\mu_{i,j}| \leq \frac{1}{2}$ for all $1 \leq i < j \leq n$
- *Lovász condition:* $\|\boldsymbol{b}_i^*\|_2^2 \leq 2\|\boldsymbol{b}_{i+1}^*\|_2^2$ for $i = 1, \ldots, n-1$

First, let us see why this definition is desirable:

Lemma 1.17. *Let $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ be an LLL-reduced basis. Then $\|\boldsymbol{b}_1\|_2 \leq 2^{n/2} \cdot \lambda_1(\Lambda(\boldsymbol{B}))$.*

Proof. Note that the 1st vector in the basis has $\boldsymbol{b}_1 = \boldsymbol{b}_1^*$. Applying the Lovász condition gives

$$\|\boldsymbol{b}_1\|_2^2 = \|\boldsymbol{b}_1^*\|_2^2 \leq 2\|\boldsymbol{b}_2^*\|_2^2 \leq \ldots \leq 2^{i-1} \cdot \|\boldsymbol{b}_i^*\|_2^2$$

On the other hand we can use Theorem 1.11 to lower bound the length of the shortest vector:

$$\lambda_1(\Lambda(\boldsymbol{B}))^2 \geq \min_{i=1,\ldots,n} \|\boldsymbol{b}_i^*\|_2^2 \geq \min_{i=1,\ldots,n} 2^{-(i-1)}\|\boldsymbol{b}_1\|_2^2 \geq 2^{-n} \cdot \|\boldsymbol{b}_1\|_2^2$$

Taking square roots then gives the claim. □

This is the algorithm that will compute an LLL-reduced basis:

LLL-algorithm

Input: A lattice basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) \in \mathbb{R}^{n \times n}$
Output: An LLL reduced basis $\tilde{\boldsymbol{B}} = (\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_n)$

(1) Compute a Gram Schmidt orthogonalization $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$ with coefficients μ_{ij} and update whenever we change the order of the basis
(2) WHILE $\boldsymbol{B}$ is not LLL-reduced DO
  (3) Apply coefficient reduction so that $|\mu_{ij}| \leq \frac{1}{2}$
  (4) If there is an index i with $\|\boldsymbol{b}_i^*\|_2^2 > 2\|\boldsymbol{b}_{i+1}^*\|_2^2$ then swap $\boldsymbol{b}_i$ and $\boldsymbol{b}_{i+1}$ in the ordering.

Obviously, the algorithm only terminates when it has found an LLL-reduced basis. Let us now prove that the algorithm terminates within a polynomial number of iterations. For this sake, we consider the *potential function*

$$\Phi(\boldsymbol{B}) = \prod_{i=1}^{n} \|\boldsymbol{b}_i^*\|_2^{n+1-i}$$

and we want to argue that it is decreasing. Intuitively, the potential function wants the vectors $\boldsymbol{b}_i^*$ with small i to be as small as possible. In particular, the point is that we swap i with $i+1$ if $\|\boldsymbol{b}_i^*\|_2$ is a lot longer than $\|\boldsymbol{b}_{i+1}^*\|_2$; one can expect that this should decrease the potential function. Let us define

$$\mathrm{Vol}_k(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$$

as the k-dimensional volume of the parallelepiped spanned by $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$. Note that by orthogonalizing the vectors, we do not change the volume of that parallelepiped (this is again *Cavalieri's principle*). Hence

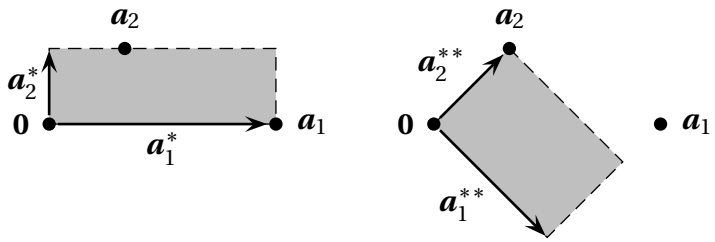
$$\mathrm{Vol}_k(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k) = \prod_{i=1}^{k} \|\boldsymbol{b}_i^*\|_2$$

We can use this to rewrite the potential function as

$$\Phi(\boldsymbol{B}) = \prod_{k=1}^{n} \mathrm{Vol}_k(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$$

We want to really understand what's going on, so here is a standalone lemma:

Lemma 1.18. *Suppose we have vectors $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ with Gram Schmidt orthogonalization $(\boldsymbol{a}_1^*, \boldsymbol{a}_2^*)$ so that $\|\boldsymbol{a}_1^*\|_2^2 \geq 2\|\boldsymbol{a}_2^*\|_2^2$ and $\mu := \frac{\langle \boldsymbol{a}_1, \boldsymbol{a}_2 \rangle}{\|\boldsymbol{a}_1\|_2^2} \leq \frac{1}{2}$. Let $(\boldsymbol{a}_2^{**}, \boldsymbol{a}_1^{**})$ be the Gram Schmidt Orthogonalization for the reverse order $(\boldsymbol{a}_2, \boldsymbol{a}_1)$. Then $\|\boldsymbol{a}_2^{**}\|_2 \leq \sqrt{\frac{3}{4}} \cdot \|\boldsymbol{a}_1^*\|_2$.*



Proof. Let us first write down the vectors in both orthogonalizations depending on $\boldsymbol{a}_1, \boldsymbol{a}_2$ and the inner product μ:

$$\boldsymbol{a}_1^* = \boldsymbol{a}_1, \qquad \boldsymbol{a}_2^* = \boldsymbol{a}_2 - \underbrace{\frac{\langle \boldsymbol{a}_1, \boldsymbol{a}_2 \rangle}{\|\boldsymbol{a}_1\|_2^2}}_{=\mu} \boldsymbol{a}_1, \qquad \text{and} \qquad \boldsymbol{a}_2^{**} = \boldsymbol{a}_2$$

This can be rewritten to $\boldsymbol{a}_2^{**} = \boldsymbol{a}_2 = \boldsymbol{a}_2^* + \mu \boldsymbol{a}_1^*$. We inspect the square of the desired ratio and get:

$$\frac{\|\boldsymbol{a}_2^{**}\|_2^2}{\|\boldsymbol{a}_1^*\|_2^2} = \frac{\|\boldsymbol{a}_2^* + \mu \cdot \boldsymbol{a}_1^*\|_2^2}{\|\boldsymbol{a}_1^*\|_2^2} \overset{\text{Pythagoras \& } \boldsymbol{a}_1^* \perp \boldsymbol{a}_2^*}{=} \frac{\|\boldsymbol{a}_2^*\|_2^2 + \mu^2 \|\boldsymbol{a}_1^*\|_2^2}{\|\boldsymbol{a}_1^*\|_2^2} \leq \underbrace{\frac{\|\boldsymbol{a}_2^*\|_2^2}{\|\boldsymbol{a}_1^*\|_2^2}}_{\leq 1/2} + \frac{1}{4} \leq \frac{3}{4}$$

□

Lemma 1.19. *In each iteration $\Phi(\boldsymbol{B})$ reduces by a constant factor.*

Proof. First, observe that making $\boldsymbol{B}$ coefficient-reduced does not change the Gram-Schmidt orthogonalization and hence leaves the potential function $\Phi(\boldsymbol{B})$ unchanged. Now, let $\boldsymbol{B}$ be a coefficient-reduced basis at the beginning of (4) and let $\tilde{\boldsymbol{B}}$ be the

basis after we swapped index i. Then the whole Gram-Schmidt orthogonalisation of $\boldsymbol{B}$ and $\tilde{\boldsymbol{B}}$ is identical — except for vectors i and $i+1$. We can hence write

$$\frac{\Phi(\tilde{\boldsymbol{B}})}{\Phi(\boldsymbol{B})} = \frac{\mathrm{Vol}_i(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{i-1},\boldsymbol{b}_{i+1})}{\mathrm{Vol}_i(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{i-1},\boldsymbol{b}_i)}$$

Let $\boldsymbol{a}_1$ be the projection of $\boldsymbol{b}_i$ on the subspace $U := \mathrm{span}\{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{i-1}\}^\perp$ and let $\boldsymbol{a}_2$ be the projection of $\boldsymbol{b}_{i+1}$ on that subspace. In the notation of the previous lemma, let $(\boldsymbol{a}_1^*,\boldsymbol{a}_2^*)$ be the orthogonalization of $(\boldsymbol{a}_1,\boldsymbol{a}_2)$ (in that order) and let $(\boldsymbol{a}_2^*,\boldsymbol{a}_1^*)$ be the orthogonalization of $(\boldsymbol{a}_2,\boldsymbol{a}_1)$ (again in that order). Then $\boldsymbol{a}_1^* = \boldsymbol{b}_i^*$ and $\boldsymbol{a}_2^{**} = \tilde{\boldsymbol{b}}_{i+1}^*$. Since the volumes of both parallelepipeds is proportional to the distance of the last vector to the subspace U, we get

$$\frac{\Phi(\tilde{\boldsymbol{B}})}{\Phi(\boldsymbol{B})} = \frac{\mathrm{vol}_i(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{i-1},\boldsymbol{b}_{i+1})}{\mathrm{vol}_i(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{i-1},\boldsymbol{b}_i)} = \frac{\|\boldsymbol{a}_2^{**}\|_2}{\|\boldsymbol{a}_1^*\|_2} \le \sqrt{\frac{3}{4}} < 1$$

using Lemma 1.18. □

Now it is easy to show that the LLL is a polynomial time algorithm. Originally, $\boldsymbol{B}$ could have been any matrix with rational entries. But we can scale any such matrix so that the entries become integral. Note that an entry B_{ij} used $\approx \log_2(|B_{ij}|)$ bits in the input. Hence, a polynomial time algorithm should have a running time that is polynomial in n and in $\log(\|\boldsymbol{B}\|_\infty)$. Moreover, the squared volume of a parallelepiped that is spanned by integral vectors will be integral.

Lemma 1.20. *Let $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k \in \mathbb{Z}^n$ be linearly independent integral vectors. Then $Vol_k(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k)^2 \in \mathbb{Z}_{\ge 1}$.*

Proof sketch. We have restricted our attention to full rank lattices so far. But there are more general definitions for lattices of lower rank. Let $\boldsymbol{B} = (\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k) \in \mathbb{Z}^{n\times k}$. Consider the lattice $\Lambda(\boldsymbol{B})$ spanned by $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k$. One can show that[5]

$$\mathrm{Vol}_k(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k) = \sqrt{\det_k(\boldsymbol{B}^T\boldsymbol{B})}$$

Since $\boldsymbol{B}^T\boldsymbol{B} \in \mathbb{Z}^{k\times k}$ is an integral matrix one has $\det_k(\boldsymbol{B}^T\boldsymbol{B}) \in \mathbb{Z}$ and the claim follows. □

The last lemma implies that in particular $\mathrm{vol}_k(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k) \ge 1$ for all $k = 1,\ldots,n$.

Lemma 1.21. *Suppose that $\boldsymbol{B} \in \mathbb{Z}^{n\times n}$. Then the LLL-algorithm applied to $\boldsymbol{B}$ takes $O(n^2 \log\max\{n, \|\boldsymbol{B}\|_\infty\})$ many iterations.*

[5] The reader may want to double-check that in case of $k = n$, this is exactly $|\det(\boldsymbol{B})|$.

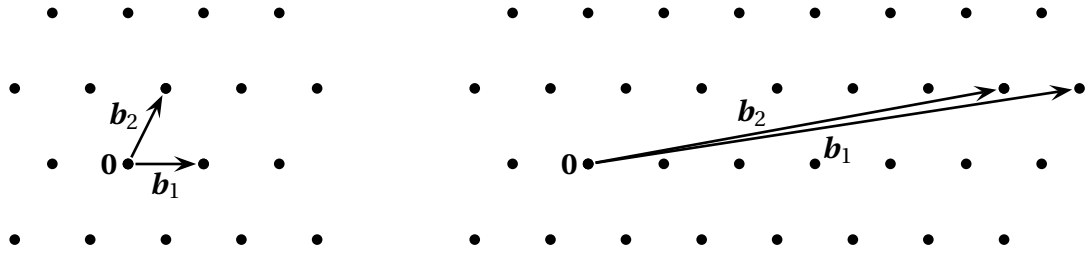
Proof. Suppose that $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are the original columns of the input matrix $\boldsymbol{B}$. Since the Gram-Schmidt process is a projection, it cannot make vectors longer. Hence $\|\boldsymbol{b}_i^*\|_2 \leq \|\boldsymbol{b}_i\|_2 \leq \sqrt{n} \cdot \|\boldsymbol{B}\|_\infty$. Hence before the first iteration, the potential function is bounded by $\Phi(\boldsymbol{B}) \leq (\sqrt{n} \cdot \|\boldsymbol{B}\|_\infty)^{n^2}$. Now, let $\tilde{\boldsymbol{B}} = (\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_n)$ be a matrix in an arbitrary iteration of the LLL algorithm. Since $\tilde{\boldsymbol{B}}$ is obtained by subtracting and permuting columns in the integral matrix $\boldsymbol{B}$, we know that $\tilde{\boldsymbol{B}} \in \mathbb{Z}^{n \times n}$. As we observed earlier, we have

$$\Phi(\tilde{\boldsymbol{B}}) = \prod_{k=1}^{n} \underbrace{\mathrm{Vol}_k(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)}_{\geq 1} \geq 1$$

Since the potential function decreases by a constant factor in each iteration, the claim follows. □

1.5.3 The orthogonality defect

We want to further discuss that the LLL-algorithm does not only find a short vector, but the LLL-reduced basis has a lot more properties. The LLL-reduced basis is really a "good" basis in the sense that at least it is approximately orthogonal.



"good" basis "bad" basis

For a lattice basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, we define the *orthogonality defect* as

$$\gamma(\boldsymbol{B}) := \frac{\prod_{i=1}^{n} \|\boldsymbol{b}_i\|_2}{\prod_{i=1}^{n} \|\boldsymbol{b}_i^*\|_2}$$

Note that $\gamma(\boldsymbol{B}) \geq 1$ and we have $\gamma(\boldsymbol{B}) = 1$ if and only if $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are pairwise orthogonal. So, $\gamma(\boldsymbol{B})$ is indeed a measure for how orthogonal a basis is. Even from a non-constructive viewpoint, it is non-trivial to argue that for every lattice there even exists a basis with $\gamma(\boldsymbol{B})$ bounded by some function of n.

Lemma 1.22. *The orthogonality defect of an LLL-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ is $\gamma(\boldsymbol{B}) \leq 2^{n^2/2}$. Moreover one has $\|\boldsymbol{b}_k\|_2 \leq 2^{k/2} \|\boldsymbol{b}_k^*\|_2$ for all $k = 1, \ldots, n$.*

Proof. Now it is important that we also have control over intermediate vectors. Recall that by the properties of the Gram-Schmidt orthogonalization, we have

$$\boldsymbol{b}_k = \boldsymbol{b}_k^* + \sum_{i=1}^{k-1} \mu_{ik} \boldsymbol{b}_i^*.$$

Taking norms, we get

$$\|\boldsymbol{b}_k\|_2^2 = \left\| \boldsymbol{b}_k^* + \sum_{i=1}^{k-1} \mu_{ik} \boldsymbol{b}_i^* \right\|_2^2 = \|\boldsymbol{b}_k^*\|_2^2 + \sum_{i=1}^{k-1} \underbrace{\mu_{ik}^2}_{\leq 1/4} \underbrace{\|\boldsymbol{b}_i^*\|_2^2}_{\leq 2^{k-i}\|\boldsymbol{b}_k^*\|_2^2} \leq \|\boldsymbol{b}_k^*\|_2^2 \cdot \underbrace{\left(1 + \frac{1}{4}\sum_{i=1}^{k-1} 2^{k-i}\right)}_{\leq 2^k} \leq 2^k \cdot \|\boldsymbol{b}_k^*\|_2^2$$

using that $\|\boldsymbol{b}_j^*\|_2^2 \leq 2\|\boldsymbol{b}_{j+1}^*\|_2^2$ for all $j = 1, \ldots, n-1$. Then being generous with the constants, we get that

$$\gamma(\boldsymbol{B}) = \prod_{k=1}^{n} \underbrace{\frac{\|\boldsymbol{b}_k\|_2}{\|\boldsymbol{b}_k^*\|_2}}_{\leq 2^{n/2}} \leq 2^{n^2/2}.$$

□

Lemma 1.23. *An LLL-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ satisfies* $\max_{i=1,\ldots,n} \|\boldsymbol{b}_i\|_2 \leq 2^n \lambda_n(\Lambda(\boldsymbol{B}))$.

Proof. Set $\Lambda := \Lambda(\boldsymbol{B})$. Consider the subspace $U := \operatorname{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}\}$ and fix a point $\boldsymbol{x} = \sum_{i=1}^{n} \boldsymbol{b}_i z_i \in \Lambda$ with $z_n \neq 0$ and $\|\boldsymbol{x}\|_2 \leq \lambda_n(\Lambda)$. Then $\lambda_n(\Lambda) \geq \|\boldsymbol{x}\|_2 \geq d(\boldsymbol{x}, U) = |z_n| \cdot \|\boldsymbol{b}_n^*\|_2 \geq \|\boldsymbol{b}_n^*\|_2$. So it remains to relate the length of the basis vectors to $\boldsymbol{b}_n^*$.

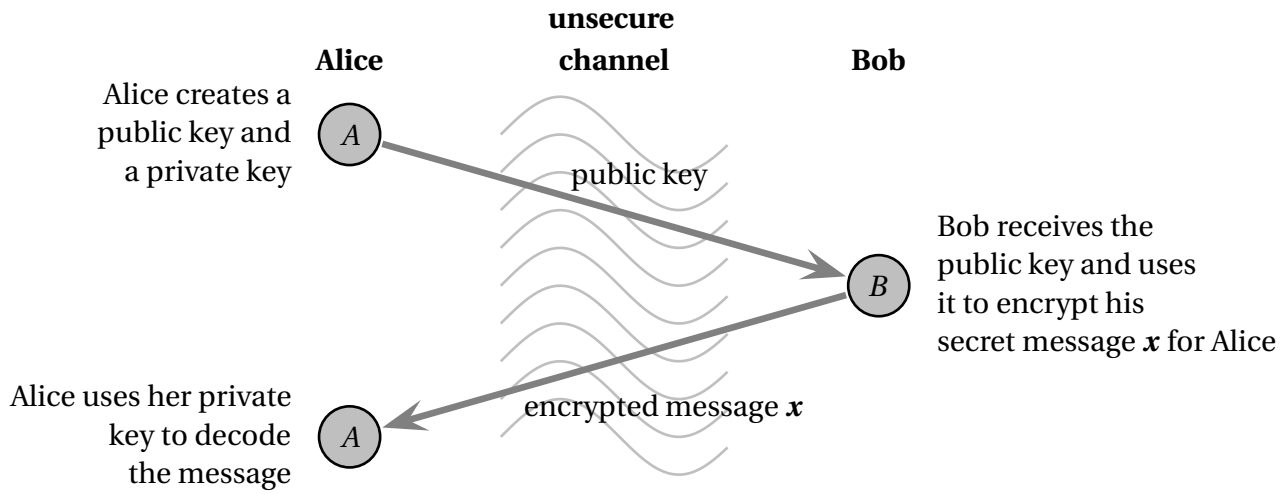
Let j be the index with $\|\boldsymbol{b}_j\|_2 = \max_{i=1,\ldots,n} \|\boldsymbol{b}_i\|_2$. Then $\|\boldsymbol{b}_j\|_2 \leq 2^{n/2}\|\boldsymbol{b}_j^*\|_2 \leq 2^n \|\boldsymbol{b}_n^*\|_2 \leq 2^n \lambda(\Lambda)$ where we use the "moreover" part of Lemma 1.22 as well as the fact that $\|\boldsymbol{b}_i^*\|_2^2 \leq 2\|\boldsymbol{b}_{i+1}^*\|_2^2$ for all $i = 1, \ldots, n-1$ as the basis is LLL-reduced. □

A different definition of a reduced basis is due to Korkine and Zolotarev and is called *K-Z-reduced basis*. Such a basis has an orthogonality defect of at most n^n and we will see the construction later in Section 1.9. However, no polynomial time algorithm is known to compute such a basis.

1.6 Breaking Knapsack Cryptosystems

The approximation guarantee of $2^{n/2}$ provided by the LLL-algorithm may sound weak, but it is already enough for a couple of very surprising applications. We want to show-case one of them here.

For a *public key cryptosystem*, the goal is that two parties A (say Alice) and B (say, Bob) can communicate a secret message over a public channel without that any third party C could decrypt it.



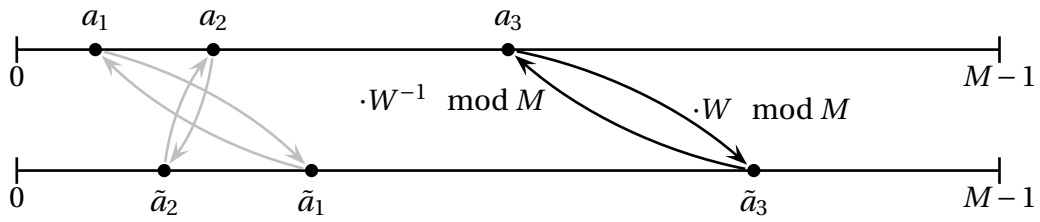
To be more precise, there will be a *public key*, which Alice would generate and then put on her webpage. Then Bob could see that public key and he would use it to encrypt a message that he wants to send to Alice. Even if a third party wiretaps the message that Bob sends to Alice and knows the public key, we want that C is still unable to decrypt the message. The important fact behind such a cryptosystem is that Alice and Bob do not need to agree on any key in advance – it suffices to communicate over the unsecure channel.

The idea behind the *Knapsack cryptosystem* is the following: Alice could create some large numbers $a_1, \ldots, a_n \in \mathbb{N}$ and publish them as public key. If Bob wants to send a secret message $\boldsymbol{x} \in \{0,1\}^n$ to Alice, he could compute the sum $S := \sum_{i=1}^n a_i x_i$ and send Alice the number S. Knapsack is an **NP**-hard problem, so Bob could hope that his message is safe enough. But in any meaningful cryptosystem, at least the intended receiver Alice should be able to decrypt the message efficiently. So, the Knapsack instance has to be "simple". One way of having an easy Knapsack instance is if the numbers a_i are *super-increasing*, that means there is a permutation $\pi : [n] \to [n]$ so that

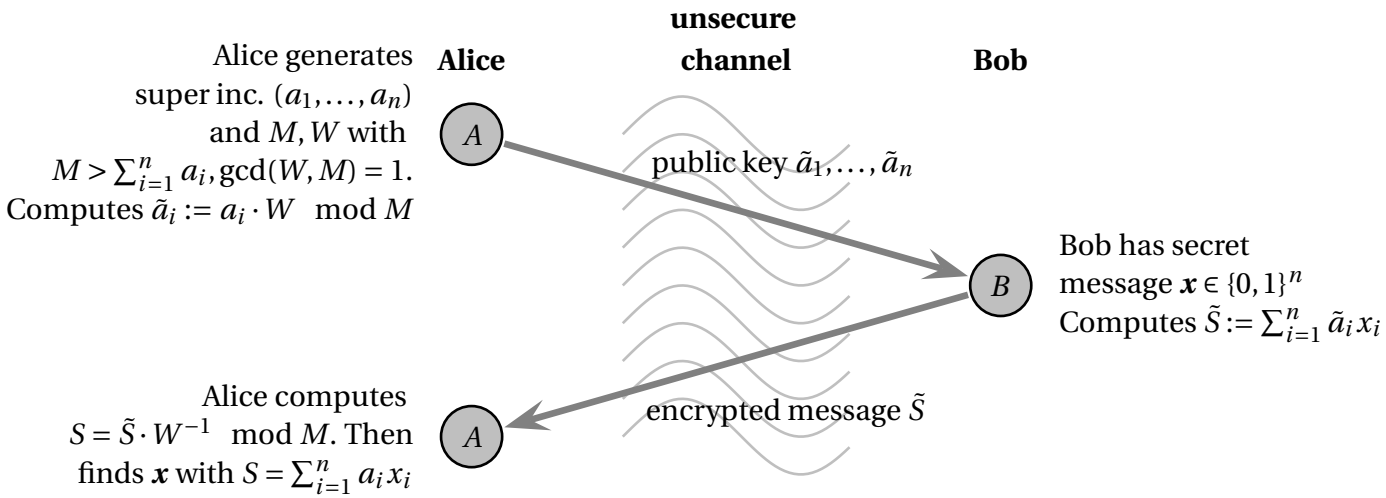
$$a_{\pi(i)} > \sum_{j:\pi(j)<\pi(i)} a_j \quad \forall\, i \in [n].$$

It is a simple exercise to show that in this case, given a sum $S = \sum_{i=1}^n a_i x_i$, one can recover the vector $\boldsymbol{x} \in \{0,1\}^n$ with a polynomial time greedy-style algorithm.

But of course, also a third party C would know how to solve such a Knapsack problem. So, we need one more ingredient to "hide" the super increasing sequence. We simply take a large number $M > \sum_{i=1}^{n} a_i$ and some random $W \in \{1, \ldots, M-1\}$ with $\gcd(M, W) = 1$ and compute numbers $\tilde{a}_i := a_i \cdot W \mod M$. Here it is helpful to remember that the map $a \mapsto a \cdot W \mod M$ is a bijection on $\{0, \ldots, M-1\}$ and the inverse function is simply $a \mapsto a \cdot W^{-1} \mod M$. The inverse W^{-1} with $W \cdot W^{-1} \equiv_M 1$ exists as M and W are coprime. We use that function to randomly "shuffle" the numbers:



Now, those numbers $\tilde{a}_1, \ldots, \tilde{a}_n$ form the public key. Note that the numbers of the super increasing sequence are now wildly mixed. Now, Bob receives the public key $\tilde{a}_1, \ldots, \tilde{a}_n$ and computes the sum $\tilde{S} = \sum_{i=1}^{n} \tilde{a}_i x_i$. Then he sends the message $\tilde{S}$ to Alice. Alice computes $S := \tilde{S} \cdot W^{-1} \mod M$ and then uses the super-increasing property to compute the unique vector $\boldsymbol{x}$ satisfying $S = \sum_{i=1}^{n} a_i x_i$ (here we use that $\sum_{i=1}^{n} a_i x_i \leq M$). Note that her *private key* consists of the pair (M, W) (and the permutation of the indices). The whole public key Knapsack crypto system can be visualized as follows:



It seemed that in order to decrypt the message without knowing the private key (the numbers M, W), one would have to solve the Knapsack problem. The remaining ingredient for a working cryptosystem would be the generation of a hard

Knapsack instance. Intuitively one might think that taking a random instance with large enough coefficients would give such a hard instance. Surprisingly, this is false due to the LLL algorithm.

1.6.1 A polynomial time algorithm to solve sparse knapsack instance

While Knapsack is **NP**-hard in the worst case, it turned out that very sparse Knapsack instances admit polynomial time algorithms. Note that here, we do not try to optimize any constant.

Theorem 1.24 (Lagarias, Odlyzko 1985)**.** *Suppose we generate a Knapsack instance by picking independently $a_1, \dots, a_n \sim \{\frac{1}{2} \cdot 2^{4n^2}, \dots, 2^{4n^2}\}$ at random. Then take a vector $\boldsymbol{x} \in \{0,1\}^n$ and compute $S := \sum_{i=1}^n a_i x_i$. Then there is a polynomial time algorithm which on input $(\boldsymbol{a}, S)$, with high probability recovers the vector $\boldsymbol{x}$.*

Proof. Let us define an $(n+1)$-dimensional basis

$$\boldsymbol{B} = (\boldsymbol{b}_0, \boldsymbol{b}_1, \dots, \boldsymbol{b}_n) = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 2^n S & -2^n a_1 & -2^n a_2 & -2^n a_3 & \dots & -2^n a_n \end{pmatrix}$$

and let us consider the lattice $\Lambda(\boldsymbol{B})$ that is spanned by the columns $\boldsymbol{b}_0, \dots, \boldsymbol{b}_n$ of this matrix. First of all, we claim that this lattice contains a very short vector, namely $\boldsymbol{b}_0 + \sum_{i=1}^n x_i \boldsymbol{b}_i = \begin{pmatrix} \boldsymbol{x} \\ 0 \end{pmatrix}$. The length of this vector is $\|\boldsymbol{x}\|_2 \leq \sqrt{n}$. Hence, we can use the LLL-algorithm to compute a vector in the lattice $\Lambda(\boldsymbol{B})$ of length at most $\|\boldsymbol{x}\|_2 \cdot 2^{n/2} \ll 2^n$. It remains to show that any such short vector is actually a multiple of $(\boldsymbol{x}, 0)$ (in fact, the LLL-algorithm returns a basis and the following claim even implies that the shortest vector in that basis must be $\pm(\boldsymbol{x}, 0)$):

Claim I. *With high probability, the only vectors $\boldsymbol{z} \in \Lambda(\boldsymbol{B})$ with $\|\boldsymbol{z}\|_2 < 2^n$ are multiples of $(\boldsymbol{x}, 0)$.*

Let $\boldsymbol{z} \in \Lambda(\boldsymbol{B})$ be a lattice vector with $\|\boldsymbol{z}\|_2 < 2^n$. We can write $\boldsymbol{z} = \alpha \boldsymbol{b}_0 + \sum_{i=1}^n y_i \boldsymbol{b}_i$ for integer coefficients $\alpha, y_1, \dots, y_n \in \mathbb{Z}$. The last coordinate of $\boldsymbol{z}$ is $2^n \cdot (\alpha S - \sum_{i=1}^n y_i a_i)$. Since the absolute value of this number has to be less than 2^n, we know that indeed $\sum_{i=1}^n y_i a_i = \alpha S$. Since $\|\boldsymbol{z}\|_2 < 2^n$, we also know that $\|\boldsymbol{y}\|_\infty \leq 2^n$. Then $\alpha \leq 2n \cdot 2^n$, since otherwise, we would have $|\alpha| S > \sum_{i=1}^n a_i y_i$.

Note that the number of triples $(\boldsymbol{x}, \boldsymbol{y}, \alpha)$ with $\boldsymbol{x} \in \{0,1\}^n$, $\|\boldsymbol{y}\|_\infty \leq 2^n$ and $|\alpha| \leq 2n \cdot 2^n$ is bounded by $2n \cdot 2^{n\cdot(n+1)} \cdot 2^n \cdot 3^{2n+1} \leq 2^{3n^2}$. So, in order to finish Claim I it suffices to show the following:

Claim II. *Fix a triple* $(\boldsymbol{x}, \boldsymbol{y}, \alpha)$ *with* $\boldsymbol{y} \notin \mathbb{Z}\boldsymbol{x}$. *Then*

$$\Pr_{\boldsymbol{a} \sim [2^{4n^2}]^n} \left[\sum_{i=1}^{n} y_i a_i = \alpha \cdot \left(\sum_{i=1}^{n} a_i x_i \right) \right] \leq \frac{1}{2} \cdot 2^{-4n^2} \quad (*)$$

Proof of Claim II. By assumption $\boldsymbol{y}$ is not a multiple of $\boldsymbol{x}$, so we know that so there is an index j with $y_j \neq \alpha x_j$. Now suppose that the value of $a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_n$ has been fixed and we just pick $a_j \sim \{\frac{1}{2} \cdot 2^{4n^2}, \ldots, 2^{4n^2}\}$ at random. Then the equation in $(*)$ can be rearranged to

$$a_j \underbrace{(y_j - \alpha x_j)}_{\neq 0} = \alpha \sum_{i \neq j} x_i a_i - \sum_{i \neq j} y_i a_i \quad (**)$$

It doesn't actually matter what the right hand side of $(**)$ is, just observe that there will be at most one choice of a_j that could satisfy the equation. The bound on the probability follows since we are drawing a_j from $\{\frac{1}{2} \cdot 2^{4n^2}, \ldots, 2^{4n^2}\}$. □

The original attack on the Knapsack cryptosystem was by Shamir. The generalized argument for low-density subset-sum problem is by Lagarias and Odlyzko with a later simplification of Frieze. For more information, we recommend the survey *The Rise and Fall of Knapsack Cryptosystems* by Odlyzko [Odl90].

1.7 The dual lattice and applications

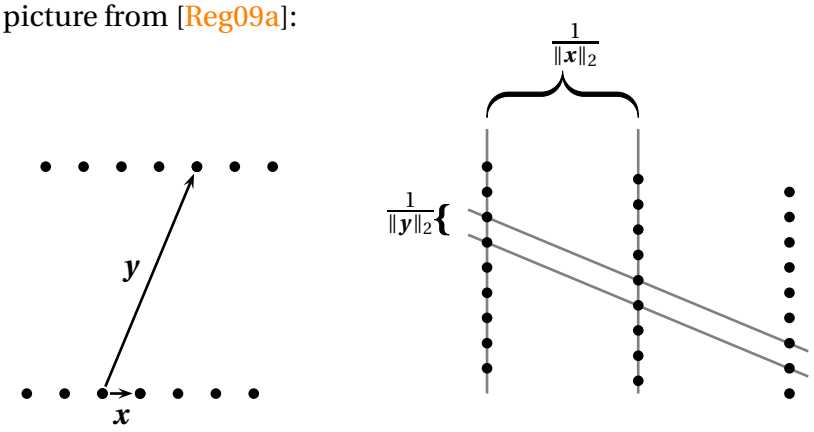
An important point is that Minkowski's Theorem is inherently *non-constructive*, that means given a symmetric convex set K with $\mathrm{Vol}_n(K) > 2^n$, we know it must contain a non-zero integer point — but the proof method does not provide a polynomial time algorithm to find that point. One might think that this is a pure artefact of the proof and a different proof technique will be algorithmic. But such a proof would have tremendous consequences. The reader may be reminded, that there are cryptosystems that rely on the assumption that it is hard to approximate the Shortest Vector Problem up to small polynomial factors. Moreover, one can show that even an approximate constructive proof for Minkowski's Theorem would imply an approximation algorithm for $\lambda_1(\Lambda)$. The proof is via an application of the concept of dual lattices.

1.7.1 Dual lattices

Let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice. Then the *dual lattice* is

$$\Lambda^* = \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{x} \rangle \in \mathbb{Z} \ \forall \boldsymbol{x} \in \Lambda\}$$

In order to get some intuition, take a lattice vector $\boldsymbol{x} \in \Lambda$. Then all dual lattice vectors lie on the hyperplanes $H_k = \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle = k\}$ with $k \in \mathbb{Z}$. In particular, each dual vector $\boldsymbol{y} \in \Lambda^*$ will either be orthogonal to $\boldsymbol{x}$ or it will have a separation of at least $\frac{1}{\|\boldsymbol{x}\|_2}$ from H_0. Loosely speaking, if lattice vectors in some direction are short in Λ, they will be long in this direction in Λ^* and vice versa. We reproduce a helpful picture from [Reg09a]:



lattice Λ dual lattice Λ^*

The next lemma will show that Λ^* is indeed a lattice and we will see that its basis is just the transpose inverse of the original basis:

Lemma 1.25. *Let $\Lambda := \Lambda(\boldsymbol{B})$ be a full rank lattice generated by $\boldsymbol{B} \in \mathbb{R}^{n \times n}$. Then $(\boldsymbol{B}^{-1})^T$ is a lattice basis for the dual lattice Λ^*.*

Proof. Let us define $\Lambda' := \{(\boldsymbol{B}^{-1})^T \boldsymbol{\mu} \mid \boldsymbol{\mu} \in \mathbb{Z}^n\}$ as the candidate lattice. For any vectors $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{\lambda} \in \Lambda$ and $\boldsymbol{y} = (\boldsymbol{B}^{-1})^T \boldsymbol{\mu}$ with coefficients $\boldsymbol{\lambda}, \boldsymbol{\mu} \in \mathbb{Z}^n$, the inner product is

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = (\boldsymbol{B}\boldsymbol{\lambda})^T ((\boldsymbol{B}^{-1})^T \boldsymbol{\mu}) = \boldsymbol{\lambda}^T \underbrace{\boldsymbol{B}^T (\boldsymbol{B}^{-1})^T}_{=\boldsymbol{I}_n} \boldsymbol{\mu} = \langle \boldsymbol{\lambda}, \boldsymbol{\mu} \rangle \in \mathbb{Z}$$

and hence $\Lambda' \subseteq \Lambda^*$.

For the other direction, take a point $\boldsymbol{y} \in \Lambda^*$. Since $(\boldsymbol{B}^{-1})^T$ has full rank, there must be a unique vector $\boldsymbol{\mu} \in \mathbb{R}^n$ with $\boldsymbol{y} = (\boldsymbol{B}^{-1})^T \boldsymbol{\mu}$. Then

$$\mathbb{Z} \overset{\text{Def. dual lattice}}{\supseteq} \{\langle \boldsymbol{x}, \boldsymbol{y} \rangle : \boldsymbol{x} \in \Lambda\} = \{\langle \boldsymbol{\lambda}, \boldsymbol{\mu} \rangle : \boldsymbol{\lambda} \in \mathbb{Z}^n\} \overset{\boldsymbol{\lambda} = \boldsymbol{e}_i}{\supseteq} \{\mu_1, \ldots, \mu_n\}$$

We see that $\boldsymbol{\mu} \in \mathbb{Z}^n$ and hence $\Lambda^* \subseteq \Lambda'$. □

We can use this insight to derive

Lemma 1.26. *For a full rank lattice $\Lambda = \Lambda(\boldsymbol{B})$ the following holds:*

(i) $(\Lambda^*)^* = \Lambda$
(ii) $\det(\Lambda^*) = \frac{1}{\det(\Lambda)}$.

Proof. The first claim follows from the last lemma with the observation that $(((\boldsymbol{B}^{-1})^T)^{-1})^T = \boldsymbol{B}$. The 2nd claim follows from $\det(\Lambda^*)\cdot\det(\Lambda) = |\det_n((\boldsymbol{B}^{-1})^T)\cdot\det_n(\boldsymbol{B})| = |\det_n(\boldsymbol{B}^{-1}\boldsymbol{B})| = |\det_n(\boldsymbol{I}_n)| = 1$. □

1.7.2 Solving Shortest Vector via Minkowski's Theorem

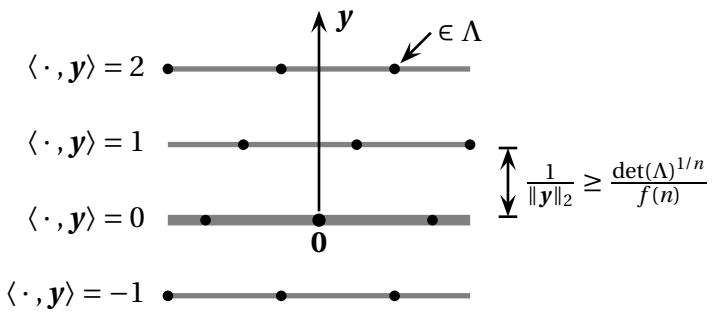
We will implicitly use the following lemma — we postpone the proof to Section 1.8.

Lemma 1.27. *Let $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ be a basis for lattice $\Lambda := \Lambda(\boldsymbol{B})$ and let $\boldsymbol{y} \in \mathbb{Q}^n$. Then there is a polynomial time algorithm to find a lattice basis $\boldsymbol{B}'$ for the sublattice $\Lambda' := \{\boldsymbol{x} \in \Lambda \mid \langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0\}$.*

Recall that Minkowski's Theorem guarantees that there is always a lattice vector $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_2 \le \sqrt{n}\cdot\det(\Lambda)^{1/n}$. Now we show the proof that an approximate version of Minkowski's theorem would imply an approximation algorithm for the Shortest Vector Problem.

Theorem 1.28 (Lenstra-Schnorr 1990)**.** *Suppose that we have a polynomial time algorithm that for any n-dimensional lattice Λ is able to find a vector $\boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}$ with $\|\boldsymbol{x}\|_2 \le f(n)\cdot\det(\Lambda)^{1/n}$, where $f(n)$ is a non-decreasing function. Then there exists a polynomial time $f(n)^2$-approximation algorithm for the Shortest Vector problem in any lattice.*

Proof. First, we use the assumed algorithm to find a lattice vector $\boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}$ of length $\|\boldsymbol{x}\|_2 \le f(n)\cdot\det(\Lambda)^{1/n}$. Is that point $\boldsymbol{x}$ a good approximation for the shortest vector? Well, not always as it is perfectly possible that $\lambda_1(\Lambda) \ll \det(\Lambda)^{1/n}$. So, we apply the algorithm again, but now to find a vector $\boldsymbol{y} \in \Lambda^* \setminus \{\boldsymbol{0}\}$ in the *dual lattice*, with $\|\boldsymbol{y}\|_2 \le f(n)\cdot\det(\Lambda^*)^{1/n} = \frac{f(n)}{\det(\Lambda)^{1/n}}$. Crucially, by definition of the dual lattice, we know that $\langle \boldsymbol{x}', \boldsymbol{y}\rangle \in \mathbb{Z}$ for all $\boldsymbol{x}' \in \Lambda$. In other words, all lattice vectors in Λ lie on hyperplanes of the form $\langle \cdot, \boldsymbol{y}\rangle \in \mathbb{Z}$.



Let $\boldsymbol{x}^* \in \Lambda \setminus \{\boldsymbol{0}\}$ be the unknown shortest vector. We distinguish two cases:

- *Case 1:* One has $\langle \boldsymbol{x}^*, \boldsymbol{y} \rangle \in \mathbb{Z} \setminus \{\boldsymbol{0}\}$. The distance of the hyperplanes $\langle \cdot, \boldsymbol{y} \rangle = \mathbb{Z}$ to each other is $\frac{1}{\|\boldsymbol{y}\|_2}$, hence we have a lower bound of $\|\boldsymbol{x}^*\|_2 \geq \frac{1}{\|\boldsymbol{y}\|_2} \geq \frac{\det(\Lambda)^{1/n}}{f(n)}$. Our found lattice vector has length $\|\boldsymbol{x}\|_2 \leq f(n) \cdot \det(\Lambda)^{1/n} \leq f(n)^2 \cdot \|\boldsymbol{x}^*\|_2$, hence $\boldsymbol{x}$ is the desired $f(n)^2$-approximation.

- *Case 2:* One has $\langle \boldsymbol{x}^*, \boldsymbol{y} \rangle = 0$. In this case, the shortest vector $\boldsymbol{x}^*$ lies in the sublattice $\Lambda' := \{\boldsymbol{x}' \in \Lambda : \langle \boldsymbol{x}', \boldsymbol{y} \rangle = 0\}$ of rank $n-1$. We apply our approximation algorithm recursively to that sublattice and inductively obtain a lattice vector $\boldsymbol{x}' \in \Lambda' \setminus \{\boldsymbol{0}\}$ of length $\|\boldsymbol{x}'\|_2 \leq f(n-1)^2 \cdot \lambda_1(\Lambda') \leq f(n)^2 \cdot \lambda_1(\Lambda)$. Here we have used that f is non-decreasing. Implicitly, we also used Lemma 1.27 to compute the basis for the sublattice Λ' in polynomial time.

While we do not know in our algorithm in which case we are in, we can compare the length $\boldsymbol{x}$ with the length of the vector provided by Case 2 and return the shorter one. □

We should remark that the rank of the lattice $\{\boldsymbol{x} \in \Lambda \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0\}$ is indeed $n-1$, that means there are indeed $n-1$ linearly independent lattice vectors in the hyperplane $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. The only prerequisite that is needed for this conclusion is that $\boldsymbol{y}$ is rational. But it is possible that the sublattice is a lot *sparser*, that means the determinant of the sublattice might be a lot larger than $\det(\Lambda)$.

1.8 The Hermite Normal Form

The question that we want to answer here is, how one can compute a lattice basis for the intersection of a lattice Λ with a subspace. For this section, we follow Chapter 4 and 5 in [Sch99]. Let us keep the situation a bit more general and consider a matrix $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m) \in \mathbb{Q}^{n \times m}$ with $m \geq n$ and *full row rank*, that means $\operatorname{span}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m) = \mathbb{R}^n$. We also want to consider the lattice $\Lambda(\boldsymbol{B}) = \{\sum_{i=1}^m \lambda_i \boldsymbol{b}_i \mid \lambda_1, \ldots, \lambda_m \in \mathbb{Z}\}$. Note that now the vectors spanning the lattice are not necessarily

linearly independent. However, we will see that there is a lattice basis $\tilde{\boldsymbol{B}} \in \mathbb{Q}^{n\times n}$ so that $\Lambda(\tilde{\boldsymbol{B}}) = \Lambda(\boldsymbol{B})$.

The first question is, how can we change the matrix $\boldsymbol{B}$ without changing the generated lattice?

Definition 1.29. The following operations on a matrix $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m) \in \mathbb{R}^{n\times m}$ are called *unimodular column operations*:

a) Exchanging columns $\boldsymbol{b}_i$ and $\boldsymbol{b}_j$ for $i \neq j$.

b) Replacing $\boldsymbol{b}_i$ by $-\boldsymbol{b}_i$.

c) Replacing $\boldsymbol{b}_i$ by $\boldsymbol{b}_i + \alpha \cdot \boldsymbol{b}_j$ for $j \neq i$ and some $\alpha \in \mathbb{Z}$.

The following is easy to see:

Corollary 1.30. *Let $\boldsymbol{B} \in \mathbb{R}^{n\times m}$ be a regular matrix and let $\tilde{\boldsymbol{B}}$ be the matrix after any number of unimodular column operations. Then $\Lambda(\boldsymbol{B}) = \Lambda(\tilde{\boldsymbol{B}})$.*

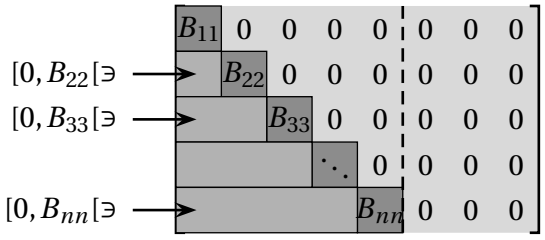
The question arises whether there is a structurally rich "normal form" that one can bring every lattice basis into. And indeed, this normal form exists:

Definition 1.31. Let $\boldsymbol{B} \in \mathbb{Q}^{n\times m}$ be a matrix. Then we say that $\boldsymbol{B}$ is in *Hermite normal form* if

i) One has $\boldsymbol{B} = (\boldsymbol{L}, \boldsymbol{0})$ where $\boldsymbol{L}$ is a lower triangular matrix

ii) $B_{ij} \geq 0$ for all $i, j \in [n]$

iii) Each diagonal entry B_{ii} is the unique maximum entry for that row i



The first observation is that as $\boldsymbol{B}$ is rational, we can scale $\boldsymbol{B}$ so that $\boldsymbol{B} \in \mathbb{Z}^{n\times m}$. Before we move on with the general case, let us discuss the special case that $\boldsymbol{B} \in \mathbb{Z}^{1\times m}$ as only *one row*. For this case, we know that the *Euclidean algorithm* can add and subtract entries until only one non-zero entry is left which can be moved

to the leftmost entry. Note that the non-zero entry will be precisely the greatest common divisor:

$$(B_{11},\dots,B_{1m}) \overset{\text{unimodular column operations}}{\mapsto} \big(\gcd(B_{11},\dots,B_{1m}),0,\dots,0\big)$$

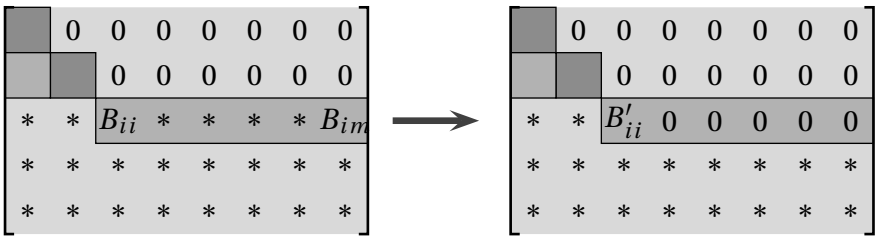
Note that the emerging row is indeed in Hermite normal form. We will now see an algorithm that is a generalization of the Euclidean algorithm to the matrix world:

Theorem 1.32. *There is an algorithm that takes any matrix $\boldsymbol{B} \in \mathbb{Z}^{n\times m}$ as input and performs poly$(n,m,\log\|\boldsymbol{B}\|_\infty)$ many unimodular row operations to obtain $\tilde{\boldsymbol{B}}$ in Hermite normal form.*

Proof. Let $\boldsymbol{B} \in \mathbb{Z}^{n\times m}$ be the input matrix. We will now apply unimodular column operations until $\boldsymbol{B}$ is in Hermite normal form:

(1) FOR $i = 1$ TO n DO

(2) Perform the Euclidean algorithm to entries $(B_{ii}, B_{i,i+1},\dots,B_{i,m})$ (actually to columns $i,\dots,m$) and obtain entries $(B'_{ii},0,\dots,0)$



(3) Add multiples of column i to columns $1,\dots,i-1$ until $0 \le B_{ij} < B_{ii}$ for all $j = 1,\dots,i-1$

We know that Euclid's algorithm takes polynomial time, hence each iteration of (2) takes a polynomial number of column operations. It is somewhat clear that the matrix that we get at the end will indeed be in Hermite normal form. □

The above argument shows that the number of iterations is bounded. But it does not immediately imply that the encoding length of any intermediate number is polynomial as well. In fact, it takes quite some effort. To get that result one has to do computations modulo M where M is the largest absolute value of any subdeterminant of $\boldsymbol{B}$. The details can be found in [Sch99].

Corollary 1.33. *For any integral matrix $\boldsymbol{A} \in \mathbb{Z}^{n\times m}$ one can compute a unimodular matrix $\boldsymbol{U} \in \mathbb{Z}^{m\times m}$ in time poly$(n,m,\log\|\boldsymbol{A}\|_\infty)$ so that $\boldsymbol{AU} = (\boldsymbol{B},\boldsymbol{0})$ is in Hermite normal form.*

One can also show the following:

Theorem 1.34. *The Hermite normal form of every matrix is unique.*

Again, for a proof, see [Sch99]. We call a subspace $U \subseteq \mathbb{R}^n$ *rational* if it has a basis whose vectors are from $\mathbb{Q}^n$ (or equivalently is $U = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = 0 \text{ for } i = 1, \ldots, k\}$ for $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k \in \mathbb{R}^n$).

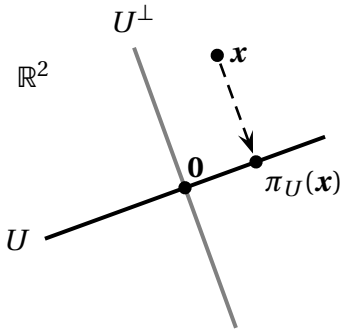
Lemma 1.35. *Let $\Lambda(\boldsymbol{B})$ be a full rank lattice with $\boldsymbol{B} \in \mathbb{Q}^{n \times n}$ and let $U \subseteq \mathbb{R}^n$ be a rational subspace. Then there is a polynomial time algorithm to compute a lattice basis $\boldsymbol{B}' \in \mathbb{Q}^{n \times \dim(U)}$ so that $\Lambda(\boldsymbol{B}') = \Lambda \cap U$.*

Proof. Let $k := \dim(U)$. After applying a (rational) linear transformation we may assume that $U = \operatorname{span}(\boldsymbol{e}_{n-k+1}, \ldots, \boldsymbol{e}_n)$. Let $\tilde{\boldsymbol{B}} = (\boldsymbol{b}_1, \ldots, \tilde{\boldsymbol{b}}_n)$ be the Hermite normal form of $\boldsymbol{B}$. Then the sublattice $\Lambda \cap U$ is spanned by the vectors $\tilde{\boldsymbol{b}}_{n-k+1}, \ldots, \tilde{\boldsymbol{b}}_n$. □

1.9 Korkine-Zolotarev reduced basis

We have seen in Section 1.5.3 that an LLL-reduced lattice basis $\boldsymbol{B}$ has an *orthogonality defect* of at most $2^{n^2/2}$. Moreover, using the LLL-algorithm one can compute such a basis in polynomial time. One might wonder whether an even better lattice basis can be obtained if we drop the requirement of a polynomial time algorithm. And indeed we will see that for any lattice Λ there is always a basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ so that the orthogonality defect $\gamma(\boldsymbol{B}) = \frac{\prod_{i=1}^n \|\boldsymbol{b}_i\|_2}{\det(\Lambda)}$ satisfies $\gamma(\boldsymbol{B}) \leq n^n$. To make this sound more impressive, consider a lattice, normalized so that $\det(\Lambda) = 1$. Then in an LLL-reduced basis, the geometric average length of a basis vector is $\gamma(\boldsymbol{B})^{1/n} \leq 2^{n/2}$. In contrast, the basis that we will discuss here will satisfy $\gamma(\boldsymbol{B})^{1/n} \leq n$. For more details on this topic we refer to the textbook of Micciancio and Goldwasser [MG02].

To fix some notation, for a subspace $U \subseteq \mathbb{R}^n$ we define $U^\perp := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \; \forall \boldsymbol{y} \in U\}$ as the *orthogonal complement*. We also write $\pi_U : \mathbb{R}^n \to U$ with $\pi_U(\boldsymbol{x}) := \operatorname{argmin}\{\|\boldsymbol{x} - \boldsymbol{y}\|_2 : \boldsymbol{y} \in U\}$ as the *orthogonal projection* of $\boldsymbol{x}$ into U. In particular for $\boldsymbol{y} \in U$ and $\boldsymbol{z} \in U^\perp$ we have $\pi_U(\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{y}$.



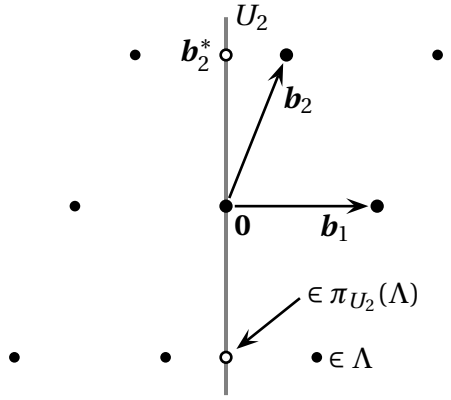
Again, for a lattice basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ we denote $(\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*)$ as the Gram-Schmidt orthogonalization with coefficients $\mu_{ij} := \frac{\langle \boldsymbol{b}_j, \boldsymbol{b}_i^* \rangle}{\|\boldsymbol{b}_i^*\|_2^2}$.

Definition 1.36. Let $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ be the basis for a full-rank lattice Λ. Define a subspace $U_i := \text{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}\}^\perp = \text{span}\{\boldsymbol{b}_i^*, \ldots, \boldsymbol{b}_n^*\}$. We call the basis $\boldsymbol{B}$ *Korkine-Zolotarev reduced* (KZ-reduced) if

(a) For all $i = 1, \ldots, n$, the vector $\boldsymbol{b}_i^*$ is a shortest lattice vector in $\pi_{U_i}(\Lambda)$.

(b) The basis $\boldsymbol{B}$ is coefficient-reduced.

To get some intuition, note that $U_1 = \mathbb{R}^n$ and $\boldsymbol{b}_1 = \boldsymbol{b}_1^*$ has to be the shortest lattice vector in Λ. Moreover, $U_2 = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{b}_1^* \rangle = 0\}$ and $\pi_{U_2}(\Lambda)$ is the $(n-1)$-dimensional lattice that is obtained by projecting Λ on U_2. Again, $\boldsymbol{b}_2^* = \pi_{U_2}(\boldsymbol{b}_2)$ is the shortest lattice vector in $\pi_{U_2}(\Lambda)$. For example, below is a KZ-reduced basis in $\mathbb{R}^2$:



It will also be convenient to remember that for all $i = 1, \ldots, n$ and $\boldsymbol{x} \in \mathbb{R}^n$ one has

$$\pi_{U_i}(\boldsymbol{x}) = \sum_{j=i}^{n} \langle \boldsymbol{x}, \boldsymbol{b}_j^* \rangle \cdot \frac{\boldsymbol{b}_j^*}{\|\boldsymbol{b}_j^*\|_2^2}$$

We know that any basis can be made coefficient-reduced in polynomial time without changing the Gram-Schmidt orthogonalization, hence (b) comes for free. On the other hand, finding a KZ-reduced must be at least **NP**-hard as it contains a shortest lattice vector as $\boldsymbol{b}_1$. Also, for an LLL-reduced basis we had the condition $\|\boldsymbol{b}^*_{i+1}\|_2^2 \geq \frac{1}{2}\|\boldsymbol{b}^*_i\|_2^2$ that we are not requiring for a KZ-reduced basis. Surprisingly, such a condition is implicitly satisfied anyway:

Lemma 1.37. *For a KZ-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ one has $\|\boldsymbol{b}^*_{i+1}\|_2^2 \geq \frac{3}{4}\|\boldsymbol{b}^*_i\|_2^2$ for all $i = 1, \dots, n-1$.*

Proof. First, we have

$$\pi_{U_i}(\boldsymbol{b}_{i+1}) = \sum_{j=i}^{n} \underbrace{\langle \boldsymbol{b}_{i+1}, \boldsymbol{b}^*_j \rangle}_{=0 \text{ if } j>i+1} \cdot \frac{\boldsymbol{b}^*_j}{\|\boldsymbol{b}^*_j\|_2^2} = \underbrace{\frac{\langle \boldsymbol{b}_{i+1}, \boldsymbol{b}^*_{i+1} \rangle}{\|\boldsymbol{b}^*_{i+1}\|_2^2}}_{=1} \boldsymbol{b}^*_{i+1} + \underbrace{\frac{\langle \boldsymbol{b}_{i+1}, \boldsymbol{b}^*_i \rangle}{\|\boldsymbol{b}^*_i\|_2^2}}_{=\mu_{i,i+1}} \boldsymbol{b}^*_i = \boldsymbol{b}^*_{i+1} + \mu_{i,i+1}\boldsymbol{b}^*_i.$$

Then we can estimate

$$\|\boldsymbol{b}^*_i\|_2^2 \overset{\boldsymbol{b}^*_i \text{ shortest vector}}{\leq} \|\pi_{U_i}(\boldsymbol{b}_{i+1})\|_2^2 \overset{\text{orthogonality}}{=} \|\boldsymbol{b}^*_{i+1}\|_2^2 + \underbrace{\mu^2_{i,i+1}}_{\leq 1/4}\|\boldsymbol{b}^*_i\|_2^2 \overset{\text{coef.-reduced}}{\leq} \|\boldsymbol{b}^*_{i+1}\|_2^2 + \frac{1}{4}\|\boldsymbol{b}^*_i\|_2^2,$$

using that $\boldsymbol{b}^*_i$ is a shortest vector in the lattice $\pi_{U_i}(\Lambda)$. Rearranging gives the claim. □

Next, we will prove that every lattice admits a KZ-reduced basis. In fact, we will provide an algorithm to find a KZ-reduced basis with n calls to a shortest vector oracle. By a slight abuse of notation, we will start with a given lattice Λ and then iteratively determine the Gram-Schmidt orthogonalization $\boldsymbol{b}^*_1, \dots, \boldsymbol{b}^*_n$ for the KZ-reduced basis in this order. Only after this, we will determine the vectors $\boldsymbol{b}_1, \dots, \boldsymbol{b}_n$.

Exponential time algorithm for KZ-reduction

Input: Full-rank lattice $\Lambda \subseteq \mathbb{R}^n$

Output: KZ-reduced basis $\boldsymbol{b}_1, \dots, \boldsymbol{b}_n$ for Λ

(1) FOR $i = 1$ TO n DO
  (2) Define $U_i := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{b}^*_j \rangle = 0 \ \forall j = 1, \dots, i-1\}$.
  (3) Compute the shortest vector in $\pi_{U_i}(\Lambda) \setminus \{\boldsymbol{0}\}$ and call it $\boldsymbol{b}^*_i$.
(4) Given the Gram-Schmidt orthogonalization $(\boldsymbol{b}^*_1, \dots, \boldsymbol{b}^*_n)$, recover a basis $\boldsymbol{B} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ that has this GS orthogonalization and satisfies $-\frac{1}{2} \leq \mu_{ij} \leq \frac{1}{2}$ for all $1 \leq i < j \leq n$ in polynomial time.

Lemma 1.38. *The algorithm computes a KZ-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ for the lattice Λ.*

Proof. It suffices to show how to extend $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$ to any basis of Λ in step (4) — then the procedure in Lemma 1.15 can make the basis coefficient-reduced in polynomial time. For $i \in \{1, \ldots, n\}$, let $\boldsymbol{b}_i \in \Lambda$ be an arbitrary vector with $\pi_{U_i}(\boldsymbol{b}_i) = \boldsymbol{b}_i^*$, which exists by choice of the $\boldsymbol{b}_i^*$. First, note that this will give n linearly independent lattice vectors and the Gram-Schmidt orthogonalization is clearly $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$. It remains to show that $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ form indeed a lattice basis of Λ. To see this, take any lattice vector $\boldsymbol{x} \in \Lambda$ and write it as $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{\lambda}$. We have to argue that $\boldsymbol{\lambda} \in \mathbb{Z}^n$. Suppose this is not the case and for some index i one has $\lambda_i \notin \mathbb{Z}$ but $\lambda_{i+1}, \ldots, \lambda_n \in \mathbb{Z}$. Without loss of generality suppose that $\lambda_i > 0$. Then also $\boldsymbol{x}' := \boldsymbol{x} - \sum_{j=i+1}^{n} \lambda_j \boldsymbol{b}_j - \lfloor \lambda_i \rfloor \boldsymbol{b}_i$ is a lattice vector and if we write $\boldsymbol{x}' = \boldsymbol{B}\boldsymbol{\lambda}'$, then $0 < \lambda_i' < 1$ and $\lambda_{i+1}' = \ldots = \lambda_n' = 0$. Considering the projection of $\boldsymbol{x}'$ we observe that

$$\pi_{U_i}(\boldsymbol{x}') = \underbrace{\pi_{U_i}\Big(\sum_{j=1}^{i-1} \lambda_j \boldsymbol{b}_j\Big)}_{=\boldsymbol{0}} + \lambda_i \underbrace{\pi_{U_i}(\boldsymbol{b}_i)}_{=\boldsymbol{b}_i^*} = \lambda_i \boldsymbol{b}_i^*$$

is non-zero lattice vector in $\pi_{U_i}(\Lambda)$ that is strictly shorter than $\boldsymbol{b}_i^*$ — this is a contradiction to the choice of $\boldsymbol{b}_i^*$. The claim then follows. □

In Chapter 2 we will discuss an algorithm that finds a shortest vector (with respect to the $\|\cdot\|_2$-norm) in time $2^{O(n)}$. This then implies the following:

Corollary 1.39. For a lattice $\Lambda := \Lambda(\tilde{\boldsymbol{B}})$ one can compute a KZ-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ in time $2^{O(n)}$ times a polynomial in the encoding length of $\tilde{\boldsymbol{B}}$.

Recall that for a lattice Λ, the *ith successive minimum* is the smallest value λ_i so that $\Lambda \cap \lambda_i B_2^n$ contains at least i linearly independent lattice vectors. The crucial insight is that the length of the basis vectors can be bounded be the corresponding successive minima:

Lemma 1.40. *For a KZ-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ one has $\|\boldsymbol{b}_i\|_2 \leq \sqrt{n} \cdot \lambda_i(\Lambda)$ for all $i = 1, \ldots, n$.*

Proof. First, condition (a) implies that in KZ-reduced basis, each Gram-Schmidt vector has a length that is bounded by the corresponding successive minimum:

Claim. *One has $\|\boldsymbol{b}_i^*\|_2 \leq \lambda_i$ for all $i = 1, \ldots, n$.*
Proof. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \Lambda \setminus \{\boldsymbol{0}\}$ be the linearly independent lattice vectors with

$\|\boldsymbol{v}_i\|_2 = \lambda_i$ for $i = 1,\dots,n$. Since π_{U_i} is a projection on a $\dim(U_i) = n-i+1$ dimensional space, there must be at least one index $j \in \{1,\dots,i\}$ so that $\pi_{U_i}(\boldsymbol{v}_j) \neq \boldsymbol{0}$. Since $\boldsymbol{b}_i^*$ will be the shortest non-zero lattice vector in $\pi_{U_i}(\Lambda)$, we conclude that

$$\|\boldsymbol{b}_i^*\|_2 \overset{\boldsymbol{b}_i^* \text{ shortest vector}}{\leq} \|\pi_{U_i}(\boldsymbol{v}_j)\|_2 \overset{\text{projection}}{\leq} \|\boldsymbol{v}_j\|_2 \overset{\lambda_1 \leq \dots \leq \lambda_n}{\leq} \|\boldsymbol{v}_i\|_2$$

and the claim follows. □

Since the basis $\boldsymbol{B}$ is coefficient-reduced, we can now bound the length of the basis vectors by estimating

$$\|\boldsymbol{b}_i\|_2^2 = \left\|\boldsymbol{b}_i^* + \sum_{j=1}^{i-1} \mu_{ji}\boldsymbol{b}_j^*\right\|_2^2 = \underbrace{\|\boldsymbol{b}_i^*\|_2^2}_{\leq \lambda_i^2} + \sum_{j=1}^{i-1} \underbrace{\mu_{ji}^2}_{\leq 1/4} \cdot \underbrace{\|\boldsymbol{b}_j^*\|_2^2}_{\leq \lambda_j^2 \leq \lambda_i^2} \leq n \cdot \lambda_i^2.$$

Taking square roots gives then $\|\boldsymbol{b}_i\|_2 \leq \sqrt{n} \cdot \lambda_i$. □

Now, we can easily show the main theorem for this section:

Theorem 1.41. *The orthogonality defect of a KZ-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1,\dots,\boldsymbol{b}_n)$ is $\gamma(\boldsymbol{B}) \leq n^n$.*

Proof. We know from Lemma 1.40 that the length of each basis vector is bounded by the corresponding successive minimum and from Minkowski's Second Theorem (see Section 1.4) we have a general upper bound on the product of those:

$$\prod_{i=1}^{n} \|\boldsymbol{b}_i\|_2 \overset{\text{Lem. 1.40}}{\leq} n^{n/2} \cdot \prod_{i=1}^{n} \lambda_i(\Lambda) \overset{\text{Mink. 2nd Thm}}{\leq} n^{n/2} \cdot n^{n/2} \cdot \det(\Lambda).$$

Then dividing both sides by $\det(\Lambda)$ gives $\gamma(\boldsymbol{B}) \leq n^n$. □

We can also show that the orthogonality defect of a KZ-basis is approximately optimal:

Lemma 1.42. *There is a lattice $\Lambda \subseteq \mathbb{R}^n$ for which any basis $\boldsymbol{B}$ has an orthogonality defect of $\gamma(\boldsymbol{B}) \geq (\frac{\sqrt{n}}{12})^n$.*

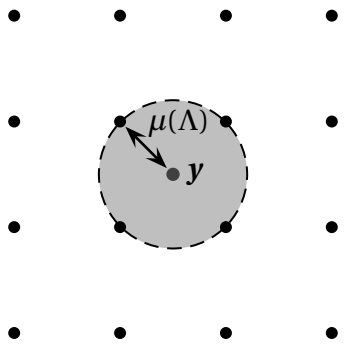
Proof. We will see in Lemma 1.46 that there is a lattice Λ with $\det(\Lambda) = 1$ and $\lambda_1(\Lambda) \geq \frac{\sqrt{n}}{12}$. Let $\boldsymbol{B} = (\boldsymbol{b}_1,\dots,\boldsymbol{b}_n)$ be any lattice basis for Λ. Then $\gamma(\boldsymbol{B}) = \frac{\prod_{i=1}^n \|\boldsymbol{b}_i\|_2}{\det(\Lambda)} \geq \lambda_1(\Lambda)^n \geq (\frac{\sqrt{n}}{12})^n$. □

1.10 The covering radius and a tight lattice for Minkowski's Theorem

We define the *covering radius* of a lattice Λ as the quantity

$$\mu(\Lambda) := \max_{\boldsymbol{y}\in\mathbb{R}^n} \min_{\boldsymbol{x}\in\Lambda} \|\boldsymbol{y}-\boldsymbol{x}\|_2$$

In words, the covering radius $\mu(\Lambda)$ gives the furthest distance of any point $\boldsymbol{y}\in\mathbb{R}^n$ to the lattice Λ.



It is not surprising that one can prove some relation to the density of a lattice.

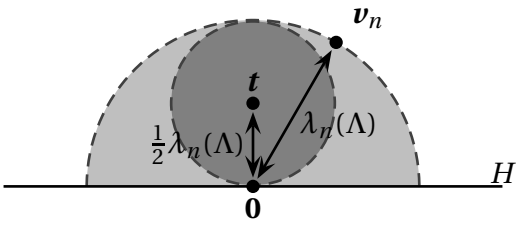
Lemma 1.43. *For any full rank lattice Λ in $\mathbb{R}^n$ one has $\mu(\Lambda) \geq \frac{\sqrt{n}}{6}\cdot \det(\Lambda)^{1/n}$.*

Proof. After scaling we may assume that $\det(\Lambda)=1$. As we know from Lemma 1.1 we have $\mathrm{Vol}_n(\frac{\sqrt{n}}{6}B_2^n) \leq (\frac{2e}{6})^n < 1$. Then the translates $\Lambda + \frac{\sqrt{n}}{6}B_2^n$ do not cover the whole $\mathbb{R}^n$. Any uncovered point $\boldsymbol{y}\in\mathbb{R}^n \setminus (\Lambda + \frac{\sqrt{n}}{6}B_2^n)$ has a distance of more than $\frac{\sqrt{n}}{6}$ to the lattice. □

Another useful lower bound for the covering radius is in terms of the n-th successive minimum.

Lemma 1.44. *For any full rank lattice $\Lambda\subseteq\mathbb{R}^n$, one has $\mu(\Lambda)\geq\frac{\lambda_n(\Lambda)}{2}$.*

Proof. Let $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\in\Lambda\setminus\{\boldsymbol{0}\}$ be the linearly independent vectors with $\lambda_i(\Lambda) = \|\boldsymbol{v}_i\|_2$. Let us abbreviate $H := \mathrm{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{n-1}\}$ as the $(n-1)$-dimensional hyperplane and in wise foresight, let us denote the normal vector of that hyperplane by $\boldsymbol{t}$, that means $H=\{\boldsymbol{x}\in\mathbb{R}^n \mid \langle\boldsymbol{x},\boldsymbol{t}\rangle = 0\}$. Let us also scale $\boldsymbol{t}$ so that $\|\boldsymbol{t}\|_2 = \frac{1}{2}\|\boldsymbol{v}_n\|_2$.

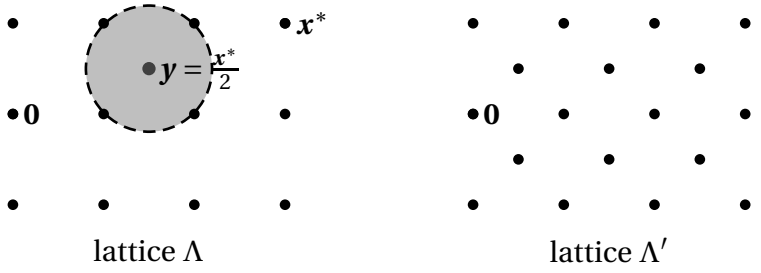


By definition we know that $\text{int}(\lambda_n(\Lambda) \cdot B_2^n \setminus H)$ does not contain a lattice point. Moreover $\boldsymbol{t} + \frac{\lambda_n(\Lambda)}{2} B_2^n \subseteq \lambda_n(\Lambda) \cdot B_2^n \setminus H$. Hence the ball $\boldsymbol{t} + \frac{\lambda_n(\Lambda)}{2} B_2^n$ does not contain a lattice point in its interior and so the covering radius is $\mu(\Lambda) \geq \frac{\lambda_n(\Lambda)}{2}$. □

We want to use the notion of covering radius to construct a lattice where Minkowski's First Theorem is tight up to constants (and then immediately Minkowski's 2nd Theorem is tight as well). Interestingly, the construction is iterative. Key argument is to show that in any lattice Λ with $\mu(\Lambda) > 2\lambda_1(\Lambda)$, we can double the density of the lattice without decreasing $\lambda_1(\Lambda)$.

Lemma 1.45. *For any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ one can construct another lattice $\Lambda' \subseteq \mathbb{R}^n$ with $\det(\Lambda') = \frac{1}{2}\det(\Lambda)$ and $\lambda_1(\Lambda') \geq \min\{\lambda_1(\Lambda), \frac{\mu(\Lambda)}{2}\}$.*

Proof. Let $\boldsymbol{y} \in \mathbb{R}^n$ be a point attaining the covering radius, i.e. $d(\boldsymbol{y}, \Lambda) = \mu(\Lambda)$. Now, let $\boldsymbol{x}^* \in \Lambda$ be a lattice point minimizing $\|2\boldsymbol{y} - \boldsymbol{x}^*\|_2$. By the definition of covering radius we have $\|\boldsymbol{y} - \frac{\boldsymbol{x}^*}{2}\|_2 = \frac{1}{2}\|2\boldsymbol{y} - \boldsymbol{x}^*\|_2 \leq \frac{\mu(\Lambda)}{2}$. That means $\frac{\boldsymbol{x}^*}{2}$ is not a lattice point. Hence $\Lambda' := \Lambda + (\Lambda + \frac{\boldsymbol{x}^*}{2})$ is a lattice with twice the density of Λ, meaning that $\det(\Lambda') = \frac{1}{2}\det(\Lambda)$. A visualization in as follows (though it seems that in $\mathbb{R}^2$ one has $\boldsymbol{y} = \frac{\boldsymbol{x}^*}{2}$ which may not necessarily be true in higher dimensions):



It remains to prove that the Λ' does not contain short vectors. For the "old" lattice vectors $\boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}$ we still have $\|\boldsymbol{x}\|_2 \geq \lambda_1(\Lambda)$. So consider a "new" vector of the form $\boldsymbol{x} - \frac{\boldsymbol{x}^*}{2}$ with $\boldsymbol{x} \in \Lambda$. Then by the reverse triangle inequality we have

$$\left\|\boldsymbol{x} - \frac{\boldsymbol{x}^*}{2}\right\|_2 \geq \underbrace{\|\boldsymbol{x} - \boldsymbol{y}\|_2}_{\geq \mu(\Lambda)} - \underbrace{\left\|\frac{\boldsymbol{x}^*}{2} - \boldsymbol{y}\right\|_2}_{\leq \mu(\Lambda)/2} \geq \frac{\mu(\Lambda)}{2}$$

Either way, the claim holds true. □

Lemma 1.46. *There exists a full rank lattice $\Lambda \subseteq \mathbb{R}^n$ with $\det(\Lambda) = 1$ and $\lambda_1(\Lambda) \geq \frac{\sqrt{n}}{12}$.*

Proof. We start with an arbitrary lattice, say $\Lambda_0 := \mathbb{Z}^n$. Then given a lattice Λ_t we use Lemma 1.45 to construct Λ_{t+1} with $\det(\Lambda_{t+1}) = \frac{1}{2}\det(\Lambda_t)$ and $\lambda_1(\Lambda_{t+1}) \geq \min\{\lambda_1(\Lambda_t), \frac{\mu(\Lambda_t)}{2}\}$. Consider the first iteration t where the length of the shortest vector strictly increases; in that case

$$\lambda_1(\Lambda_t) > \frac{\mu(\Lambda_t)}{2} \overset{\text{Lem 1.43}}{\geq} \frac{\sqrt{n}}{12}\det(\Lambda_t)^{1/n}$$

Then the scaled lattice $\Lambda := \frac{\Lambda_t}{\det(\Lambda_t)^{1/n}}$ satisfies the claim. □

We cite a stronger result without proof (see e.g. the textbook of [MH73]):

Theorem 1.47 (Conway, Thompson)**.** *There exists a lattice $\Lambda \subseteq \mathbb{R}^n$ that is self-dual (i.e. $\Lambda^* = \Lambda$) with $\lambda_1(\Lambda) = \lambda_1(\Lambda^*) \geq \Omega(\sqrt{n})$.*

1.11 Exercises

Exercise 1.1.
Let $\Lambda = \Lambda(\boldsymbol{B})$ with $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ be a lattice. Show that for any $\varepsilon > 0$ there is a radius $R := R(\varepsilon, n, \boldsymbol{B})$ so that

$$(1-\varepsilon)\cdot\frac{\mathrm{Vol}_n(R\cdot B_2^n)}{\det(\Lambda)} \leq |RB_2^n \cap \Lambda| \leq (1+\varepsilon)\cdot\frac{\mathrm{Vol}_n(R\cdot B_2^n)}{\det(\Lambda)}$$

Exercise 1.2.
Solve the following:

a) Let $K \subseteq \mathbb{R}^n$ be a symmetric convex set with $\mathrm{Vol}_n(K) > k\cdot 2^n$ for some $k \in \mathbb{N}$. Show that $|K \cap \mathbb{Z}^n| \geq k+1$.

b) Is the following claim true: For any $k \in \{1,\ldots,n\}$ there is a value $f(k)$ so that for any symmetric convex body K with $\mathrm{Vol}_n(K) > f(k)\cdot 2^n$, $K \cap \mathbb{Z}^n$ contains k linearly independent vectors.

Exercise 1.3.

This is an application of Dirichlet's Theorem: Let $\boldsymbol{a} \in (0,1]^n$ be a real vector and consider the hyperplane $H := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x} \rangle = 0\}$. Then there is a rational vector $\tilde{\boldsymbol{a}} \in \frac{\mathbb{Z}^n}{q}$ with $q \leq (2nR)^n$ so that $\tilde{H} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \tilde{\boldsymbol{a}}, \boldsymbol{x} \rangle = 0\}$ satisfies the following:

$$\forall \boldsymbol{x} \in \{-R, \ldots, R\}^n : \big(\boldsymbol{x} \in H \Rightarrow \boldsymbol{x} \in \tilde{H}\big).$$

Remark. You don't have to prove it but the same argument should also show that for all $\boldsymbol{x} \in \{-R, \ldots, R\}^n$ one has $\boldsymbol{x} \in H_{\leq} \Rightarrow \boldsymbol{x} \in \tilde{H}_{\leq}$ where $H_{\leq} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x} \rangle \leq 0\}$.

Exercise 1.4.
Let $S \subseteq \mathbb{R}^n$ be a measurable, compact set with $\mathrm{Vol}_n(S) > k$ for some $k \in \mathbb{Z}_{\geq 0}$. Then there are points $\boldsymbol{s}_0, \ldots, \boldsymbol{s}_k \in S$ with $\boldsymbol{s}_i - \boldsymbol{s}_j \in \mathbb{Z}^n \setminus \{\boldsymbol{0}\}$ for all $i \neq j$.

Exercise 1.5.
Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice. Show that $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \geq 1$, where Λ^* is the dual lattice.

Exercise 1.6.
Prove that for any lattice $\Lambda \subseteq \mathbb{R}^n$, one has $\lambda_1(\Lambda) \cdot \lambda_1(\Lambda^*) \leq n$.
Remark: A stronger theorem of Banaszczyk [Ban93a] that we will see in Chapter 4 shows that even $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \leq n$. This has an important consequence. Consider the following computational problem: Given a lattice Λ and a parameter K, distinguish the cases $\lambda_1(\Lambda) \leq L$ and $\lambda_1(\Lambda) > n \cdot L$. The consequence of this exercise is that this problem is in $\mathbf{NP} \cap \mathbf{coNP}$ in the sense that one can give an efficiently checkable proof for $\lambda_1(\Lambda) \leq L$ (simply give me a short vector) and one can also certify is $\lambda_1(\Lambda) > n \cdot L$ (give me the short dual basis). The remarkable fact is that this gap problem is not known to be in $\boldsymbol{P}$.

Exercise 1.7.
Consider the matrix

$$\boldsymbol{B} = \begin{pmatrix} 2 & 3 & 4 \\ 2 & 4 & 6 \end{pmatrix}$$

Compute the Hermite Normal form of $\boldsymbol{B}$.

Exercise 1.8.
Let $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{Z}^n$ be a vector of integer numbers. The original *Euclidean algorithm* does the following:

(1) REPEAT

  (2) Select the index i with $|a_i|$ minimal.
  (3) For all $j \neq i$ replace a_j by $\min\{|a_j + z \cdot a_i| \mid z \in \mathbb{Z}\}$.

Prove that the algorithm terminates after at most $O(\log \|\boldsymbol{a}\|_\infty)$ many iterations.

Exercise 1.9.
Let $\boldsymbol{A} \in \mathbb{Z}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{Z}^m$ with $m \leq n$ where $\boldsymbol{A}$ has full row rank. Show that in polynomial time one can compute a vector $\boldsymbol{x} \in \mathbb{Z}^n$ with $\boldsymbol{Ax} = \boldsymbol{b}$ (or decide that no such vector exists).
Remark: Use Cor. 1.33.

Exercise 1.10.
Let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice. Assume that $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n \in \Lambda$ are linearly-independent and minimize $|\det(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)|$. Prove that $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are indeed a *basis* of Λ.

Exercise 1.11.
We want to consider a relaxed version of a KZ-reduced basis. We say that a basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) \in \mathbb{R}^{n \times n}$ for a lattice Λ is *α-KZ-reduced* for $\alpha \geq 1$ if $\boldsymbol{B}$ is coefficient reduced and $\|\boldsymbol{b}_i^*\|_2 \leq \alpha \cdot \lambda_1(\pi_{U_i}(\Lambda))$ for all $i = 1, \ldots, n$. Here π_{U_i} is again the projection into $U_i := \text{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}\}^\perp$. Show that the orthogonality defect of such a basis is $\gamma(\boldsymbol{B}) \leq (\alpha n)^n$.

Exercise 1.12.
Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ be an LLL-reduced basis for Λ. Prove that for all $i \in \{1, \ldots, n\}$ one has $\|\boldsymbol{b}_i\|_2 \leq 2^{(n+1)/2} \lambda_i(\Lambda)$.
Hint. You may use following observation: Consider an LLL-reduced $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ and for some index $i \in \{1, \ldots, n\}$, define the subspace $U := \text{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}\}$ and let $\tilde{\boldsymbol{b}}_j := \Pi_{U^\perp}(\boldsymbol{b}_j)$ where $\Pi_{U^\perp}$ denotes the projection into the subspace $U^\perp$. Then $\tilde{\boldsymbol{b}}_i, \ldots, \tilde{\boldsymbol{b}}_n$ is an LLL-reduced basis for the lattice $\tilde{\Lambda} := \{\sum_{j=i}^n y_j \tilde{\boldsymbol{b}}_j : y_j \in \mathbb{Z}\}$.

Exercise 1.13.
Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice. Then for any k-dimensional sublattice $\tilde{\Lambda} \subseteq \Lambda$ one has $\det(\tilde{\Lambda}) \geq (\frac{\lambda_1(\Lambda)}{\sqrt{k}})^k$.

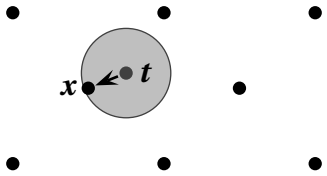
Chapter 2

The Closest Vector Problem

In this chapter, we will study a new problem:

> CLOSEST VECTOR PROBLEM (CVP)
> **Input:** A full rank lattice $\Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$ given by a regular matrix $\boldsymbol{B} \in \mathbb{Q}^{n\times n}$ and a target vector $\boldsymbol{t} \in \mathbb{R}^n$.
> **Goal:** Find the lattice vector $\boldsymbol{x} \in \Lambda(\boldsymbol{B})$ minimizing $\|\boldsymbol{x} - \boldsymbol{t}\|_2$.

A small example is as follows:



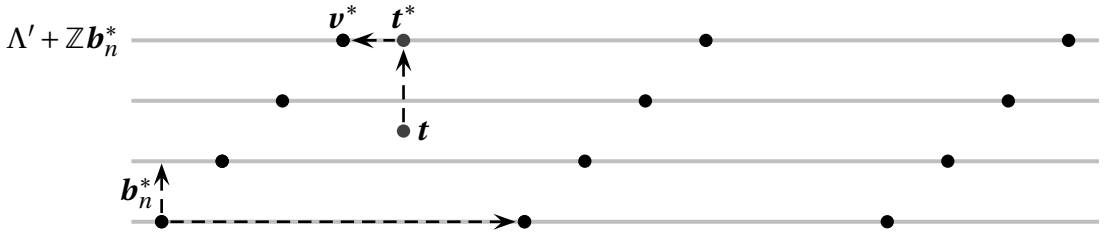
We will use the notation $\text{CVP}(\Lambda(\boldsymbol{B}), \boldsymbol{t})$ to denote the value of the optimum solution. One might imagine CVP as a more general version of the Shortest Vector problem (in the exercises, we will justify this claim). The main contents of this chapter will be the Nearest Plane algorithm by Babai and the Voronoi cell algorithm by Micciancio[1] and Voulgaris [MV10, MV13].

2.1 A $2^{O(n^2)}$-algorithm for Closest Vector

To warm up, we want to describe a simple algorithm that solves CVP in time $2^{O(n^2)}$. So, let $\boldsymbol{B}$ be the given lattice basis. In a first step, we replace $\boldsymbol{B}$ with an *LLL-reduced basis*, which only takes polynomial time, see Chapter 1. Suppose that

[1] The following link contains slides of Micciancio on the algorithm: https://cseweb.ucsd.edu/~daniele/papers/Voronoi-slides.pdf

$(\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*)$ is the *Gram-Schmidt orthogonalization* of that basis. That sequence of vectors has many properties, for example $\operatorname{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_i\} = \operatorname{span}\{\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_i^*\}$ for all i. Moreover, as $\boldsymbol{B}$ is LLL-reduced we know that $\|\boldsymbol{b}_1\|_2^2 = \|\boldsymbol{b}_1^*\|_2^2 \leq 2 \cdot \|\boldsymbol{b}_2^*\|_2^2 \leq 4\|\boldsymbol{b}_3^*\|_2^2 \leq \ldots \leq 2^{n-1}\|\boldsymbol{b}_n^*\|_2^2$. Now, consider the target vector $\boldsymbol{t}$ and let $\Lambda' = \{\sum_{i=1}^{n-1} \lambda_i \boldsymbol{b}_i : \lambda_i \in \mathbb{Z}\ \forall i = 1, \ldots, n-1\}$ be the $(n-1)$-dimensional sublattice formed by the first $n-1$ vectors. By the properties of the Gram-Schmidt orthogonalization, the whole lattice can be covered by affine subspaces of the form $\operatorname{span}(\Lambda') + k\boldsymbol{b}_n^*$ for $k \in \mathbb{Z}$. Let us call each such subspace a *layer*. If we want to find the closest lattice point $\boldsymbol{v}^* \in \Lambda$ to $\boldsymbol{t}$, then this can be done as follows: guess the right layer containing $\boldsymbol{v}^*$ and compute the orthogonal projection $\boldsymbol{t}^*$ of $\boldsymbol{t}$ on that layer. Then compute the closest vector to $\boldsymbol{t}^*$ with respect to the lower-dimensional lattice Λ'.



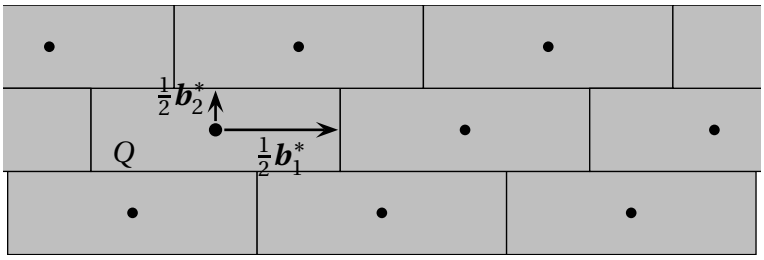
To make this recursive computation at least mildly efficient, we need to be able to bound the number of layers that we need to try out. In particular, $\boldsymbol{v}^*$ does not need to lie on the closest layer (as we can see in the picture above). But here is the crucial insight (still in the notation from above):

Lemma 2.1. *Let $\boldsymbol{t} \in \mathbb{R}^n$ be any target vector. Then the closest lattice point $\boldsymbol{v}^*$ lies on one of the at $2 \cdot 2^n$ layers that are closest to $\boldsymbol{t}$.*

This claim follows immediately from the following:

Lemma 2.2. *Let $\Lambda(\boldsymbol{B})$ be any lattice with LLL-reduced basis $\boldsymbol{B}$ and Gram-Schmidt orthogonalization $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$. For any target vector $\boldsymbol{t}$, one has $\textsc{Cvp}(\Lambda(\boldsymbol{B}), \boldsymbol{t}) \leq 2^n \|\boldsymbol{b}_n^*\|_2$.*

Proof. Let $Q := \{\sum_{i=1}^n \lambda_i \boldsymbol{b}_i^* : |\lambda_i| \leq \frac{1}{2}\ \forall i = 1, \ldots, n\}$ be a parallelepiped with center $\mathbf{0}$ whose sides are spanned by the Gram-Schmidt orthogonalization. Geometrically speaking, Q is a shifted and translated version of the fundamental parallelepiped of the basis $\boldsymbol{B}$. In particular, Q and $\mathcal{P}(\boldsymbol{B})$ have the same volume and both have the property that translates of them by lattice vectors exactly partition $\mathbb{R}^n$ (apart from a zero set).



Hence for each point $\boldsymbol{t} \in \mathbb{R}^n$, there is a lattice point $\boldsymbol{v} \in \Lambda(\boldsymbol{B})$ so that $\boldsymbol{t} \in \boldsymbol{v}+Q$. This point satisfies $\|\boldsymbol{t}-\boldsymbol{v}\|_2^2 \le \sum_{i=1}^n \|\frac{1}{2}\boldsymbol{b}_i^*\|_2^2 \le \sum_{i=1}^n 2^{n-i}\|\boldsymbol{b}_n^*\|_2^2 \le 2^n\|\boldsymbol{b}_n^*\|_2^2$. □

Theorem 2.3. *The Closest vector problem can be solved in time* $2^{O(n^2)}$.

Proof. Let $T(n)$ be the running time to solve CVP in dimension n. Then we just discussed that $T(n) \le 2 \cdot 2^n \cdot T(n-1)$. Applying induction then gives the claim. □

Obviously this is just a simple and naive algorithm. The induction can be done in a smarter way, see the $n^{O(n)} \cdot \text{poly(input)}$ algorithm by Kannan [Kan87a]. But one might already see one potential for improvement: in each of the $2 \cdot 2^n$ recursions we are solving the closest vector problem for the *same* lattice — just each time with different target vectors. If we could come up with some kind of preprocessing for the lattice, then we might reuse those computations in each of the recursions and speed up the algorithm.

An even simpler $2^{O(n^2)}$-time algorithm works for the Shortest Vector problem by computing an LLL-reduced basis $\boldsymbol{B}$ for a lattice and trying out all lattices points of the form $\boldsymbol{By}$ with $\|\boldsymbol{y}\|_\infty \le 2^{3n}$.

Theorem 2.4. *Let* $\Lambda \subseteq \mathbb{R}^n$ *be a full-rank lattice with LLL-reduced basis* $\boldsymbol{B} \in \mathbb{R}^{n \times n}$. *Then all shortest lattice vectors (with respect to the* $\|\cdot\|_2$*-norm) are contained in* $S := \{\boldsymbol{By} \mid \boldsymbol{y} \in \mathbb{Z}^n \text{ and } \|\boldsymbol{y}\|_\infty \le 2^{3n}\}$.

We will develop the proof in Exercise 2.1.

2.2 Babai's nearest plane algorithm

In this section we want to describe the so called *Nearest plane algorithm* due to Babai [Bab86] which for any lattice $\Lambda \subseteq \mathbb{R}^n$ and target vector $\boldsymbol{t} \in \mathbb{R}^n$ finds a vector $\boldsymbol{x} \in \Lambda$ with $\|\boldsymbol{x}-\boldsymbol{t}\|_2 \le 2^{n/2} \cdot \text{CVP}(\Lambda, \boldsymbol{t})$ in polynomial time. The key ingredient is (of course) to rely on an LLL-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ for the lattice Λ. Again we will use in particular the properties of the Gram Schmidt orthogonalization $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$ of that basis.

The input for the algorithm is a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n \in \mathbb{R}^m$ for the lattice and a target vector $\boldsymbol{t}$ in the span of those vectors. Due to the recursion, in later iterations the lattice will not have full rank even if that was the case for the original lattice. For a subspace $U \subseteq \mathbb{R}^m$, we will denote $\pi_U(\boldsymbol{t})$ as the *orthogonal projection* on U. For the algorithm itself, the basis does not need to be LLL-reduced, but parts of the analysis will rely on it.

NEARESTPLANEALGO
Input: Basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ for a lattice Λ; target vector $\boldsymbol{t} \in \operatorname{span}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$
Output: Lattice vector $\boldsymbol{x} \in \Lambda$

(1) Set $U := \operatorname{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}\}$

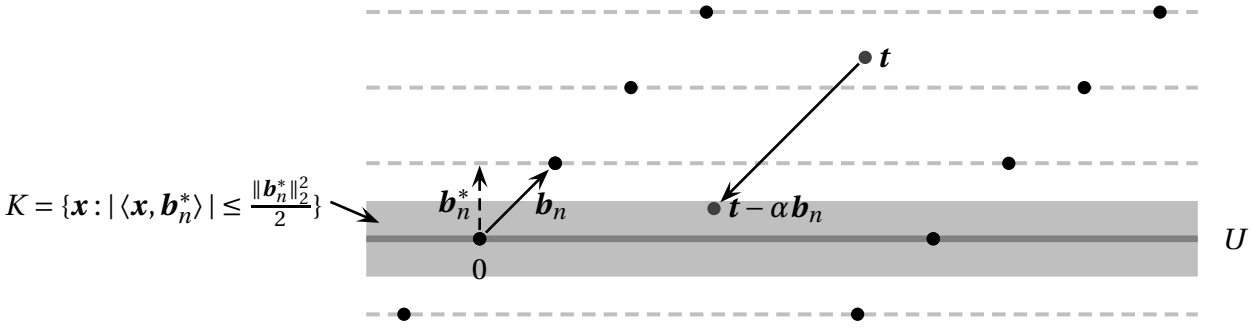
(2) Let $K := \left\{\boldsymbol{x} \in \mathbb{R}^m \mid |\langle \boldsymbol{x}, \frac{\boldsymbol{b}_n^*}{\|\boldsymbol{b}_n^*\|_2}\rangle| \leq \frac{1}{2}\|\boldsymbol{b}_n^*\|_2\right\}$

(3) Compute $\alpha \in \mathbb{Z}$ so that $\boldsymbol{t} - \alpha \boldsymbol{b}_n \in K$

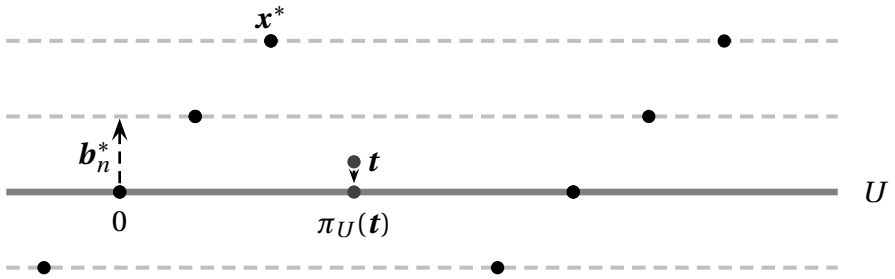
(4) If $n = 1$ then return $\alpha \boldsymbol{b}_n$

(5) Return NEARESTPLANEALGO$((\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}), \pi_U(\boldsymbol{t} - \alpha \boldsymbol{b}_n)) + \alpha \boldsymbol{b}_n$

To get some intuition about the algorithm consider the subspace $U := \operatorname{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}\}$ and note that Λ is contained in the translates of that subspace that are of the form $\alpha \boldsymbol{b}_n^* + U$ with $\alpha \in \mathbb{Z}$. The set K are all the points for which the closest translate is the one going through the origin. Then the algorithm translates $\boldsymbol{t}$ so that $t - \alpha \boldsymbol{b}_n$ is closeest to U, then one computes the projection onto U and recurses.



Replacing $\boldsymbol{t}$ by $\boldsymbol{t} - \alpha \boldsymbol{b}_n$ does not change the approximation error, hence for the analysis we will be able to assume that $\alpha = 0$ to simplify the notation. Note that $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}$ is again an LLL-reduced basis of the sublattice $\Lambda' := \{\sum_{i=1}^{n-1} y_i \boldsymbol{b}_i : \boldsymbol{y} \in \mathbb{Z}^{n-1}\} = \Lambda \cap U$.



First, we prove an *absolute* bound on the distance of t to the produced vector — this itself does not yet imply a bound relative to the optimum value $\text{CVP}(\Lambda, t)$ but it will be useful later.

Lemma 2.5. *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice with basis $B = (b_1, \ldots, b_n)$ and let $t \in \text{span}(B)$. Then $x := \text{NEARESTPLANEALGO}(B, t)$ satisfies $\|x - t\|_2^2 \leq \frac{1}{4}\sum_{i=1}^n \|b_i^*\|_2^2$. Moreover, if B is LLL-reduced then $\|x - t\|_2^2 \leq 2^{n-2}\|b_n^*\|_2^2$.*

Proof. As discussed above we may assume for simplicity that $\alpha = 0$ which means that $t \in K$. Then

$$\begin{aligned} \|x - t\|_2^2 &\overset{\text{orthogonality}}{=} \|t - \pi_U(t)\|_2^2 + \|x - \pi_U(t)\|_2^2 \\ &\overset{(*)}{\leq} \underbrace{\langle t, \frac{b_n^*}{\|b_n^*\|_2}\rangle^2}_{\leq \|\frac{1}{2}b_n^*\|_2^2} + \frac{1}{4}\sum_{i=1}^{n-1}\|b_i^*\|_2^2 \leq \frac{1}{4}\sum_{i=1}^{n}\|b_i^*\|_2^2 \end{aligned} \tag{2.1}$$

In $(*)$ we use $\|t - \pi_U(t)\|_2^2 = \langle t, \frac{b_n^*}{\|b_n^*\|_2}\rangle^2$ on the left hand side and we apply induction on the right hand side with the target vector $\pi_U(t) \in \text{span}\{b_1, \ldots, b_{n-1}\}$.

For the "moreover part" we may assume that the basis is LLL-reduced and so $\|b_i^*\|_2^2 \leq 2\|b_{i+1}^*\|_2^2$ which means that $\|b_i^*\|_2^2 \leq 2^{n-i}\|b_n^*\|_2^2$. Then using (2.1) we have

$$\|x - t\|_2^2 \leq \frac{1}{4}\sum_{i=1}^{n}\|b_i^*\|_2^2 \leq \frac{\|b_n^*\|_2^2}{4}\underbrace{\sum_{i=1}^{n}2^{n-i}}_{\leq 2^n} \leq 2^{n-2}\|b_n^*\|_2^2$$

□

Finally, we prove an approximation ratio of $2^{n/2}$. The intuition is as follows: either the value of $\text{CVP}(\Lambda, t)$ is at least $\frac{1}{2}\|b_n^*\|_2$ and the absolute guarantee from Lemma 2.5 is good enough for a $2^{n/2}$ approximation. Or the value of $\text{CVP}(\Lambda, t)$ is less than $\frac{1}{2}\|b_n^*\|_2$, but then the algorithm recurses on the "correct" translate that contains the optimum solution and the algorithm does not make any error in the first recursion step.

Theorem 2.6. *Let Λ be a lattice with LLL-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ and let $\boldsymbol{t} \in \mathrm{span}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ be a target vector. Then $\boldsymbol{x} :=$ NearestPlaneAlgo$(\boldsymbol{B}, \boldsymbol{t})$ satisfies $\|\boldsymbol{x} - \boldsymbol{t}\|_2 \leq 2^{n/2}\mathrm{Cvp}(\Lambda, \boldsymbol{t})$.*

Proof. We fix an optimum solution $\boldsymbol{x}^* \in \Lambda$, i.e. $\|\boldsymbol{x}^* - \boldsymbol{t}\|_2 = \mathrm{Cvp}(\Lambda, \boldsymbol{t})$. We prove the claim by induction over n. Again we may assume that $\alpha = 0$ (and $t \in K$) as since shifting $\boldsymbol{t}$ by a lattice point does not affect the approximation ratio. If it happens that the optimum value satisfies $\|\boldsymbol{x}^* - \boldsymbol{t}\|_2 \geq \frac{1}{2}\|\boldsymbol{b}_n^*\|_2$, then by Lemma 2.5 we know that our solution satisfies $\|\boldsymbol{x} - \boldsymbol{t}\|_2 \leq 2^{n/2-1}\|\boldsymbol{b}_n^*\|_2 \leq 2^{n/2-1} \cdot 2\|\boldsymbol{x}^* - \boldsymbol{t}\|_2 = 2^{n/2}\|\boldsymbol{x}^* - \boldsymbol{t}\|_2$ and we are done. Otherwise we have $\|\boldsymbol{x}^* - \boldsymbol{t}\|_2 < \frac{1}{2}\|\boldsymbol{b}_n^*\|_2$. Recall that $|\langle \boldsymbol{t}, \boldsymbol{b}_n^* \rangle| \leq \frac{1}{2}\|\boldsymbol{b}_n^*\|_2^2$ and so $|\langle \boldsymbol{x}^*, \boldsymbol{b}_n^* \rangle| < \|\boldsymbol{b}_n^*\|_2^2$. As $\boldsymbol{x}^* \in \Lambda$, this means that $\boldsymbol{x}^*$ has to lie on the shift going through the origin, i.e. $\boldsymbol{x}^* \in \Lambda \cap U$. In particular, $\boldsymbol{x}^*$ is still an optimum solution for $\mathrm{Cvp}((\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}), \pi_U(\boldsymbol{t}))$ and still $\boldsymbol{x} :=$ NearestPlaneAlgo$((\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}), \pi_U(\boldsymbol{t}))$. Then by induction $\|\boldsymbol{x} - \pi_U(\boldsymbol{t})\|_2 \leq 2^{(n-1)/2}\|\boldsymbol{x}^* - \pi_U(\boldsymbol{t})\|_2$. Hence

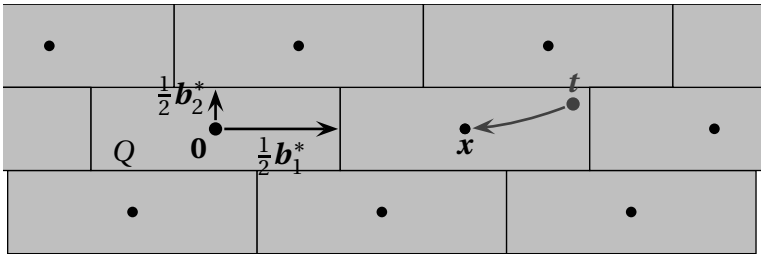
$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{t}\|_2^2 &\overset{\text{orthog.}}{=} \|\boldsymbol{t} - \pi_U(\boldsymbol{t})\|_2^2 + \|\boldsymbol{x} - \pi_U(\boldsymbol{t})\|_2^2 \\ &\leq 2^{n-1}\left(\|\boldsymbol{t} - \pi_U(\boldsymbol{t})\|_2^2 + \|\boldsymbol{x}^* - \pi_U(\boldsymbol{t})\|_2^2\right) \\ &\overset{\text{orthog.}}{=} 2^{n-1}\|\boldsymbol{x}^* - \boldsymbol{t}\|_2^2 \end{aligned}$$

which satisfies the claim. □

A second look at the algorithm also reveals which lattice point the algorithm actually finds:

Corollary 2.7. Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice with basis $\boldsymbol{B}$. Denote $Q := \{\sum_{i=1}^n y_i \boldsymbol{b}_i^* : \boldsymbol{y} \in [-\frac{1}{2}, \frac{1}{2}]^n\}$. Then $\boldsymbol{x} :=$ NearestPlaneAlgo$(\boldsymbol{B}, \boldsymbol{t})$ gives a lattice point so that $\boldsymbol{t} \in \boldsymbol{x} + Q$.

One can easily see the claim by induction; we skip the proof here.



Another useful property is that if the target vector $\boldsymbol{t}$ is close enough to the lattice then one can even compute an optimum solution in polynomial time. We leave the proof as an exercise (see Exercise 2.5).

Lemma 2.8. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $\boldsymbol{t} \in \mathbb{R}^n$ be a vector with $\mathrm{CVP}(\Lambda, \boldsymbol{t}) < 2^{-n/2-1}\lambda_1(\Lambda)$. Then one can find a vector $\boldsymbol{x} \in \Lambda$ with $\|\boldsymbol{x}-\boldsymbol{t}\|_2 = \mathrm{CVP}(\Lambda, \boldsymbol{t})$ in polynomial time.*

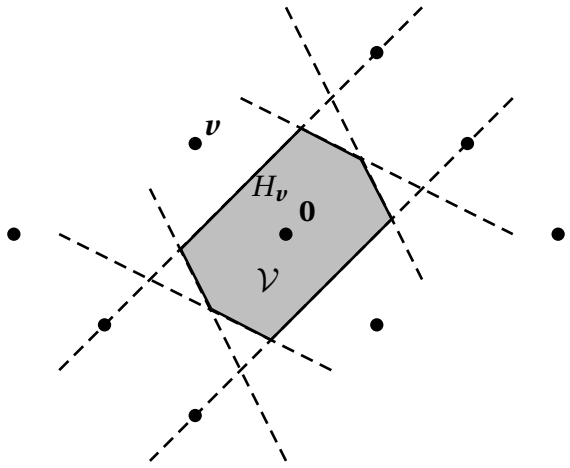
2.3 The Voronoi cell algorithm of Micciancio and Voulgaris

In this section, we want to explain an important concept that is the base for the algorithm of Micciancio and Voulgaris [MV10, MV13].

2.3.1 The Voronoi cell

For a full rank lattice $\Lambda := \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$, the *(closed) Voronoi cell* is the set of points in $\mathbb{R}^n$ that are closer to $\mathbf{0}$ than to any other lattice point (or at equal distance). Formally, we define

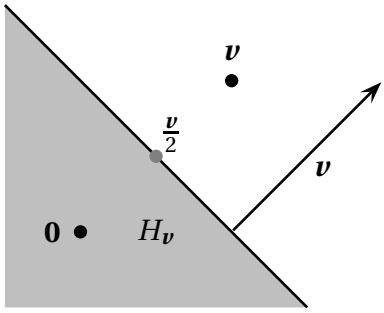
$$\mathcal{V} = \left\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}-\boldsymbol{v}\|_2 \ \forall \boldsymbol{v} \in \Lambda \setminus \{\mathbf{0}\}\right\}$$



Note that the set of points that are closer to $\mathbf{0}$ than to $\boldsymbol{v} \in \Lambda$ (or at equals distance) are exactly the set of points in the closed halfspace

$$H_{\boldsymbol{v}} = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}-\boldsymbol{v}\|_2\right\} = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{v}\rangle \leq \frac{1}{2}\|\boldsymbol{v}\|_2^2\right\}$$

Geometrically, the normal vector of this halfspace is $\boldsymbol{v}$ and it contains the point $\frac{\boldsymbol{v}}{2}$ on its boundary:

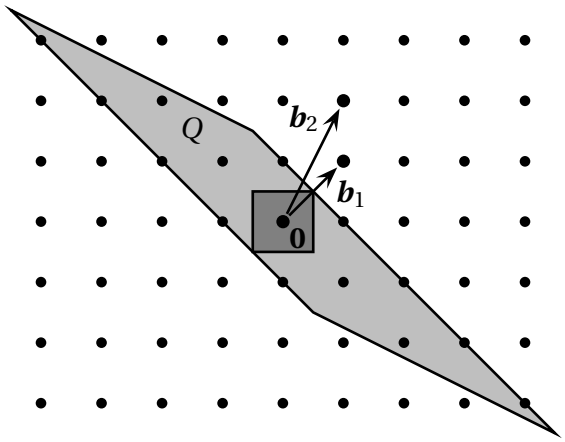


Using this definition, we can write $\mathcal{V}$ as intersection of all those halfspaces

$$\mathcal{V} = \bigcap_{\boldsymbol{v} \in \Lambda / \{\mathbf{0}\}} H_{\boldsymbol{v}}$$

From that definition, we see that $\mathcal{V}$ is a symmetric, closed convex set. The set must be *full-dimensional* as it contains the open ball of radius $\frac{1}{2}\lambda_1(\Lambda)$ around the origin.

We can give a brief argument why $\mathcal{V}$ is defined by *finitely* many half-spaces and hence is an *(open) polytope*. Consider the set $Q := \bigcap_{\boldsymbol{v} \in \{\pm \boldsymbol{b}_1, \ldots, \pm \boldsymbol{b}_n\}} H_{\boldsymbol{v}}$ where $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ is any basis for Λ. Then $\mathcal{V} \subseteq Q$ and as $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ are linearly independent, the set Q will be bounded.



Then for any $\boldsymbol{v} \in \Lambda \setminus 2Q$ one has that $\mathcal{V} \subseteq Q \subseteq H_{\boldsymbol{v}}$. Hence the only Voronoi-relevant vectors are the finitely many vectors in $\Lambda \cap 2Q$.

Every irredundant halfspace $H_{\boldsymbol{v}}$ describing the cell must induce a $(n-1)$-dimensional facet of $\mathcal{V}$. Hence there is a unique minimal set $R \subseteq \Lambda \setminus \{\mathbf{0}\}$ so that $\mathcal{V} = \bigcap_{\boldsymbol{v} \in R} H_{\boldsymbol{v}}$. We call that set R the *Voronoi-relevant vectors*.

We summarize a few properties:

Lemma 2.9 (Geometry of the Voronoi cell)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice with Voronoi cell $\mathcal{V}$.*

(a) *The set $\mathcal{V}$ is convex, full-dimensional, bounded and symmetric and $\|\cdot\|_{\mathcal{V}}$ is a norm.*

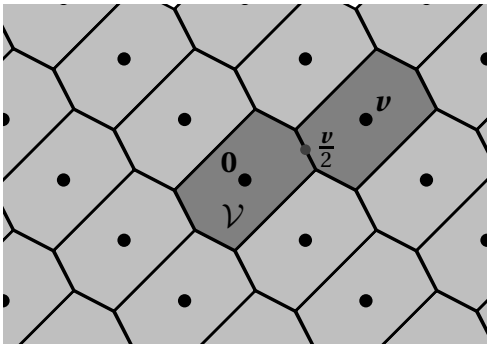
(b) *If $\mathcal{V} \cap (\boldsymbol{v} + \mathcal{V}) \neq \emptyset$ for $\boldsymbol{v} \in \Lambda$, then $\frac{\boldsymbol{v}}{2} \in \mathcal{V} \cap (\boldsymbol{v} + \mathcal{V})$.*

(c) *If $\boldsymbol{w} \in \Lambda$ is Voronoi-relevant, then the unique closest lattice vectors to $\frac{\boldsymbol{w}}{2}$ are $\mathbf{0}$ and $\boldsymbol{w}$.*

(d) *The translates $\boldsymbol{x} + \mathcal{V}$ with $\boldsymbol{x} \in \Lambda$ form a tiling of $\mathbb{R}^n$ (i.e. $\Lambda + \mathcal{V}$ covers every point in $\mathbb{R}^n$ exactly once except for the points of measure zero that have equal minimum distance to several lattice points).*

(e) *One has $\frac{\lambda_1(\Lambda)}{2} B_2^n \subseteq \mathcal{V} \subseteq n\lambda_n(\Lambda) B_2^n$.*

Proof. (a) and (c) follow from the above discussion. (d) is clear. For (b), let $\boldsymbol{x} \in \mathbb{R}^n$ be a point with $\|\boldsymbol{x}\|_{\mathcal{V}}, \|\boldsymbol{x} - \boldsymbol{v}\|_{\mathcal{V}} \leq 1$. Then $\|\frac{\boldsymbol{v}}{2}\|_{\mathcal{V}} \leq \|\frac{\boldsymbol{v}}{2} - \frac{\boldsymbol{x}}{2}\|_{\mathcal{V}} + \|\frac{\boldsymbol{x}}{2}\|_{\mathcal{V}} \leq \frac{1}{2} + \frac{1}{2} = 1$. We will show (e) in the exercises. □



Now, things are getting a little more interesting as we show that there is a limited number of Voronoi relevant vectors:

Lemma 2.10. *The number of Voronoi relevant vectors is $|R| \leq 2^{n+1}$.*

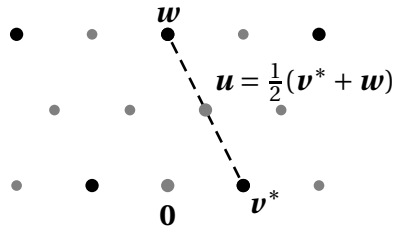
Proof. Recall that $\Lambda := \Lambda(\boldsymbol{B})$ where $\boldsymbol{B}$ has the columns $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$. Let $X := \{\sum_{i=1}^n \lambda_i \boldsymbol{b}_i \mid \lambda_i \in \{0,1\}\ \forall i = 1, \ldots, n\}$ be the vertices of the fundamental parallelepiped of the basis $\boldsymbol{B}$. Note that $|X| = 2^n$. Fix a vector $\boldsymbol{v} \in X$ and consider the translated lattice $\boldsymbol{v} + 2\Lambda$. We claim that *at most two* of the lattice vectors in that translated lattice are Voronoi relevant. In fact, we claim something stronger:
Claim. *Apart from $\boldsymbol{v}^* := argmin\{\|\boldsymbol{x}\|_2 : \boldsymbol{x} \in \boldsymbol{v} + 2\Lambda\}$ and $-\boldsymbol{v}^*$, there is no other Voronoi relevant vector in $\boldsymbol{v} + 2\Lambda$.*
Proof. Suppose that $\boldsymbol{w} \in \boldsymbol{v} + \Lambda$ is Voronoi-relevant. Let $\boldsymbol{v}^*$ be the shortest lattice

vector of the same parity as $\boldsymbol{w}$, that means $\boldsymbol{v}^* = \operatorname{argmin}\{\|\boldsymbol{x}\|_2 : \boldsymbol{x} \in \boldsymbol{v} + 2\Lambda\}$. We assume for the sake of contradiction that $\boldsymbol{w} \notin \{\boldsymbol{v}^*, -\boldsymbol{v}^*\}$.

We know that $\boldsymbol{u} := \frac{1}{2}(\boldsymbol{v}^* + \boldsymbol{w}) \in \Lambda$. We claim that the distance of the midpoint $\frac{\boldsymbol{w}}{2}$ to $\boldsymbol{u}$ is not larger than the distance to $\mathbf{0}$ and $\boldsymbol{w}$:

$$\left\| \frac{\boldsymbol{w}}{2} - \boldsymbol{u} \right\|_2 = \left\| \frac{\boldsymbol{w}}{2} - \frac{1}{2}(\boldsymbol{v}^* + \boldsymbol{w}) \right\|_2 = \frac{1}{2}\|\boldsymbol{v}^*\|_2 \leq \frac{1}{2}\|\boldsymbol{w}\|_2$$

This is a contradiction to Lemma 2.9.(c), which says that the closest lattice vectors to $\frac{\boldsymbol{w}}{2}$ are exactly $\mathbf{0}$ and $\boldsymbol{w}$.



□

The last lemma also provides us with a possibility to compute the Voronoi relevant vectors:

Lemma 2.11. *The Voronoi relevant vectors for a Voronoi cell $\mathcal{V}$ for lattice $\Lambda(\boldsymbol{B})$ can be computed by $2^{O(n)}$ many* CVP *calls with the same lattice $\Lambda(\boldsymbol{B})$ (but different target vectors).*

Proof. For every $\boldsymbol{\lambda} \in \{0,1\}^n$ we need to find the shortest vector in the shifted lattice $\boldsymbol{t} + 2\Lambda$ with $\boldsymbol{t} := \sum_{i=1}^n \lambda_i \boldsymbol{b}_i$. That is the same as solving $\text{CVP}(2\Lambda(\boldsymbol{B}), -\boldsymbol{t})$ which is the same as $\text{CVP}(\Lambda(\boldsymbol{B}), -\frac{\boldsymbol{t}}{2})$. □

It seems that if we want to use the Voronoi cell in an algorithm for CVP, we first have to solve $2^{O(n)}$ instances of CVP. This seems utterly stupid — but it will work!

2.3.2 Computing a closest vector

We can now come to the main algorithm which on input $\boldsymbol{t} \in \mathbb{R}^n$ computes a closest lattice vector $\boldsymbol{x} \in \Lambda(\boldsymbol{B})$. Here, we assume that we do have a description of the Voronoi cell in form of the Voronoi relevant vectors R. Note that this task is equivalent to finding a lattice vector $\boldsymbol{x} \in \Lambda(\boldsymbol{B})$ so that $\boldsymbol{t} - \boldsymbol{x} \in \mathcal{V}$. The algorithm is now as follows: starting at $\boldsymbol{t}$, subtract iteratively multiples of lattice vectors in R until we reach a point in the cell $\mathcal{V}$. Then the sum of the subtracted vectors will be the optimum solution $\boldsymbol{x}$. A detailed description is as follows:

Closest vector algorithm

Input: A lattice basis $\boldsymbol{B} \in \mathbb{R}^{n\times n}$, a target vector $\boldsymbol{t} \in \mathbb{R}^n$, the list of Voronoi relevant vectors $R \subseteq \Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}$ describing the Voronoi cell $\mathcal{V}$

Output: The lattice vector $\boldsymbol{x} \in \Lambda(\boldsymbol{B})$ minimizing $\|\boldsymbol{t}-\boldsymbol{x}\|_2$

(1) Set $\boldsymbol{s} := \boldsymbol{t}$

(2) WHILE $\boldsymbol{s} \notin \mathcal{V}$ DO

  (3) Compute the minimal value $\delta > 0$ so that $\boldsymbol{s} \in \delta \cdot \mathcal{V}$

  (4) Write $\delta = 2^k \cdot \alpha$ with $k \in \mathbb{Z}_{\geq 0}$ and $1 \leq \alpha < 2$

  (5) Find the vector $\boldsymbol{v} \in R$ so that $\boldsymbol{s}$ lies on the boundary of $\delta H_{\boldsymbol{v}}$

  (6) Update $\boldsymbol{s} := \boldsymbol{s} - 2^k \boldsymbol{v}$

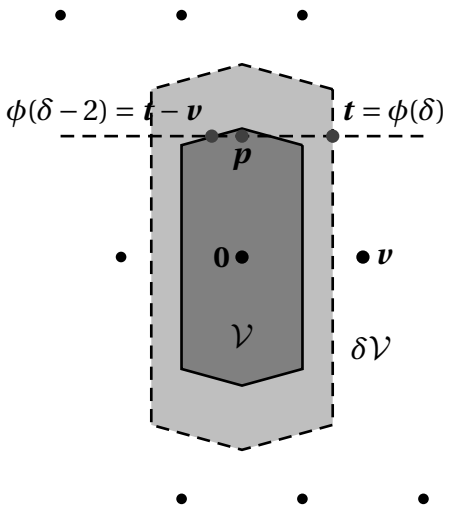
(7) Return $\boldsymbol{t} - \boldsymbol{s}$

For the analysis, it suffices to consider the situation with $k = 0$. In other words, we assume that the original vector $\boldsymbol{t}$ satisfies $\boldsymbol{t} \in 2\mathcal{V}$. We will show that within $2^{O(n)}$ iterations, we reach a point in $\mathcal{V}$. The general claim follows from the fact that $2^k\mathcal{V}$ is exactly the Voronoi cell of the scaled lattice $\Lambda(2^k\boldsymbol{B})$. Hence the algorithm iterates only over polynomially many choices of k.

So, let us understand what happens in one iteration.

Lemma 2.12. *Let $\boldsymbol{t} \in (2\mathcal{V}) \setminus \mathcal{V}$ and let $1 \leq \delta < 2$ be the minimal value so that $\boldsymbol{t} \in \delta\mathcal{V}$. Let $\boldsymbol{v} \in R$ be the Voronoi relevant vector so that $\boldsymbol{t}$ lies on the boundary of $\delta H_{\boldsymbol{v}}$. Then $\boldsymbol{t} - \boldsymbol{v} \in 2\mathcal{V}$ and $\|\boldsymbol{t}-\boldsymbol{v}\|_2 < \|\boldsymbol{t}\|_2$.*

Proof. Let $\boldsymbol{p} := \boldsymbol{t} - \delta\frac{\boldsymbol{v}}{2}$; note that $\boldsymbol{p} \perp \boldsymbol{v}$. Consider the line $\phi : \mathbb{R} \to \mathbb{R}^n$ with $\phi(\lambda) := \boldsymbol{p} + \lambda \cdot \frac{\boldsymbol{v}}{2}$ that satisfies $\phi(\delta) = \boldsymbol{t}$ and $\phi(\delta-2) = \boldsymbol{t}-\boldsymbol{v}$. Then $\|\phi(\lambda)\|_2$ is a symmetric, convex function which is minimized for $\lambda = 0$. Hence $\|\boldsymbol{t}-\boldsymbol{v}\|_2 = \|\phi(\delta-2)\|_2 < \|\phi(\delta)\|_2$ since $|\delta-2| < |\delta|$.



By assumption we have $\frac{\boldsymbol{t}}{\delta} \in \boldsymbol{v} + \mathcal{V}$ and so by the triangle inequality

$$\|\boldsymbol{t}-\boldsymbol{v}\|_{\mathcal{V}} \leq \underbrace{\left\|\frac{\boldsymbol{t}}{\delta}-\boldsymbol{v}\right\|_{\mathcal{V}}}_{\leq 1} + \left(1-\frac{1}{\delta}\right)\underbrace{\|\boldsymbol{t}\|_{\mathcal{V}}}_{\leq \delta} \leq \delta$$

□

The algorithm computes shorter and shorter vectors in $2\mathcal{V}$, so it will definitely terminate in finite time. It remains to show that only $2^{O(n)}$ iterations are needed.

Lemma 2.13. *Let $\mathcal{V}$ be the Voronoi cell of a lattice $\Lambda \subseteq \mathbb{R}^n$ and let $\boldsymbol{t} \in 2\mathcal{V}$. Then $|(\boldsymbol{t}-\Lambda) \cap 2\mathcal{V}| \leq 2^{O(n)}$.*

Proof. Since $\boldsymbol{t} \in 2\mathcal{V}$, it suffices to show that $|\Lambda \cap 4\mathcal{V}| \leq 2^{O(n)}$. Suppose $|\Lambda \cap 4\mathcal{V}| > 4^n$. By a counting argument, there must be two distinct vectors $\boldsymbol{x}, \boldsymbol{y} \in 4\mathcal{V} \cap \Lambda$ so that $\boldsymbol{x}-\boldsymbol{y} \in 4\Lambda \setminus \{\mathbf{0}\}$. Then $\boldsymbol{v} := \frac{1}{4}(\boldsymbol{x}-\boldsymbol{y})$ has $\boldsymbol{v} \in \Lambda$ and $\boldsymbol{v} \in \mathcal{V}$. But there is no lattice vector in $\mathcal{V}$ except $\mathbf{0}$. □

The lemma shows that for every fixed value of k, the algorithm iteratively finds shorter vectors and hence will not revisit a vector. Then after at most $2^{O(n)}$ iterations, the value of k will be decreased by one.

2.3.3 Putting things together

Setting up the recursion is not completely trivial in this case. Let us define two running times that we want to analyze

$$\begin{aligned} T_{\text{Voronoi}}(n) &= \text{time to compute the Voronoi cell in an } n\text{-dim. lattice} \\ T_{\text{CVP}}(n,k) &= \text{time to solve } k \text{ many CVP in the same } n\text{-dim. lattice} \end{aligned}$$

We can get the following sequence of recursions[2]:

$$\begin{aligned} T_{\text{Voronoi}}(n) &\overset{(1)}{\leq} T_{\text{CVP}}(n, 2^{O(n)}) \overset{(2)}{\leq} T_{\text{CVP}}(n-1, 2^{O(n)} \cdot 2^{O(n)}) \\ &\overset{(3)}{\leq} T_{\text{Voronoi}}(n-1) + 2^{O(n)} \cdot 2^{O(n)} \cdot 2^{O(n)}. \end{aligned}$$

Here we use in (1) that we can compute the Voronoi cell with $2^{O(n)}$ CVP computations in the same lattice. In (2), we use the dimension reduction argument from Section 2.1 saying that a CVP computation can be reduced to $2^{O(n)}$ CVP computations in the same $(n-1)$-dimensional lattice. Finally, in (3) we use that to solve k many CVP computations in the same lattice, we need to compute the Voronoi cell only *once* and then run a $2^{O(n)}$-time algorithm for each of the k target vectors. From the recursion we conclude that $T_{\text{Voronoi}}(n) \leq \sum_{k=1}^{n} 2^{O(k)} = 2^{O(n)}$ and $T_{\text{CVP}}(n,1) \leq 2^{O(n)}$ as well.

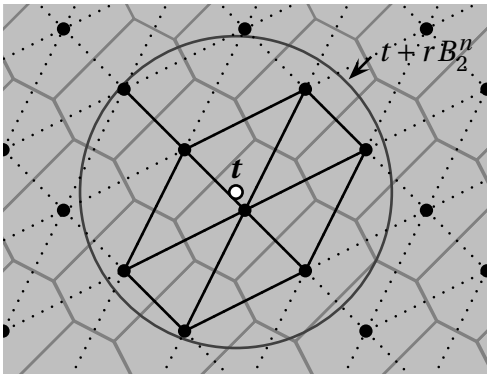
2.4 Enumerating lattice points

As we discussed, the algorithm by Micciancio and Voulgaris finds the closest lattice vector to a given target vector $\boldsymbol{t}$. But it is also possible to modify the algorithm in order to enumerate *all* lattice points close to $\boldsymbol{t}$. Additionally we will gain some geometric insights:

Theorem 2.14. *Given a lattice $\Lambda = \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$, an ellipsoid $\mathcal{E}$ and a vector $\boldsymbol{t} \in \mathbb{R}^n$, one can enumerate the points $S := \Lambda \cap (\mathcal{E} + \boldsymbol{t})$ in time $2^{O(n)} \cdot (|S|+1)$ times a polynomial in the encoding length of $\boldsymbol{B}$, $\mathcal{E}$ and $\boldsymbol{t}$.*

[2] As so often in the literature of lattice algorithms we have omitted polynomial factors depending on the encoding length of the lattice. For generally one could let $T_{\text{Voronoi}}(n,L)$ denote the time to compute the Voronoi cell in an n-dimensional lattice with L bits encoding length. Similar one can extend $T_{\text{Cvp}}(n,k,L)$ to be the time it takes to solve k many CVP's for target vectors t with $\|t\|_{\vec{\mathcal{V}}} \leq 2^{\text{poly}(n,L)}$ in a lattice of encoding length L. Then indeed $T_{\text{Voronoi}}(n,L) \overset{(1)}{\leq} T_{\text{CVP}}(n, 2^{O(n)}, L) \overset{(2)}{\leq} T_{\text{CVP}}(n-1, 2^{O(n)} \cdot 2^{O(n)}, L)$
$\overset{(3)}{\leq} T_{\text{Voronoi}}(n-1,L) + 2^{O(n)} \cdot 2^{O(n)} \cdot 2^{O(n)} \cdot \text{poly}(n,L)$.

Proof. After applying a linear transformation we may assume that $\mathcal{E} = rB_2^n$ for some $r > 0$. Let $R \subseteq \Lambda$ be the Voronoi-relevant vectors. Recall that $|R| \leq 2^{n+1}$ and moreover, we can compute R in time $2^{O(n)}$, see Lemma 2.10. Next, we define an infinite graph $G = (\Lambda, E)$ whose vertices are the lattice points and where the edges are of the form $E = \{\{\boldsymbol{x}, \boldsymbol{y}\} \mid \boldsymbol{x}, \boldsymbol{y} \in \Lambda \text{ and } \boldsymbol{x} - \boldsymbol{y} \in R\}$. Note that the graph has a degree of $|R|$.



Our argument rests crucially on the following insight:
Claim. *For any $\boldsymbol{t} \in \mathbb{R}^n$ and $r > 0$, the induced subgraph $G[S]$ with $S := \Lambda \cap (\boldsymbol{t} + rB_2^n)$ is connected.*
Proof of claim. By perturbing $\boldsymbol{t}$ and marginally increasing r without including new lattice points, we may assume that the closest lattice vector to $\boldsymbol{t}$ is unique. Let us call that lattice vector $\boldsymbol{x}^* \in \Lambda$. Next, fix any $\boldsymbol{x}_0 \in S$. We claim that there is a path from $\boldsymbol{x}_0$ to $\boldsymbol{x}^*$ in $G[S]$. Suppose that $\boldsymbol{x}_0 \neq \boldsymbol{x}^*$ since otherwise there is nothing to show. Then $\boldsymbol{x}_0 \notin \boldsymbol{t} + \mathcal{V}$. From Lemma 2.12 we know that there is a Voronoi-relevant vector $\boldsymbol{v} \in R$ so that $\|(\boldsymbol{x}_0 + \boldsymbol{v}) - \boldsymbol{t}\|_2 < \|\boldsymbol{x} - \boldsymbol{t}\|_2$. We set $\boldsymbol{x}_1 := \boldsymbol{x}_0 + \boldsymbol{v} \in S$ and note that $\{\boldsymbol{x}_0, \boldsymbol{x}_1\} \in E$. Then we can continue a path $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in S$ with $\|\boldsymbol{x}_0 - \boldsymbol{t}\|_2 > \|\boldsymbol{x}_1 - \boldsymbol{t}\|_2 > \|\boldsymbol{x}_2 - \boldsymbol{t}\|_2 > \ldots$. That path must be finite and terminate in $\boldsymbol{x}^*$.

To conclude the argument, we note that $\boldsymbol{x}^*$ can be found in time $2^{O(n)}$ and one can explore the connected graph $G[S]$ by trying the directions R from each discovered point, paying a running time of at most $2^{O(n)}$ per point. □

2.5 Exercises

Exercise 2.1.
Let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice with LLL-reduced basis $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ and Gram Schmidt orthogonalization $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$. We abbreviate $\mu_{i,j} = \frac{\langle \boldsymbol{b}_j, \boldsymbol{b}_i^* \rangle}{\|\boldsymbol{b}_i^*\|_2^2}$. First, we fix an arbitrary $\boldsymbol{x} =$

$\boldsymbol{B}\boldsymbol{y}$ with $\boldsymbol{y} \in \mathbb{R}^n$.

(i) Prove that $\|\boldsymbol{x}\|_2^2 = \sum_{k=1}^n \|\boldsymbol{b}_k^*\|_2^2 \cdot \left(y_k + \sum_{j>k} \mu_{k,j} y_j\right)^2$

(ii) Prove that for all $k \in \{1,\ldots,n\}$ one has $\|\boldsymbol{x}\|_2^2 \geq 2^{-k} \|\boldsymbol{b}_k\|_2^2 \cdot \max\{|y_k| - \frac{1}{2}\sum_{j>k} |y_j|, 0\}^2$.

Now fix a $\boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}$ with $\|\boldsymbol{x}\|_2 = \lambda_1(\Lambda)$ and let $\boldsymbol{y} \in \mathbb{Z}^n$ be so that $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}$.

(iii) Prove that for all $k \in \{1,\ldots,n\}$ one has $|y_k| \leq \max\{2^{(k+2)/2}, \sum_{j>k} |y_j|\}$.

(iv) Prove that for all $k \in \{1,\ldots,n\}$ one has $|y_k| \leq 2^{3n-k}$.

Exercise 2.2.
One might wonder whether the algorithm can be modified to work with different norms, say with $\|\cdot\|_\infty$. To examine this, consider the points $(0,0)$ and $(2,1)$ in $\mathbb{R}^2$ and draw the region of points that are closer to $(0,0)$ than to $(2,1)$ with respect to the $\|\cdot\|_\infty$-norm. What do you think, can you guarantee that Voronoi cells with respect to $\|\cdot\|_\infty$ are convex?

Exercise 2.3.
In this exercise, we want to show that CVP is not harder than CVP in the sense that we can use an oracle for CVP to solve the CVP problem. We denote $\text{CVP}(\boldsymbol{B}', \boldsymbol{t}) := \text{argmin}\{\|\boldsymbol{x} - \boldsymbol{t}\|_2 : \boldsymbol{x} \in \Lambda(\boldsymbol{B}')\}$. Suppose that $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ is the input basis for our CVP problem. Consider the following algorithm:

(1) FOR $i = 1$ TO n DO

  (2) Set $\boldsymbol{v}_i := \text{CVP}((\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}, 2\boldsymbol{b}_i, \boldsymbol{b}_{i+1}, \ldots, \boldsymbol{b}_n), \boldsymbol{b}_i)$.

(3) Return the shortest vector in $\{\boldsymbol{v}_i - \boldsymbol{b}_i \mid i = 1, \ldots, n\}$

Note that the algorithm calls the CVP oracle only n times on a lattice of dimension n. Prove that the algorithm returns the shortest vector in $\Lambda(\boldsymbol{B}) \setminus \{\boldsymbol{0}\}$.
Remark: There is no natural reduction known that goes the other way. Both problems are **NP**-hard, so there will be *some* reduction. But any known reduction from CVP to CVP causes at least a quadratic blowup in the dimension.

Exercise 2.4.
Let $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ be a regular matrix. Prove that $\mathcal{V} \subseteq n \cdot \lambda_n(\Lambda) \cdot B_2^n$ where $\mathcal{V}$ is the Voronoi cell of the lattice $\Lambda := \Lambda(\boldsymbol{B})$.

Exercise 2.5.
For a lattice Λ let us write $\text{CVP}(\Lambda, t)$ as the value of the closest vector problems. We have seen in an earlier exercise that for any lattice Λ, the number of lattice vectors of length, say $2 \cdot \lambda_1(\Lambda)$ is bounded by $2^{O(n)}$. Here we want to show that this is not true anymore for

CVP. To be precise, for any function $f(n)$, construct a lattice in n dimensions and a point $\boldsymbol{t}$ so that $|\{\boldsymbol{x} \in \Lambda \mid \|\boldsymbol{x}-\boldsymbol{t}\|_2 \leq 2 \cdot \text{CVP}(\Lambda, \boldsymbol{t})\}| \geq f(n)$.

Exercise 2.6.
Give a formal proof for the following claim: For any lattice basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ with Gram-Schmidt orthogonalization $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*$ define $Q := \{\sum_{i=1}^n \lambda_i \boldsymbol{b}_i^* \mid \frac{1}{2} \leq \lambda_i < \frac{1}{2} \; \forall i \in [n]\}$. Then $\boldsymbol{x}+Q$ for $\boldsymbol{x} \in \Lambda(\boldsymbol{B})$ form a *tiling* of $\mathbb{R}^n$ (meaning that for each $\boldsymbol{t} \in \mathbb{R}^n$ there is exactly one $\boldsymbol{x} \in \Lambda(\boldsymbol{B})$ with $\boldsymbol{t} \in \boldsymbol{x}+Q$).

Exercise 2.7.
Prove the statement from Lemma 2.8: Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $\boldsymbol{t} \in \mathbb{R}^n$ be a vector with $\text{CVP}(\Lambda, \boldsymbol{t}) < 2^{-n/2-1}\lambda_1(\Lambda)$. Then one can find a vector $\boldsymbol{x} \in \Lambda$ with $\|\boldsymbol{x}-\boldsymbol{t}\|_2 = \text{CVP}(\Lambda, \boldsymbol{t})$ in polynomial time.

Exercise 2.8.
For the CLOSEST VECTOR PROBLEM WITH PREPROCESSING (CVPP), an algorithm is first given the basis $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ of a full rank lattice $\Lambda := \Lambda(\boldsymbol{B})$. Then the algorithm may perform any computation of any length. Then the algorithm is given a target vector $\boldsymbol{t} \in \mathbb{R}^n$ and has to find a lattice vector $\boldsymbol{x} \in \Lambda$ so that $\|\boldsymbol{x}-\boldsymbol{t}\|_2$ is minimized. Prove that there is an algorithm for CVPP (whose 2nd stage runs in polynomial time) that finds a $O(n^{1.5})$-approximation.

Chapter 3

The Sieving Algorithm

In this chapter, we discuss the *Sieving algorithm* due to Ajtai, Kumar and Sivakumar [AKS01]. Later on the algorithm has appeared in many variations in the literature. In terms of exposition we follow again the lecture notes of Regev [Reg09a], though we keep it more general and state the algorithm somewhat differently. Here we make no attempt to optimize any constants.

In the following, we fix a full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ and a symmetric convex body $K \subseteq \mathbb{R}^n$. We define $\lambda_1(\Lambda, K) = \min\{\|\boldsymbol{x}\|_K : \boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}\}$ as the length of the shortest non-zero vector with respect to the norm $\|\cdot\|_K$. The main result of this chapter will be the following:

Theorem 3.1. *Given a lattice $\Lambda \subseteq \mathbb{R}^n$ and a symmetric convex body $K \subseteq \mathbb{R}^n$, there is a randomized algorithm that with high probability finds a vector attaining $\min\{\|\boldsymbol{x}\|_K : \boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}\}$ in time $2^{O(n)}$.*

The original work of [AKS01] and the notes by Regev [Reg09a] restrict their attention to the Euclidean norm $\|\cdot\|_2$, but unlike the Voronoi cell algorithm from Chapter 2.3, the AKS algorithm does not actually use any property specific to the Euclidean case.

3.1 The algorithm

Since we can approximate any norm $\|\cdot\|_K$ by a Euclidean norm[1], we may use the LLL-algorithm to determine $\lambda_1(\Lambda, K)$ up to a factor of $2^{O(n)}$; then by trying out

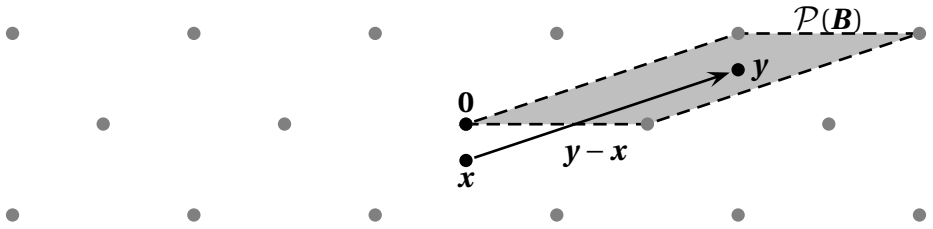
[1] To be precise, one can rephrase John's theorem as follows: For any symmetric convex body $K \subseteq \mathbb{R}^n$ there is a positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ so that $\|\boldsymbol{A}\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_K \leq \sqrt{n} \cdot \|\boldsymbol{A}\boldsymbol{x}\|_2$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

$O(n)$ candidate values and rescaling the lattice we may assume that $1 \le \lambda_1(\Lambda, K) < 1.5$. We will use the (non-trivial) fact that given a symmetric convex body one can compute a uniform random sample $\boldsymbol{x} \sim K$ in polynomial time (up to a negligible statistical error); see [DFK91] for details.

Next, fix any basis $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ for the lattice Λ (the basis does not need to be LLL-reduced). Recall that $\mathcal{P}(\boldsymbol{B}) = \{\boldsymbol{B}\boldsymbol{\lambda} : \boldsymbol{\lambda} \in [0,1)^n\}$ denotes the *fundamental parallelepiped*. For a vector $\boldsymbol{x} \in \mathbb{R}^n$ we define $\boldsymbol{y} := \boldsymbol{x} \mod \mathcal{P}(\boldsymbol{B})$ as the unique point so that $\boldsymbol{y} \in \mathcal{P}(\boldsymbol{B})$ and $\boldsymbol{x} - \boldsymbol{y} \in \Lambda$. That point can be efficiently computed as follows: write $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{\lambda}$ for some $\boldsymbol{\lambda} \in \mathbb{R}^n$. Then

$$\boldsymbol{y} = \sum_{i=1}^{n} (\lambda_i - \lfloor \lambda_i \rfloor) \cdot \boldsymbol{b}_i \in \mathcal{P}(\boldsymbol{B}) \quad \text{and} \quad \boldsymbol{y} - \boldsymbol{x} = \sum_{i=1}^{n} -(\lfloor \lambda_i \rfloor) \boldsymbol{b}_i \in \Lambda$$

In particular if we draw $\boldsymbol{x}$ at random, then $\boldsymbol{y} - \boldsymbol{x}$ is a random lattice vector, though it is not entirely clear what properties that distribution would have. In particular if the basis is far from orthogonal, then even if $\boldsymbol{x}$ is rather short, the vector $\boldsymbol{y} - \boldsymbol{x}$ can be very long.



The sieving algorithm then works as follows:

Sieving algorithm

Input: Lattice $\Lambda = \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$, norm $\|\cdot\|_K$, $N \in \mathbb{N}$
Output: Sequence $(\boldsymbol{x}_t, \boldsymbol{y}_t)_{t \in T}$ where $T \subseteq [N]$

(1) SET $L := \emptyset$ and $T := \emptyset$
(2) FOR $t = 1$ TO N DO
  (3) Draw $\boldsymbol{x} \sim K$ uniformly at random
  (4) Compute $\boldsymbol{y} := \boldsymbol{x} \mod \mathcal{P}(\boldsymbol{B})$
  (5) WHILE $\|\boldsymbol{y}\|_K > 4$ DO
    (6) IF there is an $(\boldsymbol{x}', \boldsymbol{y}') \in L$ with $\|\boldsymbol{y} - \boldsymbol{y}'\|_K \le \frac{1}{2}\|\boldsymbol{y}\|_K$ THEN
      (7) Replace $\boldsymbol{y}$ by $\boldsymbol{y} - (\boldsymbol{y}' - \boldsymbol{x}')$
    (8) ELSE
      (9) Add $(\boldsymbol{x}, \boldsymbol{y})$ to L
      (10) GOTO (2) (and increase t)
  (11) SET $(\boldsymbol{x}_t, \boldsymbol{y}_t) := (\boldsymbol{x}, \boldsymbol{y})$
  (12) Add t to T

An informal description of the algorithm is as follows: we generate a random pair $(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{x} \sim K$ and $\boldsymbol{y} - \boldsymbol{x} \in \Lambda$. Then we iteratively try to reduce the length of the vector by subtracting a lattice vector $\boldsymbol{y}' - \boldsymbol{x}'$ with $(\boldsymbol{x}', \boldsymbol{y}') \in L$. An important subtlety is that we measure the length of the current iterate as $\|\boldsymbol{y}\|_K$ and not as the length of the lattice vector $\boldsymbol{y} - \boldsymbol{x}$. Even though the difference is only $\|\boldsymbol{y} - (\boldsymbol{y} - \boldsymbol{x})\|_K \leq 1$, this will be crucial in the analysis. The reduction stops when the length reaches a threshold of 4. If the reduction fails, then we add the reduced pair $(\boldsymbol{x}, \boldsymbol{y})$ to the list L. Note that the algorithm generates an output sequence $(\boldsymbol{x}_t, \boldsymbol{y}_t)_{t \in T}$ where $\|\boldsymbol{y}_t\|_K \leq 4$ and $\boldsymbol{x}_t \in K$.

3.2 The analysis

Now we start the formal analysis.

Lemma 3.2. *At any point in the algorithm, the current pair $(\boldsymbol{x}, \boldsymbol{y})$ and any pair added to L satisfy $\boldsymbol{y} - \boldsymbol{x} \in \Lambda$.*

Proof. When we initialize the pair in (3), we have $\boldsymbol{y} = \boldsymbol{x} \mod \mathcal{P}(\boldsymbol{B})$ and so $\boldsymbol{y} - \boldsymbol{x} \in \Lambda$ as discussed previously. In any reduction step where $\boldsymbol{y} - \boldsymbol{x} \in \Lambda$ and $\boldsymbol{y}' - \boldsymbol{x}' \in \Lambda$, also the new pair satisfies $\boldsymbol{y} - (\boldsymbol{y}' - \boldsymbol{x}') - \boldsymbol{x} \in \Lambda$. □

In the following it will be convenient to abbreviate $M := \max_{i=1,\ldots,n} \|\boldsymbol{b}_i\|_K$. At certain points we will assume that $M \geq 1$. We verify that the WHILE loop actually terminates quickly:

Lemma 3.3. *The WHILE loop of (5) terminates after $O(\log(nM))$ iterations.*

Proof. Consider an iteration of the WHILE loop with current iterate $(\boldsymbol{x}, \boldsymbol{y})$ and suppose $(\boldsymbol{x}', \boldsymbol{y}') \in L$ is the pair with $\|\boldsymbol{y} - \boldsymbol{y}'\|_K \leq \frac{1}{2}\|\boldsymbol{y}\|_K$ that is being used in the reduction. Then the next iterate has length

$$\|\boldsymbol{y} - (\boldsymbol{y}' - \boldsymbol{x}')\|_K \leq \|\boldsymbol{y} - \boldsymbol{y}'\|_K + \|\boldsymbol{x}'\|_K \leq \frac{1}{2}\underbrace{\|\boldsymbol{y}\|_K}_{\geq 4} + 1 \leq \frac{3}{4}\|\boldsymbol{y}\|_K,$$

meaning the length decreases geometrically. At any moment in the WHILE loop, the iterate $(\boldsymbol{x}, \boldsymbol{y})$ satisfies $4 \leq \|\boldsymbol{y}\|_K \leq \sum_{i=1}^n \|\boldsymbol{b}_i\|_K \leq nM$ which then gives the claim. □

We also need a standard packing argument:

Lemma 3.4. *Let $K \subseteq \mathbb{R}^n$ be any symmetric convex body and let $Z \subseteq \alpha K$ so that $\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_K \geq \beta$ for all distinct $\boldsymbol{z}_1, \boldsymbol{z}_2 \in Z$. Then $|Z| \leq (\frac{2\alpha}{\beta} + 1)^n$.*

Proof. The interior of the translates $\boldsymbol{z}+\frac{\beta}{2}K$ are disjoint for all $\boldsymbol{z}\in Z$ while they are contained in $(\alpha+\frac{\beta}{2})K$. Comparing the volumes gives

$$|Z|\cdot\Big(\frac{\beta}{2}\Big)^n \mathrm{Vol}_n(K) = \mathrm{Vol}_n\Big(\bigcup_{\boldsymbol{z}\in Z}\Big(\boldsymbol{z}+\frac{\beta}{2}K\Big)\Big) \le \mathrm{Vol}_n\Big(\Big(\alpha+\frac{\beta}{2}\Big)K\Big) = \Big(\alpha+\frac{\beta}{2}\Big)^n \mathrm{Vol}_n(K)$$

Rearranging gives $|Z| \le (\frac{\alpha+\beta/2}{\beta/2})^n = (\frac{2\alpha}{\beta}+1)^n$. □

We can use this packing argument to show that the list L does not grow too large:

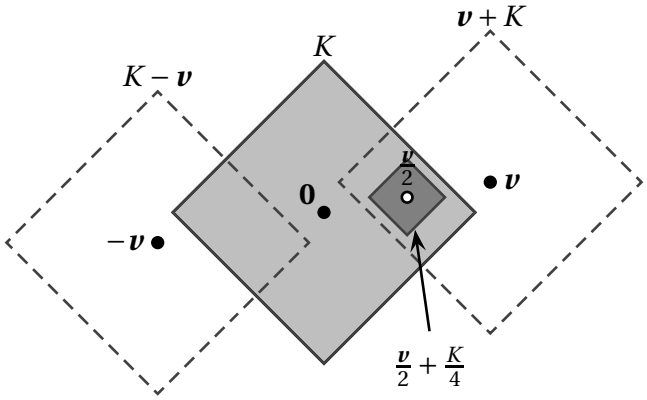
Lemma 3.5. *At any point in the algorithm one has* $|L| \le 9^n \cdot O(\ln(nM))$.

Proof. For some radius α consider $R_\alpha := \{\boldsymbol{y}' \mid (\boldsymbol{x}',\boldsymbol{y}') \in L \text{ and } \frac{\alpha}{2} \le \|\boldsymbol{y}'\|_K \le \alpha\}$. Consider the moment when the algorithm adds a pair $(\boldsymbol{x},\boldsymbol{y})$ to L so that $\boldsymbol{y}\in R_\alpha$. Then $\|\boldsymbol{y}-\boldsymbol{y}'\|_K > \frac{1}{2}\|\boldsymbol{y}\|_K \ge \frac{\alpha}{4}$ for all $\boldsymbol{y}'\in R_\alpha$. Then by induction one can easily prove that all $\boldsymbol{y}_1,\boldsymbol{y}_2\in R_\alpha$ satisfy $\|\boldsymbol{y}_1-\boldsymbol{y}_2\|_K \ge \frac{\alpha}{4}$. Hence by Lemma 3.4 we have $|R_\alpha| \le (\frac{2\alpha}{\alpha/4}+1)^n = 9^n$ for all α. Accounting for the different length classes we have $|L| \le \sum_{\alpha\in 2^{\mathbb{Z}}:8\le\alpha\le 2nM} |R_\alpha| \le 9^n\cdot O(\log(nM))$ which gives the claim. □

Lemma 3.5 then guarantees that the number of successful iterations is $|T| \ge N - 9^n\cdot\Theta(\ln(nM))$. Moreover, we already know that for each $t\in T$ one has $\boldsymbol{y}_t - \boldsymbol{x}_t \in 5K\cap\Lambda$. But so far we did not exclude the possibility that all such lattice vectors are $\mathbf{0}$. We now come to the ingenious argument that guarantees that this is not the case and indeed the output of the algorithm can be used to derive the shortest vector. More precisely, we will use the randomness coming from the initialization $\boldsymbol{x}\sim K$ to argue that for every short enough lattice vector $\boldsymbol{v}$ there is a lattice vector $\boldsymbol{w}$, so that the algorithm returns $\boldsymbol{y}_{t_1}-\boldsymbol{x}_{t_1}=\boldsymbol{w}$ and $\boldsymbol{y}_{t_2}-\boldsymbol{x}_{t_2}=\boldsymbol{w}+\boldsymbol{v}$ for some iterations t_1,t_2. In the following, we use the term "overwhelming probability" when the chance of failure is $2^{-n^{\omega(1)}}$.

Lemma 3.6. *Assume* $1\le\lambda_1(\Lambda,K)\le 1.5$ *and fix any* $\boldsymbol{v}\in 1.5K\cap\Lambda$. *Then if* $N\ge 48^n\cdot C\ln(nM)$ *for a large enough constant* C, *then with overwhelming probability,* $\boldsymbol{v}$ *is contained in the set of differences* $\{(\boldsymbol{y}_i-\boldsymbol{x}_i)-(\boldsymbol{y}_j-\boldsymbol{x}_j) : i,j\in T\}$.

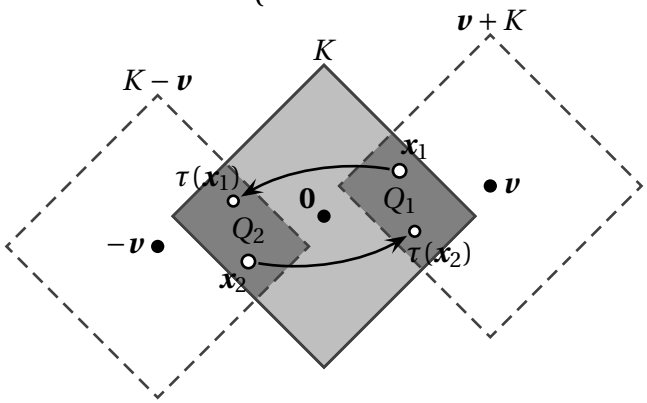
Proof. Consider the regions $Q_1 := K\cap(K+\boldsymbol{v})$ and $Q_2 := K\cap(K-\boldsymbol{v})$. First we prove that the regions are large enough for our purpose:
Claim I. *One has* $Vol_n(Q_1) = Vol_n(Q_2) \ge 4^{-n} Vol_n(K)$.
Proof of Claim I. By symmetry one clearly has $\mathrm{Vol}_n(Q_1)=\mathrm{Vol}_n(Q_2)$. Observe that $\frac{\boldsymbol{v}}{2}+\frac{K}{4}\subseteq K\cap(K+\boldsymbol{v})$ since for any $\boldsymbol{x}\in K$ one has $\|\frac{\boldsymbol{v}}{2}+\frac{\boldsymbol{x}}{4}\|_K \le \frac{1}{2}\|\boldsymbol{v}\|_K+\frac{1}{4}\|\boldsymbol{x}\|_K \le \frac{3}{4}+\frac{1}{4}=1$ by the triangle inequality and so $\frac{\boldsymbol{v}}{2}+\frac{K}{4}\subseteq K$. Similarly $\|(\frac{\boldsymbol{v}}{2}+\frac{\boldsymbol{x}}{4})-\boldsymbol{v}\|_K \le 1$ for $\boldsymbol{x}\in K$ and so $\frac{\boldsymbol{v}}{2}+\frac{K}{4}\subseteq \boldsymbol{v}+K$.



Then $\mathrm{Vol}_n(Q_1) \geq \mathrm{Vol}_n(\frac{1}{4}K) = 4^{-n}\mathrm{Vol}_n(K)$. □

We define a bijection $\tau : K \to K$ with

$$\tau(\boldsymbol{x}) := \begin{cases} \boldsymbol{x}+\boldsymbol{v} & \text{if } \boldsymbol{x} \in Q_2 \\ \boldsymbol{x}-\boldsymbol{v} & \text{if } \boldsymbol{x} \in Q_1 \\ \boldsymbol{x} & \text{otherwise} \end{cases}$$



Since $\boldsymbol{v} \in \Lambda$, we have $\boldsymbol{x} \mod \mathcal{P}(\boldsymbol{B}) = \tau(\boldsymbol{x}) \mod \mathcal{P}(\boldsymbol{B})$ for all $\boldsymbol{x} \in K$. Purely for the sake of analysis we will modify the algorithm in a way that does not change its behavior but will allow us to prove its correctness. We replace step (3) with the following:

(3') Draw $\boldsymbol{x} \sim K$. With probability 1/2, replace $\boldsymbol{x}$ by $\tau(\boldsymbol{x})$.

We call the operation of replacing $\boldsymbol{x}$ by $\tau(\boldsymbol{x})$ "tossing $\boldsymbol{x}$". Observe that tossing does not change the behaviour of the algorithm since $\tau(\boldsymbol{x})$ is still uniform from K. Also note that the modified algorithm needs access to $\boldsymbol{v}$ in order to compute τ; hence indeed this is not an algorithm that we could actually implement.

Recall that $\boldsymbol{x} \mod \mathcal{P}(\boldsymbol{B}) = \tau(\boldsymbol{x}) \mod \mathcal{P}(\boldsymbol{B})$ and so the initial choice of $\boldsymbol{y}$ does not depend on whether $\boldsymbol{x}$ was tossed. Then let us defer the decision whether $\boldsymbol{x}$ is tossed; we will make that decision when a pair $(\boldsymbol{x}, \boldsymbol{y})$ is added to L and at the very end of the algorithm we make that decision for each $\boldsymbol{x}_t$ with $t \in T$.

Consider the iterations $T^* := \{t \in T \mid \boldsymbol{x}_t \in Q_1 \cup Q_2\}$ and for each lattice vector $\boldsymbol{w} \in \Lambda$, we denote $T^*_{\boldsymbol{w}} := \{t \in T^* \mid \boldsymbol{w} = \boldsymbol{y}_t - \boldsymbol{x}_t\}$ (here we mean the vectors *before* the tossing decision is made).

Claim II. *With overwhelming probability, there is a vector $\boldsymbol{w} \in \Lambda$ so that $|T^*_{\boldsymbol{w}}| \geq (\frac{12}{11})^n$.*

Proof of Claim. In each of the iterations t, we sample $\boldsymbol{x} \sim K$ independently. By Claim I, we have an expected number of $\frac{N}{4^n}$ iterations where $\boldsymbol{x} \in Q_1 \cap Q_2$. By a standard Chernov bound argument, the actual number of such iterations is at least $\frac{1}{2} \cdot \frac{N}{4^n}$ with overwhelming probability. Conditioning on this event we have $|T^*| \geq \frac{1}{2}\frac{N}{4^n} - 9^n \cdot O(\ln(nM)) \geq 12^n$ if C is chosen large enough.

Next, denote $W := \{\boldsymbol{w} \in \Lambda : T^*_{\boldsymbol{w}} \neq \emptyset\}$. Note that for any t one has $\|\boldsymbol{y}_t - \boldsymbol{x}_t\|_K \leq \|\boldsymbol{y}_t\|_K + \|\boldsymbol{x}_t\|_K \leq 5$, implying that $W \subseteq 5K$. Moreover for distinct $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$ one has $\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_K \geq \lambda_1(\Lambda, K) \geq 1$. Then we can bound the size of W as $|W| \leq (2 \cdot 5 + 1)^n = 11^n$ using Lemma 3.4. Then there must be some lattice vector $\boldsymbol{w} \in \Lambda$ so that $|T^*_{\boldsymbol{w}}| \geq \frac{|T^*|}{|W|} \geq \frac{12^n}{11^n}$. □

Fix the vector $\boldsymbol{w} \in \Lambda$ from Claim II and condition on the event $|T^*_{\boldsymbol{w}}| \geq (\frac{12}{11})^n$ to happen. Consider our modified hypothetical algorithm which at the very end, for each $t \in T$ flips an independent coin to decide whether to write $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ or $(\tau(\boldsymbol{x}_t), \boldsymbol{y}_t)$ in the output list. Again with overwhelming probability, there will be an iteration $t_1 \in T^*_{\boldsymbol{w}}$ which is tossed (say with $\boldsymbol{x}_{t_1} \in Q_1$ for symmetry reasons) and another iteration $t_2 \in T^*_{\boldsymbol{w}}$ which is not tossed. Then their difference vector is

$$(\boldsymbol{y}_{t_1} - \tau(\boldsymbol{x}_{t_1})) - (\boldsymbol{y}_{t_2} - \boldsymbol{x}_{t_2}) = (\underbrace{\boldsymbol{y}_{t_1} - \boldsymbol{x}_{t_1}}_{=\boldsymbol{w}} + \boldsymbol{v}) - \underbrace{(\boldsymbol{y}_{t_2} - \boldsymbol{x}_{t_2})}_{=\boldsymbol{w}} = \boldsymbol{v}$$

as claimed. □

This concludes the analysis of the sieving algorithm. As indicated earlier, the algorithm is surprisingly flexible and can be used to find not just the shortest vector. It can also provide the following guarantee (where we use the "natural" constants coming out of the algorithm):

Theorem 3.7. *Let $0 < \varepsilon \leq \frac{1}{2}$ and consider a lattice $\Lambda := \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$ and a symmetric convex body $K \subseteq \mathbb{R}^n$. Then there is a randomized algorithm with running time $(1/\varepsilon)^{\Theta(n)} \cdot \log(nM)$ (where $M := \max_{i=1,\ldots,n} \|\boldsymbol{b}_i\|_K$), which returns a random set Z with the following properties: (A) $Z \subseteq 10K \cap \Lambda$; (B) for each $\boldsymbol{v} \in 1.5K \cap \Lambda$ one has $\Pr[\exists \boldsymbol{z} \in Z : \|\boldsymbol{v} - \boldsymbol{z}\|_K \leq \varepsilon] \geq 1 - 2^{-n^{\omega(1)}}$.*

This statement can be used for an approximation algorithm for the Closest Vector problem.

Theorem 3.8. *Let $0 < \varepsilon \le \frac{1}{2}$. Given a lattice $\Lambda := \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$, a symmetric convex body $K \subseteq \mathbb{R}^n$ and a target vector $\boldsymbol{t}$, there is a randomized algorithm that with high probability finds a $(1+\varepsilon)$-approximate solution to $\min\{\|\boldsymbol{t}-\boldsymbol{x}\|_K : \boldsymbol{x} \in \Lambda\}$ in time $(1/\varepsilon)^{O(n)}$ times a polynomial in the encoding length of $\boldsymbol{B}$.*

We will leave the details of both claims as exercises. Similarly one can also find approximate ith successive minima:

Theorem 3.9 ([BN09])**.** *Let $0 < \varepsilon \le \frac{1}{2}$ and consider a lattice $\Lambda := \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$ and a symmetric convex body $K \subseteq \mathbb{R}^n$. Then in time $(1/\varepsilon)^{O(n)}$ times a polynomial in the encoding length of $\boldsymbol{B}$ one can find linearly independent vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \Lambda$ so that $\|\boldsymbol{v}_i\|_K \le (1+\varepsilon) \cdot \lambda_i(\Lambda, K)$ for all $i = 1, \ldots, n$.*

3.3 Exercises

Exercise 3.1.
Prove Theorem 3.7.

Exercise 3.2.
In this exercise, we want to explain how to prove Theorem 3.8. Fix $0 < \varepsilon \le \frac{1}{2}$ and consider a full rank lattice $\Lambda := \Lambda(\boldsymbol{B})$ with $\boldsymbol{B} \in \mathbb{R}^{n \times n}$, a symmetric convex body $K \subseteq \mathbb{R}^n$ and a target vector $\boldsymbol{t} \in \mathbb{R}^n$. Let $\boldsymbol{x}^* \in \Lambda$ be a minimizer to $\min\{\|\boldsymbol{t}-\boldsymbol{x}\|_K : \boldsymbol{x} \in \Lambda\}$. We assume that (after scaling) we have $\|\boldsymbol{t}-\boldsymbol{x}^*\|_K = 1$. We extend the setting by one dimension and define a $(n+1)$-dimensional lattice $\tilde{\Lambda} := \Lambda(\tilde{\boldsymbol{B}})$ given by $\tilde{\boldsymbol{B}} := \begin{pmatrix} \boldsymbol{B} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{pmatrix}$. We also define a symmetric convex body $\tilde{K} \subseteq \mathbb{R}^{n+1}$ so that $\|(\boldsymbol{x}, x^{(n+1)})\|_{\tilde{K}} := (1-\varepsilon)\|\boldsymbol{x}\|_K + \varepsilon|x^{(n+1)}|$. Apply Theorem 3.7 to $\tilde{\Lambda}$ and $\tilde{K}$ with parameter $\tilde{\varepsilon} := \frac{\varepsilon}{2}$. Show that from the random set $\tilde{Z}$ you can extract a $(1+2\varepsilon)$-approximation to the Closest vector problem (with high probability).

Exercise 3.3.
Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then for any $t > 0$, $|\Lambda \cap t \cdot \lambda_1(\Lambda, K) \cdot K| \le (2t+1)^n$.

Chapter 4

Banaszczyk's Transference Theorems

Recall that for a lattice Λ we denoted $\Lambda^* := \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z}\ \forall \boldsymbol{x} \in \Lambda\}$ as the *dual lattice*. Also recall that $\lambda_i(\Lambda)$ gives the minimum value so that there are i linearly independent vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i \in \Lambda$ with $\|\boldsymbol{v}_1\|_2, \ldots, \|\boldsymbol{v}_i\|_2 \leq \lambda_i(\Lambda)$. The goal for this chapter is to show a relation between short vectors in a lattice Λ and the dual lattice Λ^*. We will see in an exercise that using lattice basis reduction one can prove that always $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \leq 2^{O(n^2)}$. This is already a remarkable statement in the sense that knowing the length $\lambda_1(\Lambda)$ of a single vector in the primal lattice gives an upper bound on n vectors in the dual lattice. However, the bound is exponentially large and hence quite weak. Here, we prove the following:

Theorem 4.1 (Banaszczyk '93 [Ban93b])**.** *For any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\frac{1}{2} \leq \lambda_1(\Lambda) \cdot \mu(\Lambda^*) \leq n$.*

Then as promised we may infer the following[1]:

Corollary 4.2. *For any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ one has $1 \leq \lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \leq 2n$.*

Proof. The lower bound $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \geq 1$ follows because for any $\boldsymbol{x} \in \Lambda, \boldsymbol{y} \in \Lambda^*$ one has $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z}$. For the upper bound we use $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \leq 2\lambda_1(\lambda) \cdot \mu(\Lambda^*) \leq 2n$ using Lemma 1.45. □

Recall that Minkowski's theorem implies that $\lambda_1(\Lambda) \leq \sqrt{n} \cdot \det(\Lambda)^{1/n}$ and $\lambda_1(\Lambda^*) \leq \sqrt{n} \cdot \det(\Lambda^*)^{1/n}$ which then already gives the qualitatively weaker bound of $\lambda_1(\Lambda) \cdot \lambda_1(\Lambda^*) \leq n$. Also note that the upper bound in Cor 4.2 is asymptotically tight since by Theorem 1.47 there is a lattice Λ with $\lambda_i(\Lambda) \geq \Omega(\sqrt{n})$ and $\lambda_i(\Lambda^*) \geq \Omega(\sqrt{n})$ for all $i = 1, \ldots, n$.

[1] We note that the original paper of [Ban93b] shows a bound of n rather than $2n$.

The technique used for proving Banaszczyk's Theorem are fundamentally different from the techniques we have seen so far. They rely on *Fourier analysis* and the *Discrete Gaussian*. This chapter is a reproduction of the fantastic lecture notes of Regev [Reg09a] plus the original paper [Ban96] and plus some invaluable input from Stefan Steinerberger.

4.1 Fourier analysis

The idea behind Fourier analysis is to express a function f in a different basis (the Fourier basis). Many insights can be derived from this view that are hidden otherwise.

4.1.1 The Fourier Transform

A classical object of study in functional analysis is the Fourier transform.

Definition 4.3. For a function $f : \mathbb{R}^n \to \mathbb{C}$ with $\int_{\mathbb{R}^n} |f(\boldsymbol{x})|\, d\boldsymbol{x} < \infty$ we define the *Fourier transform* as the function $\hat{f} : \mathbb{R}^n \to \mathbb{C}$ with

$$\hat{f}(\boldsymbol{y}) := \int_{\mathbb{R}^n} f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle}\, d\boldsymbol{x}$$

It is not hard to see that some technical conditions on function f are needed. In the proof of Banaszczyk's Theorem we will apply the Fourier transform only to a family of functions f that are continuous and decay exponentially. Hence we will never run into any convergence problem. Occasionally we will call such functions "*nice enough*" without making this more formal.

A popular view is to consider the function $f(\boldsymbol{x})$ as a "signal" and the Fourier coefficient $\hat{f}(\boldsymbol{y})$ gives the amplitudes of the "frequency" $\boldsymbol{y}$ of that signal. There is also an explicit way to assemble the "frequencies" to recover the "signal".

Theorem 4.4 (Fourier Inversion Formula)**.** *For a continuous function $f : \mathbb{R}^n \to \mathbb{C}$ with $\int_{\mathbb{R}^n} |f(\boldsymbol{x})| d\boldsymbol{x} < \infty$ and $\int_{\mathbb{R}^n} |\hat{f}(\boldsymbol{y})| d\boldsymbol{y} < \infty$ one has*

$$f(\boldsymbol{x}) = \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle}\, d\boldsymbol{y} \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

Proof. Before we start with the main argument, we prove a claim that will be useful:

Claim. *For $\boldsymbol{v} \in \mathbb{R}^n$ and $\beta > 0$ one has $\int_{\mathbb{R}^n} e^{2\pi i \langle \boldsymbol{v}, \boldsymbol{y} \rangle - \beta \|\boldsymbol{y}\|_2^2} d\boldsymbol{y} = \exp(-\frac{\|\boldsymbol{v}\|_2^2 \pi^2}{\beta}) \cdot \left(\sqrt{\frac{\pi}{\beta}}\right)^n$.*

Proof of Claim. By symmetry we may assume that $\boldsymbol{v} = \alpha \boldsymbol{e}_1$ for some $\alpha \geq 0$. Then one can factor the integral into products as

$$\int_{\mathbb{R}^n} e^{2\pi i \langle \boldsymbol{v}, \boldsymbol{y} \rangle - \beta \|\boldsymbol{y}\|_2^2} d\boldsymbol{y} = \underbrace{\left(\int_{\mathbb{R}} e^{2\pi i \alpha y_1 - \beta y_1^2} dy_1 \right)}_{= \exp(-\frac{\pi^2 \alpha^2}{\beta}) \cdot \sqrt{\frac{\pi}{\beta}}} \cdot \underbrace{\left(\int_{\mathbb{R}} e^{-\beta z^2} dz \right)^{n-1}}_{= \sqrt{\frac{\pi}{\beta}}} = e^{-\alpha^2 \pi^2 / \beta} \cdot \left(\frac{\pi}{\beta} \right)^{n/2} \qquad \square$$

Now fix an $\boldsymbol{x} \in \mathbb{R}^n$. Naively one might try to start the proof by inserting the definition of $\hat{f}(\boldsymbol{y})$ into the right hand side of the expression $\int_{\mathbb{R}^n} \hat{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} d\boldsymbol{y}$. The problem is that the order of the emerging double integral cannot be swapped due to convergence issues. The trick is to first multiply the expression with a *dampening factor* $e^{-\varepsilon \|\boldsymbol{y}\|_2^2}$ to deal with the convergence issue. By the assumption on f and $\hat{f}$, the error that we make will go to 0 as $\varepsilon \to 0$ which then gives the claim. Please note that in the interest of time and space we do not spell out all the limits in the argument. We write

$$\begin{aligned}
\int_{\mathbb{R}^n} \hat{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} d\boldsymbol{y} &\overset{\varepsilon \to 0}{\approx} \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \cdot e^{-\varepsilon \|\boldsymbol{y}\|_2^2} d\boldsymbol{y} \\
&\overset{\text{Def } \hat{f}}{=} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(\boldsymbol{x}') \cdot e^{-2\pi i \langle \boldsymbol{x}', \boldsymbol{y} \rangle} d\boldsymbol{x}' \right) \cdot e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \varepsilon \|\boldsymbol{y}\|_2^2} d\boldsymbol{y} \\
&= \int_{\mathbb{R}^n} f(\boldsymbol{x}') \cdot \left(\int_{\mathbb{R}^n} e^{2\pi i \langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{y} \rangle - \varepsilon \|\boldsymbol{y}\|_2^2} d\boldsymbol{y} \right) d\boldsymbol{x}' \\
&\overset{\text{Claim I w. } \beta := \varepsilon, \boldsymbol{v} := \boldsymbol{x} - \boldsymbol{x}'}{=} \left(\sqrt{\frac{\pi}{\varepsilon}} \right)^n \int_{\mathbb{R}^n} f(\boldsymbol{x}') \cdot \exp\left(-\frac{\pi^2}{\varepsilon} \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 \right) d\boldsymbol{x}' \\
&\overset{f \text{ cont., } \varepsilon \to 0}{\approx} \left(\sqrt{\frac{\pi}{\varepsilon}} \right)^n f(\boldsymbol{x}) \int_{\mathbb{R}^n} \exp\left(-\frac{\pi^2}{\varepsilon} \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 \right) d\boldsymbol{x}' \\
&\overset{\text{shift}}{=} \left(\sqrt{\frac{\pi}{\varepsilon}} \right)^n f(\boldsymbol{x}) \int_{\mathbb{R}^n} \exp\left(-\frac{\pi^2}{\varepsilon} \|\boldsymbol{x}'\|_2^2 \right) d\boldsymbol{x}' \\
&\overset{\text{Claim I w. } \beta := \frac{\pi^2}{\varepsilon}, \boldsymbol{v} := \boldsymbol{0}}{=} \left(\sqrt{\frac{\pi}{\varepsilon}} \right)^n \cdot f(\boldsymbol{x}) \cdot \left(\sqrt{\frac{\pi}{(\pi^2/\varepsilon)}} \right)^n = f(\boldsymbol{x})
\end{aligned}$$

$\square$

Additionally we will need a variant of the Fourier transform that is custom tailored to lattices.

4.1.2 The Fourier series representation

In the following, let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice. We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is Λ-*periodic* if $f(\boldsymbol{x}) = f(\boldsymbol{x} + \boldsymbol{y})$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \Lambda$. In other words, shifting

the argument of f by a lattice point leaves the value invariant. For example, the function $\mathrm{dist}(\boldsymbol{x}, \Lambda) := \min\{\|\boldsymbol{x}-\boldsymbol{y}\|_2 : \boldsymbol{y} \in \Lambda\}$ that gives the distance of a point to the nearest lattice point is a natural Λ-periodic function. Now we want to define a discrete version of the Fourier transform:

Definition 4.5. Let $\Lambda = \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$ be a full-rank lattice and let $f : \mathbb{R}^n \to \mathbb{C}$ be a Λ-periodic function. Define the *Fourier series* $\tilde{f} : \Lambda^* \to \mathbb{C}$ as

$$\tilde{f}(\boldsymbol{y}) := \frac{1}{\det(\Lambda)} \cdot \int_{\mathcal{P}(\boldsymbol{B})} f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x} \quad \forall \boldsymbol{y} \in \Lambda^*.$$

Loosely speaking, the difference to the Fourier transform is that the integral is over a bounded region rather than $\mathbb{R}^n$ and that $\tilde{f}(\boldsymbol{y})$ is only defined for dual lattice vectors. Observe that the definition itself includes a concrete basis $\boldsymbol{B}$ for the lattice. We leave it as an exercise to prove that the values $\tilde{f}(\boldsymbol{y})$ do not depend on the chosen basis. It might be useful to keep in mind that one can equivalently write

$$\tilde{f}(\boldsymbol{y}) := \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{P}(\boldsymbol{B})} \left[f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \right] \quad \forall \boldsymbol{y} \in \Lambda^*$$

where $\boldsymbol{x} \sim \mathcal{P}(\boldsymbol{B})$ means we take a uniform sample from the fundamental region $\mathcal{P}(\boldsymbol{B})$.

Now we come to the discrete analogue of Theorem 4.4.

Theorem 4.6 (Fourier series representation)**.** *Let $\Lambda = \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$ be a full-rank lattice and let $f : \mathbb{R}^n \to \mathbb{C}$ be a nice enough Λ-periodic function and let $\tilde{f} : \Lambda^* \to \mathbb{C}$ be its Fourier series. Then*

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \Lambda^*} \tilde{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

Before we give a formal proof, we will explain what it means and why it makes sense. Fix a full-rank lattice $\Lambda := \Lambda(\boldsymbol{B})$ and consider the set $V := \{f : \mathbb{R}^n \to \mathbb{C} \mid f \text{ is } \Lambda\text{-periodic}\}$[2] which forms a vector space. We can define a *(complex) inner product* $\langle \cdot, \cdot \rangle_V$ for this vector space which for function $f, g \in V$ has the value

$$\langle f, g \rangle_V := \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{P}(\boldsymbol{B})} \left[f(\boldsymbol{x}) \cdot \overline{g(\boldsymbol{x})} \right] = \frac{1}{\det(\Lambda)} \int_{\mathcal{P}(\boldsymbol{B})} f(\boldsymbol{x}) \cdot \overline{g(\boldsymbol{x})} \, d\boldsymbol{x},$$

where $\overline{a + bi} := a - bi$ gives the *complex conjugate*. For a dual vector $\boldsymbol{y} \in \Lambda^*$, we define a function $\chi_{\boldsymbol{y}} : \mathbb{R}^n \to \mathbb{C}$ with $\chi_{\boldsymbol{y}}(\boldsymbol{x}) := \exp(2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle)$. Note that each function $\chi_{\boldsymbol{y}}$ is also Λ-periodic since for $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{z} \in \Lambda$ one has

$$\chi_{\boldsymbol{y}}(\boldsymbol{x} + \boldsymbol{z}) = \underbrace{\exp(2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle) \cdot \exp(2\pi i \underbrace{\langle \boldsymbol{z}, \boldsymbol{y} \rangle}_{\in \mathbb{Z}})}_{=1} = \chi_{\boldsymbol{y}}(\boldsymbol{x}).$$

[2] I suppose for a formal argument one should be adding some "niceness" conditions here

Next, for two dual vectors $\boldsymbol{y}, \boldsymbol{y}' \in \Lambda^*$ we can verify that the inner product of the corresponding functions is

$$\begin{aligned}
\langle \chi_{\boldsymbol{y}}, \chi_{\boldsymbol{y}'} \rangle_V &= \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{P}(\boldsymbol{B})} \left[\exp(2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle) \cdot \exp(-2\pi i \langle \boldsymbol{x}, \boldsymbol{y}' \rangle) \right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{P}(\boldsymbol{B})} \left[\exp(2\pi i \langle \boldsymbol{x}, \boldsymbol{y} - \boldsymbol{y}' \rangle) \right] \\
&= \underbrace{\prod_{j=1}^{n} \mathop{\mathbb{E}}_{\lambda_j \sim [0,1]} \left[\exp\left(2\pi i \lambda_j \underbrace{\langle \boldsymbol{B}^j, \boldsymbol{y} - \boldsymbol{y}' \rangle}_{\in \mathbb{Z}}\right) \right]}_{=1 \text{ if } \langle \boldsymbol{B}^j, \boldsymbol{y} - \boldsymbol{y}' \rangle = 0, = 0 \text{ o.w.}} = \begin{cases} 1 & \text{if } \boldsymbol{y} - \boldsymbol{y}' = \boldsymbol{0} \\ 0 & \text{if } \boldsymbol{y} - \boldsymbol{y}' \neq \boldsymbol{0} \end{cases}
\end{aligned}$$

Hence the family $\{\chi_{\boldsymbol{y}}\}_{\boldsymbol{y} \in \Lambda^*}$ is an infinite *orthonormal* set of functions contained in V. Next, we note that the inner product of a function $f \in V$ with one of the functions $\chi_{\boldsymbol{y}}$ for $\boldsymbol{y} \in \Lambda^*$ is indeed

$$\langle f, \chi_{\boldsymbol{y}} \rangle_V = \frac{1}{\det(\Lambda)} \int_{\mathcal{P}(\boldsymbol{B})} f(\boldsymbol{x}) \cdot \exp(-2\pi i \langle \boldsymbol{y}, \boldsymbol{x} \rangle) d\boldsymbol{x} \overset{\text{Def. } \tilde{f}}{=} \tilde{f}(\boldsymbol{y})$$

Now, we have not proven that $\{\chi_{\boldsymbol{y}}\}_{\boldsymbol{y} \in \Lambda^*}$ is indeed an orthonormal basis of V. But if we accept that as a fact, then the only way to write f as a linear combination in terms of that basis is

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \Lambda^*} \langle f, \chi_{\boldsymbol{y}} \rangle \cdot \chi_{\boldsymbol{y}}(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \Lambda^*} \tilde{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{y}, \boldsymbol{x} \rangle} \qquad \forall \boldsymbol{x} \in \mathbb{R}^n$$

And that identity is precisely the Fourier inversion formula!

4.1.3 The proof of the Fourier Series Representation

Now, we come to the formal proof of the *Fourier Series Representation Theorem* in form of Theorem 4.6. Here we follow the exposition due to Stefan Steinerberger. After applying a linear transformation, it will suffice to consider the lattice $\Lambda = \mathbb{Z}^n$ which conveniently has the dual lattice $\Lambda^* = \mathbb{Z}^n$. Recall that for a function $f : \mathbb{R}^n \to \mathbb{C}$ and $\boldsymbol{y} \in \mathbb{Z}^n$ we have the *Fourier series coefficient*

$$\tilde{f}(\boldsymbol{y}) := \int_{[0,1]^n} f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x}$$

It will be convenient to abbreviate a function $F : \mathbb{R}^n \to \mathbb{C}$ with

$$F(\boldsymbol{x}) := \sum_{\boldsymbol{y} \in \mathbb{Z}^n} \tilde{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle}$$

and the *truncation* $F_k : \mathbb{R}^n \to \mathbb{C}$ with

$$F_k(\boldsymbol{x}) := \sum_{\boldsymbol{y}\in\mathbb{Z}^n : \|\boldsymbol{y}\|_\infty \le k} \tilde{f}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{x},\boldsymbol{y}\rangle}$$

Then the main goal will be to prove:

Theorem 4.7. *If $f : \mathbb{R}^n \to \mathbb{C}$ is sufficiently often differentiable, then $f(\boldsymbol{x}) = F(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Here the vague term "sufficiently often differentiable" will mean that there is some (large enough) polynomial $p(n)$ so that for all $\boldsymbol{\alpha} \in \mathbb{Z}^n_{\ge 0}$ with $\|\boldsymbol{\alpha}\|_1 \le p(n)$, the derivative

$$\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f(\boldsymbol{x})$$

exists.

Upper bounding the Fourier Series Coefficients

Note that by definition one has $|\tilde{f}(\boldsymbol{y})| = |\mathbb{E}_{\boldsymbol{x}\sim\mathcal{P}(\boldsymbol{B})}[f(\boldsymbol{x})\cdot e^{-2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle}]| \le \mathbb{E}_{\boldsymbol{x}\sim\mathcal{P}(\boldsymbol{B})}[|f(\boldsymbol{x})|]$. However, one can prove that the Fourier series coefficients are indeed quickly decaying. Note that the constant $C_{f,\boldsymbol{\alpha}}$ in the upcoming bound will depend on the function f and on $\boldsymbol{\alpha}$ but crucially *not* on the Fourier coefficient $\boldsymbol{y}$.

Lemma 4.8. *Let $\boldsymbol{\alpha} \in \mathbb{Z}^n_{\ge 0}$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be $\|\boldsymbol{\alpha}\|_1$-times continuously differentiable. Then there is a constant $C_{f,\boldsymbol{\alpha}} > 0$ so that*

$$|\tilde{f}(\boldsymbol{y})| \le C_{f,\boldsymbol{\alpha}} \cdot \prod_{i=1}^n \frac{1}{|y_i|^{\alpha_i}} \quad \forall\, \boldsymbol{y} \in \mathbb{Z}^n \text{ with } y_i = 0 \Rightarrow \alpha_i = 0$$

Proof. Fix a vector $\boldsymbol{\alpha} \in \mathbb{Z}^n_{\ge 0}$ and an index i with $\alpha_i > 0$. Then for $\boldsymbol{y} \in \mathbb{Z}^n$ with $y_i \neq 0$ one we can write

$$\begin{aligned}
|\tilde{f}(\boldsymbol{y})| &= \left| \int_{[0,1]^n} f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x},\boldsymbol{y}\rangle}\, d\boldsymbol{x} \right| \\
&\overset{\frac{1}{a}(e^{ax})' = e^{ax}}{=} \left| \frac{1}{2\pi i y_i} \int_{[0,1]^n} f(\boldsymbol{x}) \cdot \left(\frac{\partial}{\partial x_i} e^{-2\pi i \langle \boldsymbol{x},\boldsymbol{y}\rangle} \right) d\boldsymbol{x} \right| \\
&\overset{\text{int. by parts}}{=} \left| \frac{1}{2\pi i y_i} \int_{[0,1]^n} \left(\frac{\partial}{\partial x_i} f(\boldsymbol{x}) \right) e^{-2\pi i \langle \boldsymbol{x},\boldsymbol{y}\rangle}\, d\boldsymbol{x} \right|
\end{aligned}$$

where we use integration by parts[3]. If we iterate this argument α_1-times for the

[3] Which is $\int_0^1 u(x)\,v'(x)\,dx = u(1)\,v(1) - u(0)\,v(0) - \int_0^1 u'(x)\,v(x)\,dx$. Note that if — as in our case — the functions u and v are 1-periodic, then $u(1)\,v(1) - u(0)\,v(0) = 0$.

first coordinate, α_2-times for the second coordinate etc, then we obtain the representation

$$\begin{aligned} |\tilde{f}(\boldsymbol{y})| &= \left|\Big(\prod_{i=1}^{n} \frac{1}{(2\pi i y_i)^{\alpha_i}}\Big) \cdot \int_{[0,1]^n} \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f(\boldsymbol{x}) e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x}\right| \\ &\leq \prod_{i=1}^{n} \frac{1}{|y_i|^{\alpha_i}} \cdot \underbrace{\sup_{\boldsymbol{x} \in [0,1]^n} \left|\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f(\boldsymbol{x})\right|}_{=:C_{f,\boldsymbol{\alpha}}}. \end{aligned}$$

as desired. Here the supremum exists as the derivative is continuous and $[0,1]^n$ is compact. □

We use this to obtain a rough but convenient upper bound on the size of the Fourier series coefficients:

Lemma 4.9. *Let $f : \mathbb{R}^n \to \mathbb{C}$ be sufficiently often differentiable. Then there is a constant $C_f > 0$ so that*

$$|\tilde{f}(\boldsymbol{y})| \leq C_f \cdot \Big(\frac{1}{\|\boldsymbol{y}\|_\infty}\Big)^{2n} \quad \forall \boldsymbol{y} \in \mathbb{Z}^n \setminus \{\boldsymbol{0}\}$$

Proof. Fix $\boldsymbol{y} \in \mathbb{Z}^n \setminus \{\boldsymbol{0}\}$. Let $i \in [n]$ be an index with $|y_i| = \|\boldsymbol{y}\|_\infty$. Choose $\boldsymbol{\alpha} := 2n \cdot \boldsymbol{e}_i$. Then by Lemma 4.8 one has

$$|\tilde{f}(\boldsymbol{y})| \leq C_{f,\boldsymbol{\alpha}} \cdot \frac{1}{|y_i|^{2n}} = C_{f,\boldsymbol{\alpha}} \cdot \Big(\frac{1}{\|\boldsymbol{y}\|_\infty}\Big)^{2n}$$

Note that the promised constant C_f will then be the maximum over the n constants for the different coordinates $i = 1, \ldots, n$. □

Convergence of F_k to F

Next, we will prove that the truncation F_k converges uniformly to F as $k \to \infty$ (though that itself does not yet tell us whether f and F are the same).

Lemma 4.10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be sufficiently often differentiable. Then $\lim_{k\to\infty} F_k(\boldsymbol{x}) = F(\boldsymbol{x})$ uniformly.*

Proof. For the sake of convinience we assume $n \geq 2$. We can bound the error

between the functions as

$$\begin{aligned}
|F_k(\boldsymbol{x}) - F(\boldsymbol{x})| &= \Big| \sum_{\boldsymbol{y}\in\mathbb{Z}^n:\|\boldsymbol{y}\|_\infty>k} \tilde{f}(\boldsymbol{y}) \cdot e^{2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle} \Big| \\
&\leq \sum_{\boldsymbol{y}\in\mathbb{Z}^n:\|\boldsymbol{y}\|_\infty>k} \underbrace{|\tilde{f}(\boldsymbol{y})|}_{\leq C_f/\|\boldsymbol{y}\|_\infty^{2n}} \cdot \underbrace{|e^{2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle}|}_{\leq 1} \\
&\overset{\text{Lem 4.9}}{\leq} C_f \sum_{\ell>k} \underbrace{|\{\boldsymbol{y}\in\mathbb{Z}^n : \|\boldsymbol{y}\|_\infty \leq \ell\}|}_{\leq (2\ell+1)^n} \cdot \frac{1}{\ell^{2n}} \\
&\leq C_f \sum_{\ell>k} \Big(\frac{2\ell+1}{\ell^2}\Big)^n \overset{k\to\infty, n\geq 2}{\longrightarrow} 0
\end{aligned}$$

□

Note that the $\boldsymbol{\alpha}$-th derivative of F_k is

$$\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} F_k(\boldsymbol{x})\, d\boldsymbol{x} = \sum_{\boldsymbol{y}\in\mathbb{Z}^n:\|\boldsymbol{y}\|_\infty\leq k} \Big(\prod_{i=1}^n (2\pi i y_i)^{\alpha_i} \Big) \cdot \tilde{f}(\boldsymbol{y}) \cdot e^{2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle}$$

Then repeating the arguments in Lemma 4.9 and Lemma 4.10 with a larger $\boldsymbol{\alpha}$, we can also force that the derivatives of F_k converge uniformly. We skip the details in order to not be repetitive:

Corollary 4.11. Let $f : \mathbb{R}^n \to \mathbb{R}$ be sufficiently often differentiable. Then F is twice differentiable and Lipschitz continuous.

The multidimensional Fejér kernel

A crucial ingredient of the main argument will be the existence of a function that has finite Fourier support but almost all the mass is concentrated around a single point.

Lemma 4.12. *There is a family of functions $\phi_k : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with the following properties:*

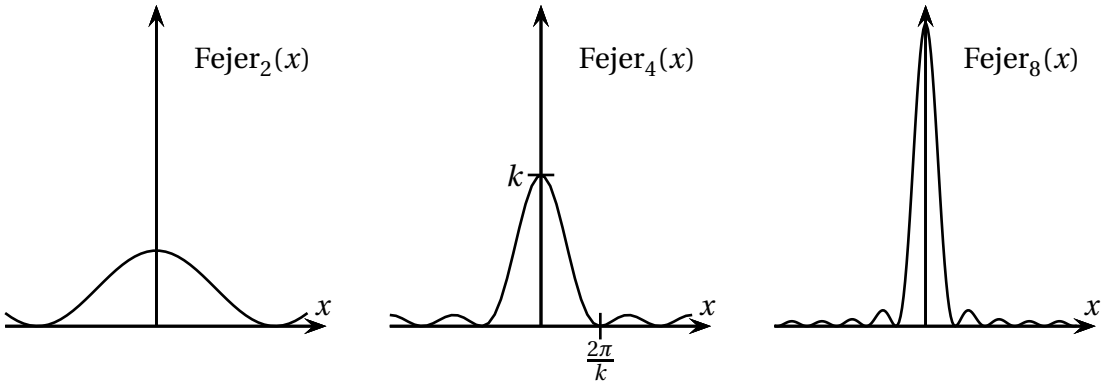
a) One has $\phi_k(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\mathbb{Z}^d:\|\boldsymbol{y}\|_\infty\leq k} a_{\boldsymbol{y}} \cdot e^{2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle}$ for some coefficients $a_{\boldsymbol{y}} \in \mathbb{R}$.

b) One has $\int_{\mathbb{R}^n} \phi_k(\boldsymbol{x})\, d\boldsymbol{x} = 1$.

c) There are $\varepsilon_k > 0$ so that $\lim_{k\to 0} \varepsilon_k = 0$ and $\int_{[-\varepsilon_k,\varepsilon_k]^n} \phi_k(\boldsymbol{x})\, d\boldsymbol{x} \geq 1 - \varepsilon_k$.

Proof. First we discuss the 1-dimensional case. We define the *Fejér kernel* as the function $\text{Fejer}_k : \mathbb{R} \to \mathbb{R}$ with

$$\text{Fejer}_k(x) = \sum_{|\ell| \le k} \left(1 - \frac{|\ell|}{k}\right) \cdot e^{i\ell x} = \frac{1}{k} \cdot \left(\frac{1-\cos(kx)}{1-\cos(x)}\right) = \frac{1}{k}\left(\frac{\sin(\frac{kx}{2})}{\sin\frac{x}{2}}\right)^2$$



For space reasons, we will not verify these identities here. Note that from the last characterization it becomes obvious that $\text{Fejer}_k(x) \ge 0$ for all $x \in \mathbb{R}$. Also without a proof we claim that indeed $\int_{\mathbb{R}} \text{Fejer}_k(x)dx = 1$ and for some $\delta_k > 0$ with $\lim_{k\to\infty} \delta_k = 0$ one has $\int_{[-\delta_k,\delta_k]} \text{Fejer}_k(x)\ dx \ge 1 - \delta_k$ meaning that most mass is concentrated around the origin. Now we choose $\phi_k : \mathbb{R}^n \to \mathbb{R}$ as the Cartesian product of the Fejér kernel:

$$\phi_k(\boldsymbol{x}) := \prod_{i=1}^{n} \text{Fejer}_k(x_i)$$

Note that in particular $\int_{[-\delta_k,\delta_k]^n} \phi_k(\boldsymbol{x})d\boldsymbol{x} \ge (1-\delta_k)^n$ and ϕ_k satisfies the claim. □

It will be convenient to obtain a shifted version of the kernel where the mass is concentrated around a point $\boldsymbol{x}_0$:

Lemma 4.13. *For $k \in \mathbb{N}$, $\boldsymbol{x}_0 \in \mathbb{R}^n$ there is a family of functions $\phi_{k,\boldsymbol{x}_0} : \mathbb{R}^n \to \mathbb{R}_{\ge 0}$ with the following properties:*

a) One has $\phi_{k,\boldsymbol{x}_0}(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\mathbb{Z}^d : \|\boldsymbol{y}\|_\infty \le k} b_{\boldsymbol{y}} \cdot e^{2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle}$ for some coefficients $b_{\boldsymbol{y}} \in \mathbb{C}$.

b) One has $\int_{\mathbb{R}^n} \phi_{k,\boldsymbol{x}_0}(\boldsymbol{x})d\boldsymbol{x} = 1$.

c) There are $\varepsilon_k > 0$ so that $\lim_{k\to 0} \varepsilon_k = 0$ and $\int_{\boldsymbol{x}_0+[-\varepsilon_k,\varepsilon_k]^n} \phi_k(\boldsymbol{x})d\boldsymbol{x} \ge 1 - \varepsilon_k$.

Proof. Simply define $\Phi_{k,\boldsymbol{x}_0}(\boldsymbol{x}) := \Phi_k(\boldsymbol{x}_0 + \boldsymbol{x})$. The only thing worth noting is that the individual summands change to $a_{\boldsymbol{y}} \cdot e^{2\pi i\langle \boldsymbol{x}_0+\boldsymbol{x},\boldsymbol{y}\rangle} = b_{\boldsymbol{y}} \cdot e^{2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle}$ where the shift makes the coefficients $b_{\boldsymbol{y}} := a_{\boldsymbol{y}} e^{2\pi i\langle \boldsymbol{x}_0,\boldsymbol{y}\rangle} \in \mathbb{C}$ complex. □

Completing the proof

We have almost everything to finish the proof. Recall the functions $\chi_{\boldsymbol{y}} : \mathbb{R}^n \to \mathbb{C}$ with $\chi_{\boldsymbol{y}}(\boldsymbol{x}) := e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle}$ that we have introduced earlier. In the main argument we would like to use *Plancherel's Theorem* which ideally should be saying that for $\mathbb{Z}^n$-periodic functions $f, g : \mathbb{R}^n \to \mathbb{C}$ one has

$$\int_{[0,1]^n} f(\boldsymbol{x}) \overline{g(\boldsymbol{x})} d\boldsymbol{x} = \sum_{\boldsymbol{y} \in \mathbb{Z}^n} \tilde{f}(\boldsymbol{y}) \tilde{g}(\boldsymbol{y}) \quad (*)$$

The expression on the left of $(*)$ is an inner product and the value of the inner product should be equal to the sum of the products of the coordinates in any orthonormal basis, which is the expression on the right hand side of $(*)$. The issue is we have not yet proven that $\{\chi_{\boldsymbol{y}}\}_{\boldsymbol{y} \in \mathbb{Z}^n}$ is a *basis* for the space of sufficiently differentiable functions. But this won't be necessary as long as we know that *one* of the functions f and g is spanned by finitely functions $\chi_{\boldsymbol{y}}$:

Lemma 4.14 (Semi-finite version of Plancherel's Theorem)**.** *Let $\mathcal{S} \subseteq \mathbb{Z}^n$ be finite and let $f, g : \mathbb{R}^n \to \mathbb{C}$ be two $\mathbb{Z}^n$-periodic functions where f is Lipschitz-continuous and g is of the form $g(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in \mathcal{S}} a_{\boldsymbol{y}} \chi_{\boldsymbol{y}}(\boldsymbol{x})$ for some $a_{\boldsymbol{y}} \in \mathbb{C}$. Then*

$$\int_{[0,1]^n} f(\boldsymbol{x}) \cdot \overline{g(\boldsymbol{x})} d\boldsymbol{x} = \sum_{\boldsymbol{y} \in \mathcal{S}} \tilde{f}(\boldsymbol{y}) \cdot a_{\boldsymbol{y}}$$

Proof. We can simply write

$$\int_{[0,1]^n} f(\boldsymbol{x}) \cdot \overline{g(\boldsymbol{x})} d\boldsymbol{x} \overset{\mathcal{S} \text{ finite}}{=} \sum_{\boldsymbol{y} \in \mathcal{S}} a_{\boldsymbol{y}} \underbrace{\int_{[0,1]^n} f(\boldsymbol{x}) \cdot \overline{\chi_{\boldsymbol{y}}(\boldsymbol{x})} d\boldsymbol{x}}_{= \tilde{f}(\boldsymbol{y})} = \sum_{\boldsymbol{y} \in \mathcal{S}} a_{\boldsymbol{y}} \tilde{f}(\boldsymbol{y})$$

□

Finally we prove the main result of this section:

Proof of Theorem 4.7. Recall that the goal is to prove that $f = F$. So suppose for the sake of contradiction that $f \neq F$, meaning there is a $\boldsymbol{x}_0 \in \mathbb{R}^n$ with $f(\boldsymbol{x}_0) \neq F(\boldsymbol{x}_0)$. By peridicity and continuity we may assume that $\boldsymbol{x}_0 \in (0,1)^n$.

On the one hand the difference $\boldsymbol{x} \mapsto f(\boldsymbol{x}) - F_k(\boldsymbol{x})$ is Lipschitz continuous and converges uniformly to $f(\boldsymbol{x}) - F(\boldsymbol{x})$ and so

$$\lim_{k \to \infty} \Big(\int_{[0,1]^n} \big(f(\boldsymbol{x}) - F_k(\boldsymbol{x}) \big) \cdot \phi_{k, \boldsymbol{x}_0}(\boldsymbol{x}) d\boldsymbol{x} \Big) = f(\boldsymbol{x}_0) - F(\boldsymbol{x}_0) \neq 0 \quad (**)$$

where we used the kernel function $\phi_{k,\boldsymbol{x}_0}$ from Lemma 4.13 which moves more and more of its mass around $\boldsymbol{x}_0$ as $k \to \infty$.

On the other hand for any $\boldsymbol{y} \in \mathbb{R}^n$ with $\|\boldsymbol{y}\|_\infty \leq k$ we have $\tilde{f}(\boldsymbol{y}) = \tilde{F}_k(\boldsymbol{y})$ and so $\widetilde{(f - F_k)}(\boldsymbol{y}) = 0$ by linearity. As $\phi_{k,\boldsymbol{x}_0}(\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\mathbb{Z}^n:\|\boldsymbol{y}\|_\infty \leq k} a_{\boldsymbol{y}} \chi_{\boldsymbol{y}}(\boldsymbol{x})$ has only finite Fourier support, we can apply Plancherel's Theorem from Lemma 4.14 and

$$\int_{[0,1]^n} (f(\boldsymbol{x}) - F_k(\boldsymbol{x})) \cdot \phi_{k,\boldsymbol{x}_0}(\boldsymbol{x}) d\boldsymbol{x} = \sum_{\boldsymbol{y}\in\mathbb{Z}^n:\|\boldsymbol{y}\|_\infty \leq k} \underbrace{(\tilde{f}(\boldsymbol{y}) - \tilde{F}_k(\boldsymbol{y}))}_{=0} \cdot a_{\boldsymbol{y}} = 0$$

Then certainly taking the limit for $k \to \infty$ will give 0, too. This is a contradiction to $(**)$. Hence $f = F$ as claimed. □

4.1.4 The Poisson Summation Formula

The *Poisson Summation Formula* shows that the sum of a function f over all lattice points is the same as the sum over the Fourier transform $\hat{f}$ over all points in the dual lattice (up to normalization factor). For a discrete set $A \subseteq \mathbb{R}^n$ we write $f(A) := \sum_{\boldsymbol{x}\in A} f(\boldsymbol{x})$. Recall that $\hat{f}$ is indeed the "regular" Fourier transform. First an auxiliary lemma:

Lemma 4.15. *For a nice enough function $f : \mathbb{R}^n \to \mathbb{C}$ and a full-rank lattice $\Lambda \subseteq \mathbb{R}^n$, the function $\varphi(\boldsymbol{x}) := \sum_{\boldsymbol{z}\in\Lambda} f(\boldsymbol{x} + \boldsymbol{z})$ is Λ-periodic and has Fourier series coefficients*

$$\tilde{\varphi}(\boldsymbol{y}) = \det(\Lambda^*) \cdot \hat{f}(\boldsymbol{y}) \quad \forall \boldsymbol{y} \in \Lambda^*$$

Proof. Note that φ is indeed Λ-periodic by construction. For any $\boldsymbol{y} \in \Lambda^*$ we have

$$\begin{aligned}
\tilde{\varphi}(\boldsymbol{y}) &\overset{\text{Def. Fourier series}}{=} \frac{1}{\det(\Lambda)} \int_{\mathcal{P}(\boldsymbol{B})} \varphi(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x} \qquad (*) \\
&\overset{\text{Def. } \varphi}{=} \frac{1}{\det(\Lambda)} \int_{\mathcal{P}(\boldsymbol{B})} \sum_{\boldsymbol{z}\in\Lambda} f(\boldsymbol{x} + \boldsymbol{z}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x} \\
&\overset{\text{swapping order}}{=} \frac{1}{\det(\Lambda)} \sum_{\boldsymbol{z}\in\Lambda} \int_{\mathcal{P}(\boldsymbol{B})} f(\boldsymbol{x} + \boldsymbol{z}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x} \\
&\overset{\boldsymbol{z}\in\Lambda \Rightarrow e^{-2\pi i \langle \boldsymbol{z}, \boldsymbol{y} \rangle} = 1}{=} \frac{1}{\det(\Lambda)} \sum_{\boldsymbol{z}\in\Lambda} \int_{\mathcal{P}(\boldsymbol{B})} f(\boldsymbol{x} + \boldsymbol{z}) \cdot e^{-2\pi i \langle \boldsymbol{x} + \boldsymbol{z}, \boldsymbol{y} \rangle} \, d\boldsymbol{x} \\
&\overset{\Lambda + \mathcal{P}(\boldsymbol{B}) = \mathbb{R}^n}{=} \underbrace{\frac{1}{\det(\Lambda)}}_{=\det(\Lambda^*)} \underbrace{\int_{\mathbb{R}^n} f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x}}_{=\hat{f}(\boldsymbol{y})} = \det(\Lambda^*) \cdot \hat{f}(\boldsymbol{y}).
\end{aligned}$$

□

Now we come to the Poisson Summation Formula itself:

Theorem 4.16 (Poisson Summation Formula for Lattices)**.** *For a nice enough function $f : \mathbb{R}^n \to \mathbb{C}$ and a full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ one has $f(\Lambda) = \det(\Lambda^*) \cdot \hat{f}(\Lambda^*)$.*

Proof. We define the function $\varphi(\boldsymbol{x}) := \sum_{\boldsymbol{z}\in\Lambda} f(\boldsymbol{x}+\boldsymbol{z})$. As φ is Λ-periodic, we may apply the Fourier Series Representation Theorem (Theorem 4.6). Then

$$\begin{aligned} f(\Lambda) &= \sum_{\boldsymbol{z}\in\Lambda} f(\boldsymbol{0}+\boldsymbol{z}) \overset{\text{Def. } \varphi}{=} \varphi(\boldsymbol{0}) \overset{\text{Fourier Series representation}}{=} \sum_{\boldsymbol{y}\in\Lambda^*} \tilde{\varphi}(\boldsymbol{y}) \cdot \underbrace{e^{2\pi i \langle \boldsymbol{0}, \boldsymbol{y}\rangle}}_{=1} \\ &\overset{\text{Lem 4.15}}{=} \sum_{\boldsymbol{y}\in\Lambda^*} \underbrace{\det(\Lambda^*)\hat{f}(\boldsymbol{y})}_{=\tilde{\varphi}(\boldsymbol{y})} = \det(\Lambda^*) \cdot \hat{f}(\Lambda^*). \end{aligned}$$

□

We leave the following extension as an exercise:

Corollary 4.17 (Shifted Poisson Summation Formula)**.** For a nice enough function $f : \mathbb{R}^n \to \mathbb{C}$, a full rank lattice $\Lambda \subseteq \mathbb{R}^n$ and a vector $\boldsymbol{s} \in \mathbb{R}^n$ one has $\sum_{\boldsymbol{x}\in\Lambda} \exp(2\pi i \langle \boldsymbol{x}, \boldsymbol{s}\rangle) \cdot f(\boldsymbol{x}) = \det(\Lambda^*) \cdot \hat{f}(\Lambda^* - \boldsymbol{s})$.

4.2 The discrete Gaussian

A crucial function in the proof of Banaszczyk's Theorem is the *discrete Gaussian* which for $s > 0$ is a function

$$\rho_s : \mathbb{R}^n \to \mathbb{R}_{\geq 0} \quad \text{with} \quad \rho_s(\boldsymbol{x}) := e^{-\pi \|\boldsymbol{x}/s\|_2^2} \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

In particular we will consider the sum $\rho_s(\Lambda)$ over a lattice. Intuitively, the quantity $\rho_s(\Lambda)$ counts the lattice points while the contribution of each point $\boldsymbol{x} \in \Lambda$ fades quickly if $\|\boldsymbol{x}\|_2$ is getting too large.

First we prove that for $s = 1$, the Fourier transform of the discrete Gaussian is again the discrete Gaussian:

Lemma 4.18. *For all $s > 0$, the Fourier transform of the discrete Gaussian is $\hat{\rho}_s(\boldsymbol{x}) = s^n \cdot \rho_{1/s}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Proof. Let us define the coordinate contribution as $g_s(x) := e^{-\pi \cdot (x/s)^2}$ for $x \in \mathbb{R}$. The following two facts can be proven using the proper integral manipulation skills. We will skip the proof here:

Fact I. *One has* $\hat{g}_s(y) = s \cdot e^{-\pi\cdot(ys)^2}$ *for all* $y \in \mathbb{R}$.

Fact II. *For any function* $f : \mathbb{R}^n \to \mathbb{R}$ *in product form* $f(\boldsymbol{x}) = \prod_{i=1}^n f_i(x_i)$ *one has* $\hat{f}(\boldsymbol{y}) = \prod_{i=1}^n \hat{f}_i(y_i)$ *for* $\boldsymbol{y} \in \mathbb{R}^n$.

Next, observe that the discrete Gaussian has a product structure and we can write

$$\rho_s(\boldsymbol{x}) = e^{-\pi\|\boldsymbol{x}/s\|_2^2} = \prod_{i=1}^n e^{-\pi(x_i/s)^2} = \prod_{i=1}^n g_s(x_i).$$

Then for $\boldsymbol{y} \in \mathbb{R}^n$ the Fourier transform is

$$\hat{\rho}_s(\boldsymbol{y}) \overset{\text{Fact II}}{=} \prod_{i=1}^n \hat{g}(y_i) \overset{\text{Fact I}}{=} \prod_{i=1}^n (s \cdot e^{-\pi\cdot(sy_i)^2}) = s^n \cdot e^{-\pi\|\frac{1}{1/s}\boldsymbol{y}\|_2^2} = s^n \cdot \rho_{1/s}(\boldsymbol{y}).$$

□

This implies a useful relation between the sum of the discrete Gaussian over a lattice and its dual lattice:

Corollary 4.19. For any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any $s > 0$ one has

$$\rho_s(\Lambda) = \det(\Lambda^*) \cdot s^n \cdot \rho_{1/s}(\Lambda^*).$$

Proof. Follows from Lemma 4.18 and the Poisson Summation Formula for Lattices from Lemma 4.16. □

It is not hard to exactly quantify the sum of the discrete Gaussian over a *shifted* lattice as well. Essentially if we shift the lattice by $\boldsymbol{u}$, then we need to "pull out" a factor of $e^{2\pi i\langle \boldsymbol{y},\boldsymbol{u}\rangle}$ for every summand.

Lemma 4.20. *For any full-rank lattice* $\Lambda \subseteq \mathbb{R}^n$*, any* $s > 0$ *and* $\boldsymbol{u} \in \mathbb{R}^n$ *one has*

$$\rho_s(\Lambda + \boldsymbol{u}) = \det(\Lambda^*) \cdot s^n \cdot \sum_{\boldsymbol{y}\in\Lambda^*} \rho_{1/s}(\boldsymbol{y}) \cdot e^{2\pi i\langle \boldsymbol{y},\boldsymbol{u}\rangle}.$$

Proof. We consider the function $f(\boldsymbol{x}) := \rho_s(\boldsymbol{x}+\boldsymbol{u})$ and write the Fourier transform as

$$\hat{f}(\boldsymbol{x}) = \int_{\mathbb{R}^n} \rho_s(\boldsymbol{y}+\boldsymbol{u}) e^{-2\pi i\langle \boldsymbol{x},\boldsymbol{y}\rangle}\, d\boldsymbol{y} = \int_{\mathbb{R}^n} \rho_s(\boldsymbol{y}) e^{-2\pi i\langle \boldsymbol{x},\boldsymbol{y}-\boldsymbol{u}\rangle}\, d\boldsymbol{y} = \hat{\rho}_s(\boldsymbol{x}) \cdot e^{2\pi i\langle \boldsymbol{x},\boldsymbol{u}\rangle} \qquad (*)$$

for $\boldsymbol{x} \in \mathbb{R}^n$. Then applying Lemma 4.16 to f gives

$$\begin{aligned}
\rho_s(\Lambda + \boldsymbol{u}) &= f(\Lambda) \overset{\text{Lemma 4.16}}{=} \det(\Lambda^*) \cdot \sum_{\boldsymbol{y}\in\Lambda^*} \hat{f}(\boldsymbol{y}) \\
&\overset{(*)}{=} \det(\Lambda^*) \cdot \sum_{\boldsymbol{y}\in\Lambda^*} \hat{\rho}_s(\boldsymbol{y}) \cdot e^{2\pi i\langle \boldsymbol{y},\boldsymbol{u}\rangle} \\
&\overset{\hat{\rho}_s(\boldsymbol{y})=s^n\rho_{1/s}(\boldsymbol{y})}{=} \det(\Lambda^*) \cdot s^n \cdot \sum_{\boldsymbol{y}\in\Lambda^*} \rho_{1/s}(\boldsymbol{y}) \cdot e^{2\pi i\langle \boldsymbol{y},\boldsymbol{u}\rangle}.
\end{aligned}$$

□

An important insight is that shifting the lattice can only *decrease* the sum over the discrete Gaussian:

Lemma 4.21. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice, $s > 0$ and let $\boldsymbol{u} \in \mathbb{R}^n$ be a shift. Then*

$$\rho_s(\Lambda + \boldsymbol{u}) \leq \rho_s(\Lambda).$$

Proof. We can estimate that

$$\begin{aligned}
\rho_s(\Lambda + \boldsymbol{u}) &\overset{\text{Lem 4.20}}{=} \left| \det(\Lambda^*) \cdot s^n \cdot \sum_{\boldsymbol{y} \in \Lambda^*} \rho_{1/s}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{y}, \boldsymbol{u} \rangle} \right| \\
&\leq \det(\Lambda^*) \cdot s^n \cdot \sum_{\boldsymbol{y} \in \Lambda^*} \rho_{1/s}(\boldsymbol{y}) \cdot \underbrace{\left| e^{2\pi i \langle \boldsymbol{y}, \boldsymbol{u} \rangle} \right|}_{\leq 1} \\
&\leq \det(\Lambda^*) \cdot s^n \cdot \underbrace{\sum_{\boldsymbol{y} \in \Lambda^*} \rho_{1/s}(\boldsymbol{y})}_{= \rho_{1/s}(\Lambda^*)} \\
&\overset{\text{Cor. 4.19}}{=} \rho_s(\Lambda)
\end{aligned}$$

which gives the claim. □

Increasing the scaling factor s for the discrete Gaussians means that the effective length of the lattice vectors is reduced and the sum $\rho_s(\Lambda)$ would increase. But we can limit the decrease and show that it cannot be more than exponential.

Lemma 4.22. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice and let $\boldsymbol{u} \in \mathbb{R}^n$ and $s \geq 1$. Then $\rho_s(\Lambda + \boldsymbol{u}) \leq s^n \cdot \rho_1(\Lambda)$.*

Proof. It suffices to prove that $\rho_s(\Lambda) \leq s^n \cdot \rho_1(\Lambda)$ — the general claim follows then from Lemma 4.21 where we showed that shifting can only decrease the sum of the discrete Gaussian. We will use the formula from Cor. 4.19 twice and obtain

$$\rho_s(\Lambda) \overset{\text{Cor. 4.19 for s}}{=} \det(\Lambda^*) \cdot s^n \cdot \sum_{\boldsymbol{y} \in \Lambda^*} \underbrace{\rho_{1/s}(\boldsymbol{y})}_{\leq \rho_1(\boldsymbol{y})} \leq \det(\Lambda^*) \cdot s^n \sum_{\boldsymbol{y} \in \Lambda^*} \rho_1(\boldsymbol{y}) \overset{\text{Cor. 4.19 for } s=1}{=} s^n \cdot \rho_1(\Lambda).$$

This gives the claim. □

One might be tempted to prove the claim for $\boldsymbol{u} = \boldsymbol{0}$ point-wise — however it is *not* true that $\frac{\rho_s(\boldsymbol{x})}{\rho_1(\boldsymbol{x})} \leq s^n$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and all $s \geq 1$. The claim only works amortized over all lattice points. As $\rho_r(\boldsymbol{x}) = \rho_1(\frac{\boldsymbol{x}}{r})$, Lemma 4.22 easily generalizes to the following:

Corollary 4.23. *for $s \geq 1$ and any $r > 0$ one has $\rho_{rs}(\Lambda + \boldsymbol{u}) \leq s^n \rho_r(\Lambda)$.*

4.3 The Proof of Banaszczyk's Theorem

Finally we come to the main part of proving the Transference Theorem of Banaszczyk's. First, if we consider the sum $\rho_1(\Lambda) = \sum_{\boldsymbol{x}\in\Lambda} e^{-\pi\|\boldsymbol{x}\|_2^2}$ then we know that always $\rho_1(\Lambda) \geq e^{-\pi\|\boldsymbol{0}\|_2^2} = 1$ due to the contribution of the origin. A useful insight is that the contribution of lattice points outside of a ball of radius $\sqrt{n}$ to $\rho_1(\Lambda)$ is always negligible:

Lemma 4.24. *For any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any vector $\boldsymbol{u} \in \mathbb{R}^n$ one has*

$$\rho_1\big((\Lambda + \boldsymbol{u}) \setminus \sqrt{n}B_2^n\big) \leq 2^{-n} \cdot \rho_1(\Lambda).$$

Proof. The proof basically uses that for long vectors the value $\rho_s(\boldsymbol{x})$ increases a lot with s, while we know that the overall sum can only grow with s^n. Then clearly long vectors could not have contributed much to the sum. More formally

$$\begin{aligned}
2^n \cdot \rho_1(\Lambda) &\overset{\text{Lem. 4.22}}{\geq} \rho_2(\Lambda + \boldsymbol{u}) \\
&\geq \rho_2\big((\Lambda + \boldsymbol{u}) \setminus \sqrt{n}B_2^n\big) \\
&= \sum_{\boldsymbol{y}\in\Lambda+\boldsymbol{u}:\|\boldsymbol{y}\|_2>\sqrt{n}} e^{-\pi\|\boldsymbol{y}/2\|_2^2} \\
&= \sum_{\boldsymbol{y}\in\Lambda+\boldsymbol{u}:\|\boldsymbol{y}\|_2>\sqrt{n}} \underbrace{e^{\frac{3}{4}\pi\|\boldsymbol{y}\|_2^2}}_{\geq 4^n} \cdot e^{-\pi\|\boldsymbol{y}\|_2^2} \\
&\geq 4^n \cdot \rho_1\big((\Lambda + \boldsymbol{u}) \setminus \sqrt{n}B_2^n\big)
\end{aligned}$$

Rearranging then gives the claim. □

Again, we state a generalization that follows from simple scaling for later reference:

Corollary 4.25. *For any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ any $r > 0$ and any vector $\boldsymbol{u} \in \mathbb{R}^n$ one has*

$$\rho_r\big((\Lambda + \boldsymbol{u}) \setminus r\sqrt{n}B_2^n\big) \leq 2^{-n} \cdot \rho_r(\Lambda).$$

An easy consequence is that in a lattice without short vectors, essentially all the Gaussian weight has to lie on the origin $\boldsymbol{0}$:

Lemma 4.26. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice with $\lambda_1(\Lambda) > \sqrt{n}$. Then $\rho_1(\Lambda \setminus \{\boldsymbol{0}\}) \leq 2 \cdot 2^{-n}$.*

Proof. Using the previous Lemma we have

$$\rho_1(\Lambda \setminus \{\boldsymbol{0}\}) \overset{\lambda_1(\Lambda)>\sqrt{n}}{=} \rho_1\big(\Lambda \setminus \sqrt{n}B_2^n\big) \overset{\text{Lem. 4.24}}{\leq} 2^{-n} \cdot \rho_1(\Lambda) = 2^{-n} \cdot \big(\underbrace{\rho_1(\boldsymbol{0})}_{=1} + \rho_1(\Lambda \setminus \{\boldsymbol{0}\})\big)$$

Rearranging for $\rho_1(\Lambda \setminus \{\mathbf{0}\})$ then gives the claim. □

We can prove that $\rho_1(\Lambda + \boldsymbol{u}) \approx \det(\Lambda^*)$ if $\rho_1(\Lambda^* \setminus \{\mathbf{0}\}) \ll 1$. For later reference we prove the statement in more generality than we need at the moment.

Lemma 4.27. *For any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any $s > 0$ and $\boldsymbol{u} \in \mathbb{R}^n$ one has*

$$|\rho_s(\Lambda + \boldsymbol{u}) - s^n \det(\Lambda^*)| \leq s^n \det(\Lambda^*) \cdot \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})$$

Proof. We write

$$\begin{aligned} |\rho_1(\Lambda + \boldsymbol{u}) - s^n \det(\Lambda^*)| &\overset{\text{Lem 4.20}}{=} s^n \det(\Lambda^*) \cdot \left| \Big(\sum_{\boldsymbol{y} \in \Lambda^*} \rho_{1/s}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{y}, \boldsymbol{u} \rangle} \Big) - 1 \right| \\ &\leq s^n \det(\Lambda^*) \sum_{\boldsymbol{y} \in \Lambda^* \setminus \{\mathbf{0}\}} \rho_{1/s}(\boldsymbol{y}) \cdot \underbrace{|e^{2\pi i \langle \boldsymbol{y}, \boldsymbol{u} \rangle}|}_{\leq 1} \\ &\leq s^n \det(\Lambda^*) \cdot \rho_{1/s}(\Lambda^* \setminus \mathbf{0}) \end{aligned}$$

□

An immediate consequence is as follows:

Corollary 4.28. *For any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any $s > 0$ and $\boldsymbol{u} \in \mathbb{R}^n$ one has*

$$1 \geq \frac{\rho_s(\Lambda + \boldsymbol{u})}{\rho_s(\Lambda)} \geq \frac{1 - \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})}{1 + \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})}$$

Proof. We know the upper bound already from Cor 4.23. The lower bound follows from

$$\frac{\rho_s(\Lambda + \boldsymbol{u})}{\rho_s(\Lambda)} \overset{\text{Lem 4.27}}{\geq} \frac{s^n \det(\Lambda^*) \cdot (1 - \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}))}{s^n \det(\Lambda^*) \cdot (1 + \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}))} = \frac{1 - \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})}{1 + \rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\})}$$

□

The next lemma gives one crucial insight: if the lattice Λ has no vector of length $\sqrt{n}$ or less, then the sum $\rho_1(\Lambda^* + \boldsymbol{u})$ over the shifted dual lattice does only marginally depend on the shift $\boldsymbol{u}$. This will then quickly imply that the dual lattice has no large "holes" and the covering radius has to be small.

Lemma 4.29. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice with $\lambda_1(\Lambda) > \sqrt{n}$. Then for any vector $\boldsymbol{u} \in \mathbb{R}^n$ one has*

$$\rho_1(\Lambda^* + \boldsymbol{u}) = (1 \pm 2 \cdot 2^{-n}) \cdot \det(\Lambda).$$

Proof. We estimate that

$$|\rho_1(\Lambda^* + \boldsymbol{u}) - \det(\Lambda)| \overset{\text{Lem 4.27}}{\leq} \det(\Lambda) \cdot \underbrace{\rho_1(\Lambda \setminus \{\mathbf{0}\})}_{\leq 2\cdot 2^{-n} \text{ by Lem. 4.26}} \leq \det(\Lambda) \cdot 2 \cdot 2^{-n}$$

□

Now we can prove Banaszczyk's result: we will show that for any lattice Λ with no non-zero vector of length at most $\sqrt{n}$, the covering radius of the dual lattice is bounded by $\sqrt{n}$.

Theorem 4.30. *For any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\lambda_1(\Lambda) \cdot \mu(\Lambda^*) \leq n$.*

Proof. We assume $n \geq 2$. After scaling the lattice appropriately it suffices to assume $\lambda_1(\Lambda) > \sqrt{n}$ and $\mu(\Lambda^*) > \sqrt{n}$ and bring this to a contradiction. From Lemma 4.29 we know that for a lattice Λ with $\lambda_1(\Lambda) > \sqrt{n}$, shifting the dual lattice has little effect on the sum of the discrete Gaussian; applying Lemma 4.29 twice gives that for any $\boldsymbol{u} \in \mathbb{R}^n$ one has

$$\frac{\rho_1(\Lambda^* - \boldsymbol{u})}{\rho_1(\Lambda^*)} \geq \frac{1 - 2 \cdot 2^{-n}}{1 + 2 \cdot 2^{-n}} \geq \frac{1}{3} \quad (*)$$

Now, fix a vector $\boldsymbol{u} \in \mathbb{R}^n$ attaining the covering radius for the dual lattice, that means $\Lambda^* \cap (\boldsymbol{u} + \sqrt{n}B_2^n) = \emptyset$ (which is equivalent to $(\Lambda^* - \boldsymbol{u}) \cap \sqrt{n}B_2^n = \emptyset$. Then

$$\frac{1}{3}\rho_1(\Lambda^*) \overset{(*)}{\leq} \rho_1(\Lambda^* - \boldsymbol{u}) \overset{\Lambda^* \cap (\boldsymbol{u} + \sqrt{n}B_2^n) = \emptyset}{=} \rho_1((\Lambda^* - \boldsymbol{u}) \setminus \sqrt{n}B_2^n) \overset{\text{Lem 4.24}}{<} 2^{-n}\rho_1(\Lambda^*)$$

which is a contradiction for $n \geq 2$. □

4.4 The Transference Theorem for arbitrary symmetric convex bodies

The goal for this section is to present a transference bound for arbitrary symmetric convex bodies rather than Euclidean balls. First we need to generalize some of the introduced notation. For a symmetric convex body $K \subseteq \mathbb{R}^n$ and a full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ we define the *ith successive minimum with respect to norm $\|\cdot\|_K$* as

$$\lambda_i(\Lambda, K) := \min\left\{r \geq 0 \mid \dim(\operatorname{span}(\Lambda \cap rK)) \geq i\right\}$$

As before, we will be particularly interested in the length of the *shortest vector* in norm $\|\cdot\|_K$ which is $\lambda_1(\Lambda, K) = \min\{\|\boldsymbol{x}\|_K : \boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}\}$. For a convex body $K \subseteq \mathbb{R}^n$ we define the *covering radius with respect to* K as

$$\mu(\Lambda, K) = \min\left\{r \geq 0 \mid \forall \boldsymbol{u} \in \mathbb{R}^n : (\boldsymbol{u} + rK) \cap \Lambda \neq \emptyset\right\}$$

Note that this definition also makes sense if K is not symmetric. But we will not need that level of generality until the next chapter. Of course, if K is symmetric then we can also write

$$\mu(\Lambda, K) = \max_{\boldsymbol{u} \in \mathbb{R}^n} \min_{\boldsymbol{x} \in \Lambda} \|\boldsymbol{x} - \boldsymbol{u}\|_K.$$

We denote $K^\circ := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 1 \ \forall \boldsymbol{y} \in K\}$ as the *polar* of K. If K is any convex body with $\boldsymbol{0} \in \text{int}(K)$, then one can show that $(K^\circ)^\circ = K$. Also note that $\|\cdot\|_{K^\circ}$ is the *dual norm* to $\|\cdot\|_K$ meaning that $\|\boldsymbol{x}\|_{K^\circ} = \sup\{\langle \boldsymbol{y}, \boldsymbol{x} \rangle : \|\boldsymbol{y}\|_K \leq 1\}$. The Euclidean ball is the only self-polar body, that means $(B_2^n)^\circ = B_2^n$. Another example is that $(B_\infty^n)^\circ = B_1^n$.

By John's Theorem, any symmetric convex body can be approximated within a $\sqrt{n}$ factor with an ellipsoid and so the result from Theorem 4.30 implies that $1 \leq \lambda_1(\Lambda, K) \cdot \mu(\Lambda^*, K^\circ) \leq O(n^{3/2})$. The goal for this section is a more powerful transference theorem for general norms that only loses a logarithmic factor compared to the Euclidean norm:

Theorem 4.31 (Banaszczyk 1996)**.** *For full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any symmetric convex body $K \subseteq \mathbb{R}^n$ one has $1 \leq \lambda_1(\Lambda, K) \cdot \mu(\Lambda^*, K^\circ) \leq O(n \log(n))$.*

4.4.1 Fourier analysis with arbitrary symmetric convex bodies

In this subsection we will generalize a few facts on Fourier analysis from earlier to deal with arbitrary symmetric convex bodies. Additionally we will develop some new arguments. First, for a symmetric convex body $K \subseteq \mathbb{R}^n$ we define

$$\beta(K) := \sup_{\substack{\Lambda \subseteq \mathbb{R}^n \\ \text{lattice}}} \sup_{\boldsymbol{u} \in \mathbb{R}^n} \frac{\rho_1((\boldsymbol{u} + \Lambda) \setminus K)}{\rho_1(\Lambda)}$$

Equivalently, $\beta(K)$ is the smallest number so that for any lattice Λ and any vector $\boldsymbol{u}$ one has $\rho_1((\boldsymbol{u} + \Lambda) \setminus K) \leq \beta(K) \cdot \rho_1(\Lambda)$. Note that always $0 < \beta(K) < 1$. In some sense a small value $\beta(K)$ — the threshold of $\beta(K) \leq \frac{1}{4}$ will work for our purposes — means that the set K is *big* and any discrete Gaussian will place a significant fraction of weight inside K. For example in Lemma 4.24 we proved that $\beta(\sqrt{n}B_2^n) \leq 2^{-n}$ meaning that $\sqrt{n}B_2^n$ satisfies our notion of a "big" set. First we generalize Lemma 4.26 and Lemma 4.29 to work with the new definition of a "big" set:

Lemma 4.32. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $K \subseteq \mathbb{R}^n$ be a symmetric convex body with $\lambda_1(\Lambda, K) > 1$ and let $\boldsymbol{u} \in \mathbb{R}^n$. Then the following holds:*

(a) One has $\rho_1(\Lambda \setminus \{\mathbf{0}\}) \leq \frac{\beta(K)}{1-\beta(K)}$.

(b) One has $|\rho_1(\Lambda^* + \boldsymbol{u}) - \det(\Lambda)| \leq \frac{\beta(K)}{1-\beta(K)} \cdot \det(\Lambda)$.

(c) One has $1 \geq \frac{\rho_1(\Lambda^*+\boldsymbol{u})}{\rho_1(\Lambda^*)} \geq 1 - 2\beta(K)$.

Proof. First we prove (a). As $\lambda_1(\Lambda, K) > 1$ we know that $\Lambda \cap K = \{\mathbf{0}\}$. Then

$$\rho_1(\Lambda \setminus \{\mathbf{0}\}) \overset{\Lambda\cap K=\{\mathbf{0}\}}{=} \rho_1(\Lambda \setminus K) \leq \beta(K) \cdot \rho_1(\Lambda) = \beta(K) \cdot (\underbrace{\rho_1(\mathbf{0})}_{=1} + \rho_1(\Lambda \setminus \{\mathbf{0}\}))$$

Rearranging gives (a). For (b), we have

$$|\rho_1(\Lambda^* + \boldsymbol{u}) - \det(\Lambda)| \overset{\text{Lem } 4.27}{\leq} \det(\Lambda) \cdot \rho_1(\Lambda \setminus \{\mathbf{0}\}) \overset{\text{Lem ??}}{\leq} \det(\Lambda) \cdot \frac{\beta(K)}{1-\beta(K)}$$

To get (c), we apply (b) twice and have

$$\frac{\rho_1(\Lambda^* + \boldsymbol{u})}{\rho_1(\Lambda^*)} \geq \frac{(1 - \frac{\beta(K)}{1-\beta(K)}) \cdot \det(\Lambda)}{(1 + \frac{\beta(K)}{1-\beta(K)}) \cdot \det(\Lambda)} = 1 - 2\beta(K)$$

□

Finally, we generalize a part of the argument used in Theorem 4.30. It may appear a little odd that we have two sets K and Q appear. In out later application Q will simply be chosen as the properly scaled polar of K. But our formulation is more general and we won't have to worry yet about the correct scaling of the polar.

Lemma 4.33. *Let $K, Q \subseteq \mathbb{R}^n$ be symmetric convex bodies with $\beta(K), \beta(Q) \leq \frac{1}{4}$. Then for any lattice $\Lambda \subseteq \mathbb{R}^n$ one has*

$$\lambda_1(\Lambda, K) \cdot \mu(\Lambda^*, Q) \leq 1$$

Proof. Suppose for the sake of contradiction that there is a lattice Λ with $\lambda_1(\Lambda, K) \cdot \mu(\Lambda^*, Q) > 1$. The left hand side is invariant under scaling of Λ so we may assume that $\lambda_1(\Lambda, K) > 1$ and $\mu(\Lambda^*, Q) > 1$. Since we have the lower bound $\mu(\Lambda^*, Q) > 1$

on the covering radius we know there is a vector $\boldsymbol{u} \in \mathbb{R}^n$ so that $(\boldsymbol{u}+\Lambda^*) \cap Q = \emptyset$. By assumption this vector satisfies

$$\frac{\rho_1((\boldsymbol{u}+\Lambda^*) \setminus Q)}{\rho_1(\Lambda^*)} \leq \beta(Q) \leq \frac{1}{4} \qquad (*)$$

On the other hand we have $\lambda_1(\Lambda, K) > 1$ which satisfies the assumption of Lemma 4.32.(c) and we have

$$\frac{\rho_1(\Lambda^* + \boldsymbol{u})}{\rho_1(\Lambda^*)} \geq 1 - 2\beta(K) \geq \frac{1}{2} \qquad (**)$$

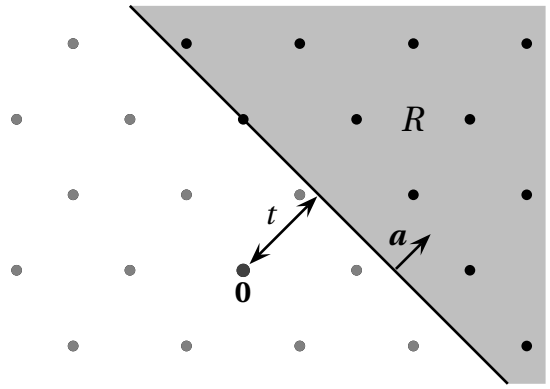
We combine both facts together and obtain:

$$\frac{1}{2} \overset{(*)}{\leq} \frac{\rho_1(\boldsymbol{u}+\Lambda^*)}{\rho_1(\Lambda^*)} \overset{(\boldsymbol{u}+\Lambda^*)\cap Q=\emptyset}{=} \frac{\rho_1((\boldsymbol{u}+\Lambda^*) \setminus Q)}{\rho_1(\Lambda^*)} \leq \frac{1}{4}$$

This is a contradiction. □

We have most of the estimates related to the discrete Gaussian in place but we need one estimate that is new.

Lemma 4.34. *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice. For a unit vector $\boldsymbol{a} \in S^{n-1}$ and $t \geq 0$ consider the region $R := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x} \rangle \geq t\}$. Then for any $\boldsymbol{u} \in \mathbb{R}^n$ one has $\rho_1((\boldsymbol{u}+\Lambda) \cap R) \leq e^{-\pi t^2} \rho_1(\Lambda)$.*



Proof. We use the standard trick in measure concentration to consider an exponentially weighted sum. We obtain

$$\begin{aligned} e^{2\pi t^2} \cdot \rho_1((\boldsymbol{u}+\Lambda) \cap R) &\overset{(*)}{\leq} \sum_{\boldsymbol{x} \in \boldsymbol{u}+\Lambda} \rho_1(\boldsymbol{x}) \cdot e^{2\pi t \langle \boldsymbol{a}, \boldsymbol{x} \rangle} \\ &= \sum_{\boldsymbol{x} \in \boldsymbol{u}+\Lambda} \exp\left(-\pi \|\boldsymbol{x}\|_2^2 + 2\pi t \langle \boldsymbol{a}, \boldsymbol{x} \rangle\right) \\ &= \sum_{\boldsymbol{x} \in \boldsymbol{u}+\Lambda} \exp\left(\pi \|t\boldsymbol{a}\|_2^2 - \pi \|\boldsymbol{x} - t\boldsymbol{a}\|_2^2\right) \\ &= e^{\pi t^2} \cdot \rho_1(\boldsymbol{u} - t\boldsymbol{a} + \Lambda) \overset{(**)}{\leq} e^{\pi t^2} \cdot \rho_1(\Lambda) \end{aligned}$$

Here we use in $(*)$ that any point $\boldsymbol{x} \in \boldsymbol{u} + \Lambda$ on the right hand side that is in R contributes $e^{2\pi t\langle \boldsymbol{a}, \boldsymbol{x}\rangle} \geq e^{2\pi t^2}$ to the left hand side. In $(**)$ we use Lemma 4.21 telling us that central shifts maximize the weight of the discrete Gaussian. Rearranging gives the claim. □

4.4.2 Properties of the discrete Gaussian

We want to make a small excursion to connect in particular the last proven lemma with a very versatile concept in probability theory that deals with random variables that have *Gaussian-type tails*. There are several ways how one could define "having a Gaussian-type tail" but luckily all candidate choices are equivalent. It will also be cleaner to focus on *mean-zero* random variables.

Lemma 4.35 (Conditions of Sub-Gaussian tails)**.** *Let $X \in \mathbb{R}$ be a random variable with $\mathbb{E}[X] = 0$. The following statements are equivalent in the sense that if condition i holds with $s_i > 0$ then there is an $s_j \in [\frac{s_i}{C}, C s_i]$ so that also condition j holds where $C > 0$ is a universal constant.*

- *Condition 1: One has $\Pr[|X| \geq t] \leq 2\exp(-t^2/s_1^2)$ for all $t \geq 0$.*
- *Condition 2: One has $\mathbb{E}[|X|^p]^{1/p} \leq s_2\sqrt{p}$ for all $p \geq 1$.*
- *Condition 3: One has $\mathbb{E}[\exp(X^2/s_3^2)] \leq 2$.*
- *Condition 4: One has $\mathbb{E}[\exp(\lambda X)] \leq \exp(s_4^2\lambda^2)$ for all $\lambda \in \mathbb{R}$.*

We refer to the wonderful exposition in Vershynin [Ver19] for details. So it is natural to pick one of the above conditions as a parameter determining the concentration behavior of a random variable:

Definition 4.36. Let $X \in \mathbb{R}$ be a random variable with $\mathbb{E}[X] = 0$. We define the *sub-gaussian norm* as as

$$\|X\|_{\psi_2} := \inf\left\{s > 0 : \mathbb{E}\left[\exp\left(\frac{X^2}{s^2}\right)\right] \leq 2\right\}$$

While it is not at all obvious, $\|\cdot\|_{\psi_2}$ is indeed a norm on the space of mean-zero random variables, i.e. $\|tX\|_{\psi_2} = |t| \cdot \|X\|_{\psi_2}$ for $t \in \mathbb{R}$ and $\|X+Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2}$ for any jointly distributed mean-zero random variables (which may even be dependent). Moreover one can prove the following properties:

Lemma 4.37. *In the following let $X_1, \ldots, X_N$ be jointly distributed mean-zero random variables.*

(i) *One has* $\mathbb{E}[\max\{|X_1|,\dots,|X_N|\}] \le O(\sqrt{\log(N)}) \cdot \max\{\|X_i\|_{\psi_2} : i \in [N]\}$

(ii) *If* $X_1,\dots,X_N$ *are independent then* $\|X_1+\dots+X_N\|_{\psi_2} \le C \cdot (\sum_{i=1}^N \|X_i\|_{\psi_2}^2)^{1/2}$ *for a universal constant* $C>0$.

The proof is not difficult when using the different conditions of Lemma 4.35. Again we refer to Vershynin [Ver19] for details.

Now back to lattices. For a full-rank lattice $\Lambda \subseteq \mathbb{R}^n$, we define the *discrete Gaussian* as the distribution $\mathcal{D}_1(\Lambda)$ that yields each vector $\boldsymbol{x} \in \Lambda$ with probability $\frac{\rho_1(\boldsymbol{x})}{\rho_1(\Lambda)}$. Despite its discrete character, we can still compare $\mathcal{D}_1(\Lambda)$ to a standard normal distribution. In fact, the Lemma 4.34 we have proven earlier implies the following:

Lemma 4.38 (Subgaussianity of Discrete Gaussian)**.** *Let* $\Lambda \subseteq \mathbb{R}^n$ *be an arbitrary full rank lattice. Then for any direction* $\boldsymbol{\theta} \in S^{n-1}$*, the random variable* $\langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$ *with* $\boldsymbol{x} \sim \mathcal{D}_1(\Lambda)$ *satisfies* $\| \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle \|_{\psi_2} \le O(1)$.

4.4.3 Convex Geometry

Next, we need to find a way to prove that any symmetric convex body K can be scaled so that both K and K° are "big". Let $N(\mathbf{0}, \boldsymbol{I}_n)$ be distribution of the n-dimensional Gaussian with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_n$. Recall that one can generate a sample $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)$ by independently sampling the coordinates $x_1,\dots,x_n \sim N(0,1)$. We introduce a well studied quantity in convex geometry which is the *ℓ-value*

$$\ell_K := \mathop{\mathbb{E}}_{\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)} \left[\|\boldsymbol{x}\|_K^2 \right]^{1/2}$$

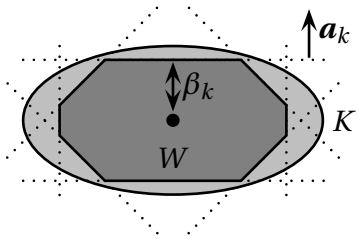
Intuitively, ℓ_K gives a notion of "average thinness" of a symmetric convex body. For example for scalars of the Euclidean ball we have $\ell_{rB_2^n} = \mathbb{E}_{\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)}[\|\boldsymbol{x}\|_{rB_2^n}^2]^{1/2} = \frac{1}{r}\mathbb{E}_{\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)}[\|\boldsymbol{x}\|_2^2]^{1/2} = \frac{\sqrt{n}}{r}$. It turns out to be very useful to consider the product $\ell_K \cdot \ell_{K^\circ}$. In the example of the scaled Euclidean ball we have $\ell_{rB_2^n} \cdot \ell_{(rB_2^n)^\circ} = \frac{\sqrt{n}}{r} \cdot r\sqrt{n} = n$ for all $r > 0$. On the other hand, it may happen that a body K is very thin in some direction and very long in another one and so the product $\ell_K \cdot \ell_{K^\circ}$ can be made as large as desired. A simple construction to see this would be the 2-dimensional ellipsoid $K := \{\binom{My_1}{y_2/M} : \boldsymbol{y} \in B_2^n\}$ with M large where $\ell_K = \Theta(M)$ and $\ell_{K^\circ} = \Theta(M)$. However, one of the deepest and most important results in convex geometry is that any symmetric convex body can be linearly transformed to that the corresponding product of ℓ-values is almost as small as it is for the Euclidean ball. The result is a combination of work by Lewis [Lew79], Pisier [Pis80] and Figiel, Tomczak-Jaegerman [FTJ79].

Theorem 4.39 ($\ell\ell^\circ$-estimate)**.** *For any symmetric symmetric convex body $K \subseteq \mathbb{R}^n$ there is an invertible linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $\ell_{T(K)} \cdot \ell_{(T(K))^\circ} \leq O(n \log n)$.*

We refer to the textbook of Artstein-Avidan, Giannopoulos and Milman[AAGM15] for details.

The next step will be to prove that a symmetric convex body K with small ℓ_K-value also has a small β-value. If we think of K as the intersection of halfspaces then we can in principle use Lemma 4.34 to bound the weight of a discrete Gaussian outside of K. The problem is that in order for that argument to be effective, K would have to be defined by few halfspaces where the allowed number depends on their distance to the origin. One may not think that this approach is feasible — after all there are bodies like $K = B_2^n$ defined by infinitely many halfspaces. Yet, any symmetric convex body allows an inner approximation of the following form:

Theorem 4.40. *For any symmetric convex body $K \subseteq \mathbb{R}^n$ there is a sequence $\{\boldsymbol{a}_k\}_{k\in\mathbb{N}} \subseteq S^{n-1}$ of unit vectors and values $\beta_k := \frac{C\sqrt{\log(2k)}}{\ell_K}$ so that the symmetric convex body $W := \{\boldsymbol{x} \in \mathbb{R}^n \mid |\langle \boldsymbol{a}_k, \boldsymbol{x}\rangle| \leq \beta_k \ \forall k \in \mathbb{N}\}$ satisfies $W \subseteq K$. Here $C > 0$ is a universal constant.*



This statement is a consequence of the Talagrand's *Majorizing Measure Theorem* [Tal87] which shows that the expected supremum of any Gaussian process is up to a constant characterized by a simpler quantity called the γ_2*-function*. For an excellent exposition on this method and more applications of it, we recommend the recent textbook of Vershynin [Ver19].

Lemma 4.41. *For any $\varepsilon > 0$ there is some $\delta > 0$ so that the following holds: For any symmetric convex body $K \subseteq \mathbb{R}^n$ with $\ell_K \leq \delta$ one has $\beta(K) \leq \varepsilon$.*

Proof. Fix a symmetric convex body $K \subseteq \mathbb{R}^n$ with $\ell_K \leq \delta$ where we will later choose δ small enough. Let $W := \{\boldsymbol{x} \in \mathbb{R}^n : |\langle \boldsymbol{a}_k, \boldsymbol{x}\rangle| \leq \beta_k \ \forall k \in \mathbb{N}\}$ be the symmetric convex body with $W \subseteq K$ from Theorem 4.40. Fix any lattice $\Lambda \subseteq \mathbb{R}^n$ and

any vector $\boldsymbol{u} \in \mathbb{R}^n$. We bound

$$\begin{aligned}
\rho_1((\Lambda+\boldsymbol{u}) \setminus K) &\overset{W \subseteq K}{\leq} \rho_1((\Lambda+\boldsymbol{u}) \setminus W) \\
&\leq \sum_{k=1}^{\infty} \rho_1\big(\{\boldsymbol{x} \in \boldsymbol{u}+\Lambda : |\langle \boldsymbol{a}_k, \boldsymbol{x}\rangle| \geq \beta_k\}\big) \\
&\overset{\text{Lem 4.34}}{\leq} \rho_1(\Lambda) \sum_{k=1}^{\infty} 2\exp\big(-\pi\beta_k^2\big) \\
&= 2\rho_1(\Lambda) \sum_{k=1}^{\infty} \exp\Big(-\underbrace{\frac{\pi C^2}{\ell_K^2}}_{=:\alpha} \log(2k)\Big) \\
&= 2\rho_1(\Lambda) \sum_{k=1}^{\infty} \frac{1}{(2k)^{\alpha}} \leq \varepsilon \cdot \rho_1(\Lambda)
\end{aligned}$$

Here we can make α as large as needed by choosing $\delta > 0$ small enough. □

4.4.4 The proof of the transference theorem for arbitrary symmetric convex bodies

Finally we have all the tools together to prove Theorem 4.31 which we restate for convenience:

Theorem (Theorem 4.31)**.** *For full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any symmetric convex body $K \subseteq \mathbb{R}^n$ one has $1 \leq \lambda_1(\Lambda, K) \cdot \mu(\Lambda^*, K^\circ) \leq O(n\log(n))$.*

Proof. Fix any symmetric convex body $K \subseteq \mathbb{R}^n$. We apply Theorem 4.39 and obtain a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $\ell_{T(K)} \cdot \ell_{(T(K))^\circ} \leq Cn\log n$. As we have to prove the statement over all lattices Λ, we can apply any linear transformation to K and to Λ without affecting whether or not the statement is true. Hence we replace K with $T(K)$ and assume from now on that $\ell_K \cdot \ell_{K^\circ} \leq Cn\log n$; moreover we may assume that K is scaled so that $\ell_K \leq \delta$ where δ is the constant required by Lemma 4.41 in order to achieve the conclusion with $\varepsilon := \frac{1}{4}$. Then $\ell_{K^\circ} \leq \frac{Cn\log(n)}{\delta}$. So we abbreviate $s := \frac{Cn\log(n)}{\delta^2}$ and set $Q := sK^\circ$. Then $\ell_Q = \frac{\ell_{K^\circ}}{s} \leq \delta$. From Lemma 4.41 we know that $\beta(K) \leq \frac{1}{4}$ and $\beta(Q) \leq \frac{1}{4}$. Applying Lemma 4.33 we learn that $\lambda_1(\Lambda, K) \cdot \mu_1(\Lambda^*, Q) \leq 1$. As $\mu(\Lambda^*, Q) = \frac{1}{s}\mu(\Lambda^*, K^\circ)$ we then have $\lambda_1(\Lambda, K) \cdot \mu_1(\Lambda^*, K^\circ) \leq s \leq O(n\log(n))$. □

In an exercise we will prove that $\mu(\Lambda, K) \geq \frac{1}{2}\lambda_n(\Lambda, K)$ still holds for every lattice and every symmetric convex body K. We may then conclude the following:

Theorem 4.42. *For full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any symmetric convex body $K \subseteq \mathbb{R}^n$ one has $1 \leq \lambda_1(\Lambda, K) \cdot \lambda_n(\Lambda^*, K^\circ) \leq O(n \log(n))$.*

4.5 Exercises

Exercise 4.1.
Prove that in any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\mu(\Lambda) \leq n \cdot \lambda_n(\Lambda)$.
Extra point: Prove that even $\mu(\Lambda) \leq O(\sqrt{n}) \cdot \lambda_n(\Lambda)$.

Exercise 4.2.
Show that the definition of the Fourier series does not depend on the chosen basis. More precisely, let $\Lambda \subseteq \mathbb{R}^n$ be a full-rank lattice and let $f : \mathbb{R}^n \to \mathbb{R}$ be a Λ-periodic function. Suppose $\boldsymbol{B}_1, \boldsymbol{B}_2$ are basis with $\Lambda = \Lambda(\boldsymbol{B}_1) = \Lambda(\boldsymbol{B}_2)$. Prove that for for all $\boldsymbol{y} \in \Lambda^*$ one has

$$\int_{\mathcal{P}(\boldsymbol{B}_1)} f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x} = \int_{\mathcal{P}(\boldsymbol{B}_2)} f(\boldsymbol{x}) \cdot e^{-2\pi i \langle \boldsymbol{x}, \boldsymbol{y} \rangle} \, d\boldsymbol{x}.$$

Exercise 4.3.
Prove the following statement: For any symmetric convex body $K \subseteq \mathbb{R}^n$ and any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\lambda_1(K, \Lambda) \cdot \lambda_1(K^\circ, \Lambda^*) \leq Cn$ where $C > 0$ is a universal constant.
Hint: You may use the following deep result of Blaschke-Santaló-Bourgain-Milman without a proof: For any symmetric convex body $K \subseteq \mathbb{R}^n$ one has

$$C_1^n \leq \frac{\mathrm{Vol}_n(K) \cdot \mathrm{Vol}_n(K^\circ)}{\mathrm{Vol}_n(B_2^n)^2} \leq 1$$

for some universal constant $C_1 > 0$.

Exercise 4.4.
Prove the following generalization of Lemma 1.44: For any full rank lattice Λ and any symmetric convex body $K \subseteq \mathbb{R}^n$ one has $\mu(\Lambda, K) \geq \frac{1}{2}\lambda_n(\Lambda, K)$.

Exercise 4.5.
Prove the Shifted Poisson Summation Formula (Cor 4.17): For a nice enough function $f : \mathbb{R}^n \to \mathbb{C}$, a full rank lattice $\Lambda \subseteq \mathbb{R}^n$ and a vector $\boldsymbol{s} \in \mathbb{R}^n$ one has $\sum_{\boldsymbol{x} \in \Lambda} \exp(2\pi i \langle \boldsymbol{x}, \boldsymbol{s} \rangle) \cdot f(\boldsymbol{x}) = \det(\Lambda^*) \cdot \hat{f}(\Lambda^* - \boldsymbol{s})$.

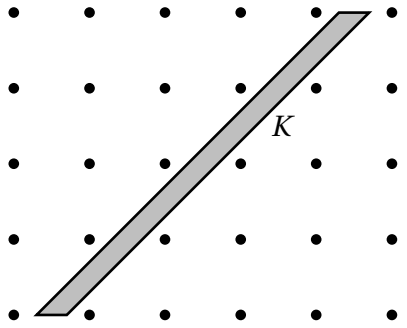
Chapter 5

The Flatness Theorem and Integer Programming

In this chapter we will discuss two applications of the Transference Theorems from Chapter 4.

5.1 The Flatness Theorem

We know from Minkowski's Theorem that a symmetric convex body K with $\mathrm{Vol}_n(K) \geq 2^n$ must contain an an integer point other than the origin. We would like to somehow generalize this to arbitrary convex bodies. But it is easy to see that there are convex bodies with arbitrarily large volume that do not intersect $\mathbb{Z}^n$.



Next, one might get the suspicion that it can only happen that $K \cap \mathbb{Z}^n = \emptyset$, if K is *thin* in some direction. It turns out that this intuition is true in a strong sense: either there is an *integer direction* in which K is thin or otherwise K intersects $\mathbb{Z}^n$.

For a vector $\boldsymbol{c} \in \mathbb{Z}^n$ and a convex body $K \subseteq \mathbb{R}^n$ we define

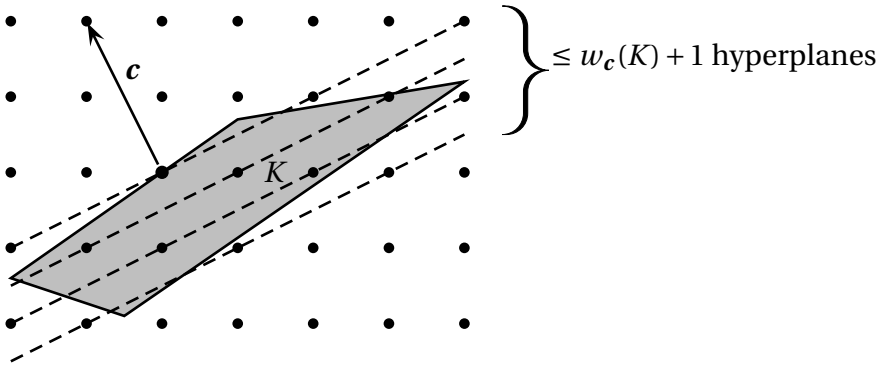
$$\mathrm{width}_{\boldsymbol{c}}(K) := \max\left\{ \langle \boldsymbol{c}, \boldsymbol{x} \rangle - \langle \boldsymbol{c}, \boldsymbol{y} \rangle : \boldsymbol{x}, \boldsymbol{y} \in K \right\}$$

as the *width of K in direction $\boldsymbol{c}$*. Moreover we define

$$\text{intwidth}(K) := \inf_{\boldsymbol{c} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}} \text{width}_{\boldsymbol{c}}(K)$$

as the *integer width* of K. Later in our application to integer programming the following insight will be crucial:

Observation 5.1. Given K and $\boldsymbol{c} \in \mathbb{Z}^n$. All points in $K \cap \mathbb{Z}^n$ are contained in at most $w_{\boldsymbol{c}}(K)+1$ many hyperplanes of the form $K \cap \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{c}^T \boldsymbol{x} = \delta\}$ with $\delta \in \mathbb{Z}$.



It is important to note that width(K) is not the geometric width — it is the geometric width *times* the length $\|\boldsymbol{c}\|_2$. In order to show that width(K) is small one has to find a *short* vector $\boldsymbol{c} \in \mathbb{Z}^n$ so that K is thin in direction $\boldsymbol{c}$.

Theorem 5.2 (Khinchine's Flatness Theorem)**.** *For any convex body $K \subseteq \mathbb{R}^n$ at least one of the following holds:*

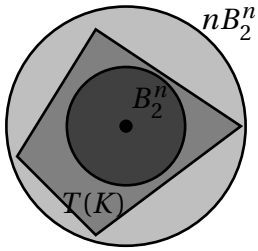
(A) One has $K \cap \mathbb{Z}^n \neq \emptyset$
(B) There is a direction $\boldsymbol{c} \in \mathbb{Z}^n$ with $\text{width}_{\boldsymbol{c}}(K) \leq \mu(\mathbb{Z}^n, K) \cdot \lambda_1(\mathbb{Z}^n, (K-K)^\circ) \leq f(n)$ where one can bound $f(n) \leq O(n^2)$.

In short: any lattice point free convex body K has intwidth$(K) \leq f(n)$. Here Khinchine gave the first bound on intwidth(K) independent on the dimension — however the polynomial upper bounds are more recent.

5.1.1 A transference bound for asymmetric bodies

We want to derive the flatness theorem from the result in Section 4.4. The issue is that those results only hold for *symmetric* convex bodies. Hence we need to be able to approximate an arbitrary convex body with a symmetric one. We remind the reader of the well-known result due to John:

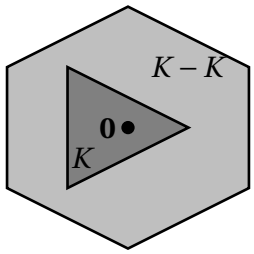
Theorem 5.3 (John (1949))**.** *For any convex body $K \subseteq \mathbb{R}^n$ there is an invertible affine linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $B_2^n \subseteq T(K) \subseteq nB_2^n$.*



Equivalently, for any convex body K there is an ellipsoid $\mathcal{E}$ and a translate $\boldsymbol{u} \in \mathbb{R}^n$ so that $\boldsymbol{u} + \mathcal{E} \subseteq K \subseteq \boldsymbol{u} + n\mathcal{E}$. The bound is tight in general, see for example an equilateral simplex. However for symmetric bodies, this can be improved:

Theorem 5.4 (John (1949))**.** *For any symmetric convex body $K \subseteq \mathbb{R}^n$ there is an invertible linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $B_2^n \subseteq T(K) \subseteq \sqrt{n}B_2^n$.*

Then we could approximate an arbitrary convex body K with an ellipsoid, as ellipsoids are symmetric. But there is another natural choice to use. For a convex body $K \subseteq \mathbb{R}^n$, we consider the *difference body* $K - K := \{\boldsymbol{x} - \boldsymbol{y} : \boldsymbol{x}, \boldsymbol{y} \in K\}$. Then $K - K$ is symmetric. Moreover if $\mathbf{0} \in K$, then $K \subseteq K - K$.



Theorem 5.5. *For any convex body $K \subseteq \mathbb{R}^n$ there is a point $\boldsymbol{u} \in K$ so that $K - \boldsymbol{u} \subseteq K - K \subseteq 2n \cdot (K - \boldsymbol{u})$.*

Proof. The claim is not affected by applying a linear transformation to K, hence we may assume by John's Theorem that $B_2^n \subseteq K \subseteq nB_2^n$. Then $K - K \subseteq 2nB_2^n \subseteq 2nK$. □

Theorem 5.6 (Asymmetric Transference Theorem)**.** *For any lattice $\Lambda \subseteq \mathbb{R}^n$ and any convex body $K \subseteq \mathbb{R}^n$ one has $1 \leq \mu(\Lambda, K) \cdot \lambda_1(\Lambda^*, (K - K)^\circ) \leq 2n^2$.*

Proof. We only prove the upper bound. After applying an affine transformation to K we may assume $B_2^n \subseteq K \subseteq nB_2^n$. From $B_2^n \subseteq K$ we know that $\mu(\Lambda, B_2^n) \geq$

$\mu(\Lambda, K)$. Moreover from $\frac{1}{2n}B_2^n = (2nB_2^n)^\circ \subseteq (K-K)^\circ$ we conclude that $\lambda_1(\Lambda^*, B_2^n) \geq \frac{1}{2n}\lambda_1(\Lambda^*, (K-K)^\circ)$. Then

$$\mu(\Lambda, K) \cdot \lambda_1(\Lambda^*, (K-K)^\circ) \leq \mu(\Lambda, B_2^n) \cdot 2n\lambda_1(\Lambda^*, B_2^n) \leq 2n^2$$

using Theorem 4.30. □

5.1.2 Proof of the Flatness Theorem

Now we can prove the Flatness Theorem:

Proof of Khinchine's Flatness Theorem (Theorem 5.2). Consider a convex body $K \subseteq \mathbb{R}^n$ with $K \cap \mathbb{Z}^n = \emptyset$. From our assumption we know that $\mu(\mathbb{Z}^n, K) > 1$. Let $\boldsymbol{c} \in \mathbb{Z}^n$ be the shortest vector with respect to the dual norm, i.e. $\|\boldsymbol{c}\|_{(K-K)^\circ} = \lambda_1(\mathbb{Z}^n, (K-K)^\circ)$. Fix $\boldsymbol{x}^* := \operatorname{argmax}\{\langle \boldsymbol{c}, \boldsymbol{x}\rangle : \boldsymbol{x} \in K\}$ and $\boldsymbol{y}^* := \operatorname{argmin}\{\langle \boldsymbol{c}, \boldsymbol{y}\rangle : \boldsymbol{y} \in K\}$. Note that $\boldsymbol{x}^* - \boldsymbol{y}^* \in K - K$. Then

$$\begin{aligned}
\operatorname{width}_{\boldsymbol{c}}(K) &= |\langle \boldsymbol{c}, \boldsymbol{x}^* - \boldsymbol{y}^* \rangle| \\
&\overset{\text{C.S.}}{\leq} \underbrace{\|\boldsymbol{c}\|_{(K-K)^\circ}}_{=\lambda_1(\mathbb{Z}^n,(K-K)^\circ)} \cdot \underbrace{\|\boldsymbol{x}^* - \boldsymbol{y}^*\|_{K-K}}_{\leq 1 \leq \mu(\mathbb{Z}^n,K)} \\
&\leq \lambda_1(\mathbb{Z}^n, (K-K)^\circ) \cdot \mu(\mathbb{Z}^n, K) \overset{\text{Thm 5.6}}{\leq} 2n^2
\end{aligned}$$

Here we use the the Cauchy Schwarz Inequality, the fact that $(\mathbb{Z}^n)^* = \mathbb{Z}^n$ and the bound from the Asymmetric Transference Theorem in Theorem 5.6. □

For the sake of completness we outline that in some sense the inequality used in the flatness theorem is exact:

Lemma 5.7. *The following holds:*

(a) If $Q \subseteq \mathbb{R}^n$ is a symmetric convex body, then intwidth$(Q) = 2\lambda_1(\mathbb{Z}^n, Q^\circ)$.

(b) If $K \subseteq \mathbb{R}^n$ is a convex body with $\mu(\mathbb{Z}^n, K) = 1$, then

$$\mu(\mathbb{Z}^n, K) \cdot \lambda_1(\mathbb{Z}^n, (K-K)^\circ) = \textit{intwidth}(K)$$

Proof. To see (a), we write

$$\operatorname{intwidth}(Q) = 2 \inf_{\boldsymbol{c} \in \mathbb{Z}^n \setminus \{\boldsymbol{0}\}} \max\{|\langle \boldsymbol{c}, \boldsymbol{x}\rangle| : \boldsymbol{x} \in Q\} = 2 \inf_{\boldsymbol{c} \in \mathbb{Z}^n \setminus \{\boldsymbol{0}\}} \|\boldsymbol{x}\|_{Q^\circ} = 2\lambda_1(\mathbb{Z}^n, Q^\circ)$$

as $\|\cdot\|_{Q^\circ}$ is the *dual norm* to $\|\cdot\|_Q$. Then for (b) we have

$$\lambda_1(\mathbb{Z}^n, (K-K)^\circ) \overset{(a)}{=} \frac{1}{2}\operatorname{intwidth}(K-K) = \operatorname{intwidth}(K)$$

□

We also want to mention without a proof, the following qualitative strengthening of the flatness theorem: either K is very flat in some direction or there must actually be many directions in which the body is at least moderately flat:

Theorem 5.8 ([KL88])**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body with $K \cap \mathbb{Z}^n = \emptyset$. Then for some $k \in \{1,\dots,n\}$ there are linearly independent vectors $c_1,\dots,c_k \in \mathbb{Z}^n$ so that $width_{c_i}(K) \leq O(k^3 \log^2(2k))$ for all $i = 1,\dots,k$.*

5.2 Application to Integer Programming

In this section we will see a beautiful application of the Flatness Theorem. Integer programming is one of the most powerful and most useful problems in discrete optimization.

INTEGER PROGRAMMING (IP)
Input: A linear inequality system $\boldsymbol{Ax} \leq \boldsymbol{b}$ with $\boldsymbol{A} \in \mathbb{Q}^{m\times n}$, $\boldsymbol{b} \in \mathbb{Q}^m$
Goal: Find a point $\boldsymbol{x} \in K \cap \mathbb{Z}^n$ where $K := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{Ax} \leq \boldsymbol{b}\}$

This problem is among the first problems that were shown to be **NP**-hard. The fundamental practical importance comes from the fact that the standard approach for operations research practicioners is to model whatever problem appears in their real-world application as an integer linear program and then solve it using quite sophisticated software tools. In this chapter, we will look at the problem from a purely theoretical perspective. For a more detailed treatment, we refer to the survey of Kannan [Kan87b].

As the integer programming problem is **NP**-hard, there is no hope for a polynomial time algorithm, but it is natural to ask whether the problem can be solved in time $T(n)\cdot \text{poly}(m, \log\|\boldsymbol{A}\|_\infty, \log\|\boldsymbol{b}\|_\infty)$ where $T(n)$ will be some exponentially growing function of the dimension; here we have implicitly assumed that $\boldsymbol{A}$ and $\boldsymbol{b}$ are scaled to be integers. In other words: *can we solve integer programming in polynomial time when the dimension is some fixed constant?* The affirmative answer due is to Lenstra [Len83]. The idea is that if we have any direction $\boldsymbol{c} \in \mathbb{Z}^n$ where say $\text{width}_{\boldsymbol{c}}(K) \leq n^{O(1)}$, then we know that all the candidate solutions $K \cap \mathbb{Z}^n$ lie on $n^{O(1)}$ many hyperplanes of the form $\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{c}, \boldsymbol{x}\rangle = \delta\}$ for $\delta \in \mathbb{Z}$. In each hyperplane we have to solve an $(n-1)$-dimensional subproblem. Overall this would result in a $n^{O(n)}$ time algorithm. There is of course the problem that there might not be a direction $\boldsymbol{c} \in \mathbb{Z}^n$ where $\text{width}_{\boldsymbol{c}}(K) \leq n^{O(1)}$. In that case the problem is easy to be solved directly by rounding the "coordinates" of a point in the center of K where "coordinates" is with respect to a short system of linearly independent vectors.

The original algorithm of Lenstra [Len83] was based on the LLL algorithm which is even a polynomial time algorithm but it only provides exponential guarantees on the width. In contrast, we will rather use the exact $2^{O(n)}$-time algorithm for finding the shortest vector as a subroutine. This gives the best known running time (up to constants in the exponent) and makes for a cleaner algorithm.

We can now give the complete algorithm:

Lenstra's algorithm for Integer Programming

Input: Polytope $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ given by matrix $\boldsymbol{A} \in \mathbb{Q}^{m\times n}$ and vector $\boldsymbol{b} \in \mathbb{Q}^m$.

Output: Either a point $K \cap \mathbb{Z}^n$ or decision that none exists.

(1) Compute linearly independent vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{Z}^n$ with $\|\boldsymbol{v}_i\|_{K-K} \leq 2\lambda_i(\mathbb{Z}^n, K-K)$ for $i = 1, \ldots, n$.

(2) Compute the center $\boldsymbol{a} \in K$ so that $\frac{1}{2n}(K-K) \subseteq K - \boldsymbol{a} \subseteq K - K$

(3) Write $\boldsymbol{a} = \sum_{i=1}^n \lambda_i \boldsymbol{v}_i$ with $\lambda_i \in \mathbb{R}$.

(4) IF $\boldsymbol{x}^* := (\sum_{i=1}^n \lfloor \lambda_i \rfloor \boldsymbol{v}_i) \in K$ THEN return $\boldsymbol{x}^*$

(5) Compute $\boldsymbol{c} \in \mathbb{Z}^n$ with $\|\boldsymbol{c}\|_{(K-K)^\circ} = \lambda_1(\mathbb{Z}^n, (K-K)^\circ)$

(6) FOR all $\delta \in \{\lceil \min\{\langle \boldsymbol{c}, \boldsymbol{x}\rangle \mid \boldsymbol{x} \in K\}\rceil, \ldots, \lfloor \max\{\langle \boldsymbol{c}, \boldsymbol{x}\rangle \mid \boldsymbol{x} \in K\}\rfloor\}$ DO

  (7) Run the Hermite normal form algorithm to find a lattice basis $\boldsymbol{B}' \in \mathbb{Q}^{n\times(n-1)}$ and an offset $\boldsymbol{d}$ with $\boldsymbol{d} + \Lambda(\boldsymbol{B}') = \{\boldsymbol{x} \in \mathbb{Z}^n \mid \boldsymbol{c}^T\boldsymbol{x} = \delta\}$

  (8) Run the algorithm recursively to find integer point in $\{\boldsymbol{x}' \in \mathbb{R}^{n-1} \mid \boldsymbol{A}(\boldsymbol{d} + \boldsymbol{B}'\boldsymbol{x}') \leq \boldsymbol{b}\}$

(9) Return any point in $K \cap \mathbb{Z}^n$ that was found or decide that none exists otherwise

In (1) we use Theorem 3.9 to compute 2-approximate successive minima in time $2^{O(n)}$.

Theorem 5.9. *For a polytope $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$, Lenstra's algorithm finds a point $\boldsymbol{x}^* \in K \cap \mathbb{Z}^n$ in time $n^{O(n)}$ times a polynomial in the encoding length of $\boldsymbol{A}$ and $\boldsymbol{b}$ (if there is any such point).*

Proof. The correctness is clear in the sense that the algorithm does an exhaustive search or terminates with finding an integral point. So it remains to bound the running time and in particular the number of recursions.

Claim I. *If the algorithm does not terminate in (4) then $\lambda_n(\mathbb{Z}^n, K-K) \geq \frac{1}{4n^2}$.*

Proof of Claim I. Assume for the sake of contradiction that $\lambda_n(\mathbb{Z}^n, K-K) < \frac{1}{4n^2}$. Then the "rounding error" was

$$\|\boldsymbol{a} - \boldsymbol{x}^*\|_{K-K} \leq \sum_{i=1}^n \|\boldsymbol{v}_i\|_{K-K} \leq 2n \cdot \lambda_n(\mathbb{Z}^n, K-K) \leq \frac{1}{2n}$$

and so $\boldsymbol{x}^* \in K$. □

Claim II. *In (6) the algorithm recurses on at most $O(n^3 \log(n))$ many subproblems.*

Proof of Claim II. By Theorem 4.42 we know that $\lambda_n(\mathbb{Z}^n, K-K) \cdot \lambda_1(\mathbb{Z}^n, (K-K)^\circ) \leq O(n \log n)$. By Claim I we know that if we reach (6) then $\lambda_n(\mathbb{Z}^n, K-K) \geq \frac{1}{4n^2}$. Hence $\|\boldsymbol{c}\|_{(K-K)^\circ} = \lambda_1(\mathbb{Z}^n, (K-K)^\circ) \leq O(n^3 \log n)$. As before we may consider $\boldsymbol{x}^* := \operatorname{argmax}\{\langle \boldsymbol{c}, \boldsymbol{x} \rangle : \boldsymbol{x} \in K\}$ and $\boldsymbol{y}^* := \operatorname{argmin}\{\langle \boldsymbol{c}, \boldsymbol{y} \rangle : \boldsymbol{y} \in K\}$ and the number of recursions is bounded by $|\langle \boldsymbol{c}, \boldsymbol{x}^* - \boldsymbol{y}^* \rangle| + 1 \leq \|\boldsymbol{c}\|_{(K-K)^\circ} \cdot \|\boldsymbol{x}^* - \boldsymbol{y}^*\|_{K-K} + 1 \leq O(n^3 \log(n))$ by Cauchy Schwarz. □

Finally note that $T(n) \leq O(n^3 \log(n)) \cdot T(n-1) + 2^{O(n)}$ where the $+2^{O(n)}$ comes from the subroutines to find shortest vectors with respect to the given norms. Overall this may be resolved to $T(n) \leq n^{O(n)}$. □

5.3 Improved Transference and Flatness bounds for non-symmetric convex bodies*

This section is intended as additional material if towards the end of the course there is time left or alternatively as additional reading material for the interested reader. In this section we will discuss a result of Banaszczyk, Litvak, Pajor and Szarek [BLPS99] which improves the bound of $O(n^2)$ appearing in both, the Flatness Theorem 5.2 and in the Asymmetric Transference Theorem 5.6 down to $O(n^{3/2}\sqrt{\log n})$[1].

5.3.1 Preliminaries

First, in order to handle asymmetric convex bodies, we need to extend some of the notation. For a convex body $K \subseteq \mathbb{R}^n$ with $\mathbf{0} \in \operatorname{int}(K)$ we define the *Gauge function*

$$\|\boldsymbol{x}\|_K := \inf\{t \geq 0 \mid \boldsymbol{x} \in tK\}$$

If K happens to be symmetric then $\|\cdot\|_K$ is simply the norm with K as unit ball. For an asymmetric convex body, $\|\cdot\|_K$ is not a symmetric function but we still have the following properties:

(1) Subadditivity: $\|\boldsymbol{x} + \boldsymbol{y}\|_K \leq \|\boldsymbol{x}\|_K + \|\boldsymbol{y}\|_K$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

(2) Positive homogeneity: $\|s\boldsymbol{x}\|_K = s\|\boldsymbol{x}\|_K$ for $\boldsymbol{x} \in \mathbb{R}^n$ and $s \geq 0$.

(3) Non-negativity: $\|\boldsymbol{x}\|_K \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

[1] To be precise the original paper puts in additional effort to remove the $\sqrt{\log(n)}$ factor but we will skip that part here.

We hope that it will not lead to confusion that we use the notation $\|\cdot\|_K$ for something else than a norm. Next, we extend the notion of the *ℓ-value*. For a convex body $K \subseteq \mathbb{R}^n$ with $\mathbf{0} \in \text{int}(K)$ and $p \geq 1$ we define

$$\ell_p(K) := \mathop{\mathbb{E}}_{\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)} \left[\|\boldsymbol{x}\|_K^p \right]^{1/p}$$

The case $\ell(K) := \ell_2(K)$ coincides with the quantity that we have defined earlier. The reason for us to introduce the ℓ-value with general $p \geq 1$ is that sometimes it is easier to work with the case $p = 2$ and in some situations it is easier to work in with $p = 1$. But conveniently, the different values $\ell_p(K)$ only differ by a constant (depending on p) anyway.

Lemma 5.10. *For any convex body $K \subseteq \mathbb{R}^n$ with $\mathbf{0} \in \text{int}(K)$ one has*

$$\frac{c}{\sqrt{p}} \ell_p(K) \leq \ell_1(K) \leq \ell_p(K) \quad \forall p \geq 1$$

where $c > 0$ is a universal constant.

The proof is not actually difficult but it requires a few additional fact from probability theory that we do not want to fully spell out for space reasons. However we give a sketch; the book of Vershynin [Ver19] is a good source to look up details.

Proof sketch. Let us abbreviate the random variable $X := \|\boldsymbol{x}\|_K$ where $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)$. Then the 2nd part of the claim translates to $\mathbb{E}[X] \leq \mathbb{E}[X^p]^{1/p}$ and this inequality is indeed true by *Jensen's inequality* and the fact that that function $z \to z^p$ is concave for $z \geq 0$.

For the 2nd direction, we scale K so that $\ell_1(K) = 1$. We claim that then $\frac{1}{3} B_2^n \subseteq K$. Suppose this was false. In that case there is a unit vector $\boldsymbol{a} \in S^{n-1}$ so that the halfspace $H := \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x} \rangle \leq \frac{1}{3}\}$ contains K. But then $\ell_1(K) \geq \ell_1(H) = \mathbb{E}_{\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)}[\max\{0, 3\langle \boldsymbol{a}, \boldsymbol{x} \rangle\}] = 3 \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty z dz = \frac{3}{\sqrt{2\pi}} > 1$ which is a contradiction. Hence the map $\boldsymbol{x} \to \|\boldsymbol{x}\|_K$ is $\frac{1}{3}$-Lipschitz with mean 1 for $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)$. Then $\|X - 1\|_{\psi_2} \leq O(1)$ by concentration for Lipschitz functions and so $\mathbb{E}[|X-1|^p]^{1/p} \leq O(\sqrt{p})$ by Lemma 4.35. □

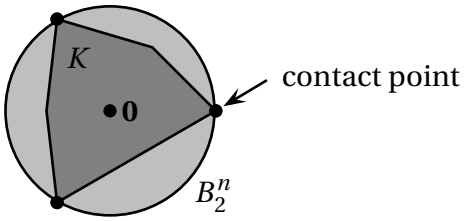
We have already a version of John's Theorem in Theorem 5.3 which sometimes is called the "Basic John's Theorem" and merely tells us that after some linear transformation one has $\frac{1}{n} B_2^n \subseteq K \subseteq B_2^n$. In contrast, the full version of John's Theorem also gives an exact characterization when B_2^n is the smallest ellipsoid containing K:

Theorem 5.11 (Full version of John's Theorem)**.** *Let $K \subseteq \mathbb{R}^n$ be a convex body with the property that* $\min\{Vol_n(\mathcal{E}) \mid K \subseteq \mathcal{E}$ *and* $\mathcal{E}$ *is ellipsoid*$\}$ *is attained for $\mathcal{E} = B_2^n$. Then there are* contact points *$\boldsymbol{u}_1, \dots, \boldsymbol{u}_m \in \mathbb{R}^n$ with the following property:*

(i) One has $\|\boldsymbol{u}_i\|_K = \|\boldsymbol{u}_i\|_2 = 1$ for all $i = 1, \dots, m$.

(ii) There are coefficients $\boldsymbol{\lambda} \in \mathbb{R}^m_{\geq 0}$ so that $\sum_{i=1}^m \lambda_i = n$, $\sum_{i=1}^m \lambda_i \boldsymbol{u}_i = \boldsymbol{0}$ and $\sum_{i=1}^m \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T = \boldsymbol{I}_n$.

(iii) One has $m \leq 2n^2$.



Here the term *contact point* means a point that lies on the intersection of the boundary of K and the boundary of B_2^n. One should think of condition (ii) as the property that there are boundary points "in all directions" which is somewhat intuitive because otherwise we could "shrink" the B_2^n and obtain a smaller ellipsoid still containing K. We highly recommend the survey of Ball [Bal97] for details on John's theorem. We also remind the reader of a standard fact in probability theory:

Lemma 5.12. *Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m \in \mathbb{R}^n$ be a finite set of vectors. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)} \left[\max_{i=1,\dots,m} |\langle \boldsymbol{u}_i, \boldsymbol{x} \rangle| \right] \leq C \sqrt{\log(2m)} \cdot \max\{\|\boldsymbol{u}_i\|_2 : i = 1, \dots, m\}$$

where $C > 0$ is a universal constant.

Again, see e.g. [Ver19] for a proof.

5.3.2 An $\ell\ell^\circ$-estimate for asymmetric bodies via John's Theorem

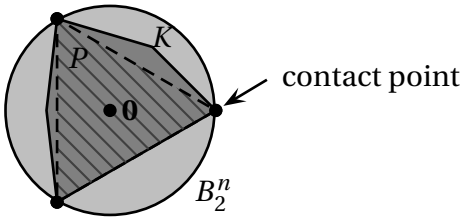
A key ingredient of previous proofs of the transference theorem was to apply a linear transformation to the body K so that the product $\ell(K) \cdot \ell(K^\circ)$ was small which intuitively means that neither K nor K° are particularly thin. The issue is that the proof of the so-called *$\ell\ell^\circ$-estimate* (Theorem 4.39) crucially relies on symmetry and does not carry over to general convex bodies. So we prove a (quantitatively weaker) statement that works for the non-symmetric case too:

Lemma 5.13. *Let $K \subseteq \mathbb{R}^n$ be a convex body with the property that $\min\{Vol_n(\mathcal{E}) \mid K \subseteq \mathcal{E}$ and $\mathcal{E}$ is ellipsoid$\}$ is attained for $\mathcal{E} = B_2^n$. Then $\ell(K) \leq O(n\sqrt{\log(2n)})$ and $\ell(K^\circ) \leq O(\sqrt{n})$. In the special case that K is a polytope with N vertices one additionally has $\ell(K^\circ) \leq O(\sqrt{\log(2N)})$.*

Proof. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ be the contact points with coefficients $\boldsymbol{\lambda} \in \mathbb{R}^m_{\geq 0}$ as guaranteed by John's Theorem (Theorem 5.11). We can upper bound the value $\|\boldsymbol{x}\|_K$ using those contact points:

Claim I. *For each $\boldsymbol{x} \in \mathbb{R}^n$ one has $\|\boldsymbol{x}\|_K \leq 2n \cdot \max_{i=1,\ldots,m} |\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle|$.*

Proof of Claim. Fix $\boldsymbol{x} \in \mathbb{R}^n$. Consider $P := \text{conv}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ and note that $P \subseteq K$ meaning that $\|\boldsymbol{x}\|_K \leq \|\boldsymbol{x}\|_P$.



Next, note that the value $\|\boldsymbol{x}\|_P$ is equal to the amount of positive weight needed to write $\boldsymbol{x}$ as a conic combination, i.e.

$$\|\boldsymbol{x}\|_P = \min\left\{\|\boldsymbol{t}\|_1 : \boldsymbol{x} = \sum_{i=1}^m t_i \boldsymbol{u}_i \text{ and } \boldsymbol{t} \in \mathbb{R}^m_{\geq 0}\right\} \quad (*)$$

Let $s \geq 0$ be a parameter that we determine later. We use the properties from John's Theorem to write

$$\begin{aligned} \boldsymbol{x} &= \overbrace{\left(\sum_{i=1}^m \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T\right)}^{=\boldsymbol{I}_n} \boldsymbol{x} = \sum_{i=1}^m \lambda_i \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle \boldsymbol{u}_i \\ &= \sum_{i=1}^m \lambda_i \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle \boldsymbol{u}_i - s \underbrace{\sum_{i=1}^m \lambda_i \boldsymbol{u}_i}_{=\boldsymbol{0}} = \sum_{i=1}^m \lambda_i \cdot \big(\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - s\big)\boldsymbol{u}_i \end{aligned}$$

In principle this looks like we can apply $(*)$ in order to bound $\|\boldsymbol{x}\|_P$ but we need to make sure that the coefficients are non-negative which means that $\lambda_i \cdot (\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - s) \geq 0$ for all $i = 1, \ldots, m$. Since $\lambda_i \geq 0$ we can choose $s := \min\{\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle : i = 1, \ldots, m\}$. Then applying $(*)$ gives

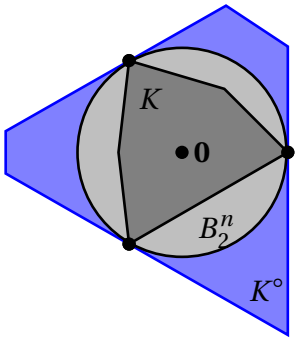
$$\|\boldsymbol{x}\|_P \leq \sum_{i=1}^m \lambda_i \cdot (\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle - s) \leq \underbrace{\sum_{i=1}^m \lambda_i}_{=n} \cdot 2 \max_{i=1,\ldots,m} |\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle| \leq 2n \cdot \max_{i=1,\ldots,m} |\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle|$$

as claimed. □

By Lemma 5.10, it suffices to prove upper bounds on $\ell_1(K)$ and $\ell_1(K^\circ)$ rather than $\ell(K)$ and $\ell(K^\circ)$. For the first quantity we use

$$\ell_1(K) = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim N(\mathbf{0},\boldsymbol{I}_n)}\left[\|\boldsymbol{x}\|_K\right] \overset{\text{Claim I}}{\leq} 2n \mathop{\mathbb{E}}_{\boldsymbol{x}\sim N(\mathbf{0},\boldsymbol{I}_n)}\left[\max_{i=1,\dots,m} |\langle \boldsymbol{u}_i, \boldsymbol{x}\rangle|\right] \overset{(**)}{\leq} O\big(n\sqrt{\log(n)}\big)$$

where in $(**)$ we use Lemma 5.12 and the fact that $\|\boldsymbol{u}_i\|_2 = 1$ for all $i = 1,\dots,m$. It remains to prove the bounds on $\ell_1(K^\circ)$. First note that $K \subseteq B_2^n$ implies that $B_2^n \subseteq K^\circ$ and so $\ell_1(K^\circ) \leq \ell_1(B_2^n) \leq O(\sqrt{n})$.



Next, we prove the bound in terms of the number of vertices of K. Let $K = \operatorname{conv}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_N\}$ be the description of K as a convex hull of its vertices where $\|\boldsymbol{a}_i\|_2 \leq 1$ for all $i = 1,\dots,N$. Then $K^\circ = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle \leq 1 \ \forall i = 1,\dots,N\}$. Hence

$$\ell_1(K^\circ) = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim N(\mathbf{0},\boldsymbol{I}_n)}\left[\max\{\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle : i = 1,\dots,N\}\right] \leq O\big(\sqrt{\log(2N)}\big)$$

again by Lemma 5.12. □

Corollary 5.14. *For any convex body $K \subseteq \mathbb{R}^n$, there exists an affine linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $\ell(T(K)) \cdot \ell((T(K))^\circ) \leq O(n^{3/2}\sqrt{\log(n)})$.*

Proof. Choose T as an affine linear map so that the minimum volume ellipsoid containing $T(K)$ happens to be B_2^n. Then Lemma 5.13 gives $\ell(K) \leq O(n\sqrt{\log(2n)})$ and $\ell(K^\circ) \leq O(\sqrt{n})$. □

Furthermore, we can recover the $\ell\ell^\circ$-estimate of $\ell(K) \cdot \ell(K^\circ) \leq O(n\log(n))$ at least for polytopes where the number of vertices is bounded by a polynomial in n.

Corollary 5.15. *For any polytope $K \subseteq \mathbb{R}^n$ with N vertices, there exists an affine linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $\ell(T(K)) \cdot \ell((T(K))^\circ) \leq O(n\sqrt{\log(n) \cdot \log(N)})$.*

The proof uses an identical argument to Cor 5.14.

5.3.3 The improved asymmetric transference theorem

We need to generalize another fact from earlier:

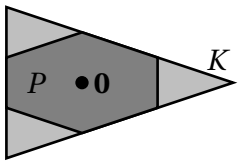
Lemma 5.16. *For any lattice $\Lambda \subseteq \mathbb{R}^n$, any convex body $K \subseteq \mathbb{R}^n$ with $\mathbf{0} \in int(K)$ and any symmetric convex body $Q \subseteq \mathbb{R}^n$ one has*

$$\mu(\Lambda, K) \cdot \lambda_1(\Lambda^*, Q) \leq C_1 \cdot \ell(K) \cdot \ell(Q)$$

where C_1 is a universal constant.

Proof. First we prove the statement for the case that both bodies K and Q are symmetric. Scale K and Q so that $\ell(K) = \delta = \ell(Q)$ where δ is the constant from Theorem 4.41 to achieve that $\beta(K), \beta(Q) \leq \varepsilon = \frac{1}{4}$. Then by Lemma 4.33 we have $\mu(\Lambda, K) \cdot \lambda_1(\Lambda^*, Q) \leq 1 \leq C_0 \cdot \ell(K) \cdot \ell(Q)$ if we choose $C_0 := \frac{1}{\delta^2}$. That concludes the argument for the case that K is symmetric.

Next, we consider the case that K is potentially asymmetric and we want to reduce it to the symmetric case. For that we consider the "symmetrizer" $P := K \cap (-K)$. Trivially P is a symmetric convex body with $P \subseteq K$.



While in terms of inclusion, P might be drastically smaller that K, the ℓ-value of P is close to the one of K:

Claim I. *One has $\ell(P) \leq 2\ell(K)$.*

Proof of Claim I. For each $\boldsymbol{x}$ one has $\|\boldsymbol{x}\|_P = \max\{\|\boldsymbol{x}\|_K, \|\boldsymbol{x}\|_{-K}\} = \max\{\|\boldsymbol{x}\|_K, \|-\boldsymbol{x}\|_K\}$. Then by symmetry of the Gaussian distribution for at least half the outcomes of $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_n)$ the values of $\|\boldsymbol{x}\|_K$ and $\|\boldsymbol{x}\|_P$ will coincide and so $\mathbb{E}_{\boldsymbol{x}\sim N(\mathbf{0},\boldsymbol{I}_n)}[\|\boldsymbol{x}\|_K^2]^{1/2} \geq \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim N(\mathbf{0},\boldsymbol{I}_n)}[\|\boldsymbol{x}\|_P^2]^{1/2}$. □

Now we conclude that

$$\mu(\Lambda, K) \cdot \lambda_1(\Lambda^*, Q) \overset{P\subseteq K}{\leq} \mu(\Lambda, P) \cdot \lambda_1(\Lambda^*, Q) \overset{\text{sym. case}}{\leq} C_0 \ell(P) \ell(Q) \overset{\text{Claim I}}{\leq} 2C_0 \ell(K) \ell(Q)$$

□

Combining Cor 5.14 and Lemma 5.16 gives:

Theorem 5.17. *For any convex body $K \subseteq \mathbb{R}^n$ and any lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\mu(\Lambda, K) \cdot \lambda_1(\Lambda^*, (K-K)^\circ) \leq O(n^{3/2}\sqrt{\log n})$.*

Combining Cor 5.15 and Lemma 5.16 gives:

Theorem 5.18. *For any full-dimensional polytope $K \subseteq \mathbb{R}^n$ with N vertices and any lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\mu(\Lambda, K) \cdot \lambda_1(\Lambda^*, (K-K)^\circ) \leq O(n\sqrt{\log(n) \cdot \log(N)})$.*

5.3.4 The Flatness Constant

We want to summarize some of the bounds that we have discussed and put them into context. The *flatness constant* in dimension n is the maximum integer width of a lattice point free convex body, i.e.

$$\text{flatness}(n) := \sup_{\substack{K \subseteq \mathbb{R}^n \text{ convex} \\ \text{with } K \cap \mathbb{Z}^n = \emptyset}} \{\text{intwidth}(K)\}$$

The best known bounds are

Theorem 5.19. *One has $(2-o(1))n \leq \text{flatness}(n) \leq \Theta(n^{4/3}\log^{O(1)}(n))$.*

We will see a lower bound of $\text{flatness}(n) \geq n$ in an exercise. The bound of $\text{flatness}(n) \geq (2-o(1))n$ is due to [MSW21]. The upper bound follows from the work of Rudelson [Rud98]. In fact, Rudelson showed that the $O(n^{3/2}\sqrt{\log(n)})$ upper bound on the $\ell\ell^\circ$-value of a convex body that we have seen in Cor 5.14 can be improved:

Lemma 5.20 (Rudelson [Rud98])**.** *For any convex body $K \subseteq \mathbb{R}^n$, there exists an affine linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ so that $\ell(T(K)) \cdot \ell((T(K))^\circ) \leq O(n^{4/3}\log^{O(1)}(n))$.*

Then with Lemma 5.16 this immediately implies the improved upper bound on the flatness constant. We also summarize the improved bounds for special cases:

Corollary 5.21. Let $K \subseteq \mathbb{R}^n$ be a convex body with $K \cap \mathbb{Z}^n = \emptyset$. Then the following holds:

(a) One has $\text{intwidth}(K) \leq \Theta(n^{4/3}\log^{O(1)}(n))$.

(b) If K is symmetric with respect to some center $\boldsymbol{u} \in K$, then $\text{intwidth}(K) \leq O(n\log n)$.

(c) If K is a polytope with N vertices then $\text{intwidth}(K) \leq O(n\sqrt{\log(n) \cdot \log(N)})$.

(d) If K is a polytope with N facets then $\text{intwidth}(K) \leq O(n\sqrt{\log(n) \cdot \log(N)})$.

Note that (d) follows immediately from Cor 5.15 by switching the roles of K and K°.

Exercises

Exercise 5.1.
Prove that for any convex body $K \subseteq \mathbb{R}^n$ and any lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\frac{1}{2}\lambda_n(\Lambda, K-K) \leq \mu(\Lambda, K) \leq 2n\lambda_n(\Lambda, K-K)$.

Exercise 5.2.
We want to prove that flatness$(n) \geq n$. For this sake consider the simplex $K := \text{conv}\{\mathbf{0}, n\boldsymbol{e}_1, \ldots, n\boldsymbol{e}_n\} \subseteq \mathbb{R}^n$.

1. Prove that $\text{int}(K) \cap \mathbb{Z}^n = \emptyset$.
2. Prove that for any $\boldsymbol{c} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}$ one has $\text{width}_{\boldsymbol{c}}(K) \geq n$.

Exercise 5.3.
Let $K \subseteq \mathbb{R}^n$ be a convex body with $\text{Vol}_n(K) \leq (\frac{c}{n})^n$ for a small enough universal constant $c > 0$. Prove that there is an $\boldsymbol{a} \in \mathbb{Z}^n$ and $\beta \in \mathbb{Z}$ so that $K \cap \mathbb{Z}^n \subseteq \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{a}, \boldsymbol{x} \rangle = \beta\}$.
Hint. You may use the following results from convex geometry without a proof:

1. **Fact 1.** For any convex body $P \subseteq \mathbb{R}^n$ one has $\text{Vol}_n(P-P) \leq 2^n \text{Vol}_n(P)$.
2. **Fact 2.** For any symmetric convex body $Q \subseteq \mathbb{R}^n$ one has $C_1^n \leq \frac{\text{Vol}_n(Q)\cdot\text{Vol}_n(Q^\circ)}{\text{Vol}_n(B_2^n)^2} \leq 1$ for some universal constant $C_1 > 0$.

Chapter 6

Lattice problems in NP ∩ coNP

The result that we discuss in this chapter will be more of a complexity-theoretic nature and is due to Aharonov and Regev [AR05]. However in terms of techniques we will make additional use of *Fourier analysis* for lattices that we learned in Chapter 4.

Recall that the LLL algorithm from Chapter 1.5 is a polynomial time algorithm that finds a $2^{n/2}$-approximation to the shortest vector $\lambda_1(\Lambda)$ in a given lattice Λ. In the other hand, assuming $\mathbf{NP} \not\subseteq \mathbf{BPTIME}(2^{\mathrm{poly}(\log(n))})$, there is no polynomial time algorithm to approximate the shortest vector within a factor of $2^{(\log n)^{1/2-\varepsilon}}$ (which is still less than n^{δ} for any constant $\delta > 0$). For most practically relevant problems one is used to the outcome that either the problem is **NP**-hard or it is solvable in polynomial time. Oddly, there is evidence that finding, say a polynomial factor approximation to the shortest vector is neither **NP**-hard nor in **P**.

In complexity theory one usually formulates problems as *decision problems* where the answer to be computed is either *yes* or *no*. For $\beta \geq 1$ consider the following problem:

GAPSVP$_\beta$
Input: A lattice $\Lambda := \Lambda(\boldsymbol{B})$
Goal: Distinguish the following cases:
- **YES**. One has $\lambda_1(\Lambda) \leq 1$
- **NO**. One has $\lambda_1(\Lambda) \geq \beta$

One may imagine that instances with $1 < \lambda_1(\Lambda) < \beta$ will not appear as input. This is also called a *promise* problem. Note that the larger β is, the easier is the problem where GAPSVP$_{2^{n/2}}$ is solvable in polynomial time.

6.1 GapSVP$_{4n}$ is in NP ∩ coNP

To warm up we describe a result that is already implicitly included in what we have proven earlier — we can certify that $\lambda_1(\Lambda)$ is large by certifying that $\lambda_n(\Lambda^*)$ is small.

Theorem 6.1. *The problem* GAPSVP$_{4n}$ *is in* **NP** ∩ **coNP**.

Proof. The part GAPSVP$_{4n} \in$ **NP** is trivial — the certificate that we are in the YES case is any lattice vector $\boldsymbol{x} \in \Lambda \setminus \{\mathbf{0}\}$ with $\|\boldsymbol{x}\|_2 \leq 1$. Note that given a vector $\boldsymbol{x}$ we can indeed decide in polynomial time whether $\boldsymbol{x} \in \Lambda$. We use the following verifier to show that GAPSVP$_{4n} \in$ **coNP**:

Input: Lattice $\Lambda(\boldsymbol{B})$
Certificate: Vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n \in \mathbb{R}^n$
Verifier: Accept if all of the following holds:
(i) $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n \in \Lambda^*$
(ii) $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ are linearly independent
(iii) $\|\boldsymbol{w}_i\|_2 \leq \frac{1}{2}$ for all $i = 1, \ldots, n$.

Now we show correctness of the verifier.
Claim I. *If $\lambda_1(\Lambda) \leq 1$ then the verifier rejects any certificate.*
Proof of Claim I. Suppose for the sake of contradiction that the verifier accepts while there is a vector $\boldsymbol{x} \in \Lambda \setminus \{\mathbf{0}\}$ with $\|\boldsymbol{x}\|_2 \leq 1$. There must be at least one i with $\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \neq 0$ and so by definition of dual lattice one has $1 \leq |\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle| \leq \|\boldsymbol{w}_i\|_2 \|\boldsymbol{x}\|_2 \leq \frac{1}{2}\|\boldsymbol{x}\|_2$. Then $\|\boldsymbol{x}\|_2 \geq 2$ which gives a contradiction. □

Claim II. *If $\lambda_1(\Lambda) \geq 4n$ then there exists a certificate that the verifier accepts.*
Proof of Claim II. Choose $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n \in \Lambda^*$ as the successive minima of the dual lattice with $\|\boldsymbol{w}_i\|_2 = \lambda_i(\Lambda^*)$. By Cor 4.2 we have $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \leq 2n$ and so $\lambda_n(\Lambda^*) \leq \frac{1}{2}$. Then the verifier would accept. □

6.2 GAPCVP with gap $O(\sqrt{n})$ is in NP ∩ coNP

Now we will come to the main result of this chapter, namely that we can decrease the gap to $O(\sqrt{n})$ even for the more general *Closest Vector problem*. Recall that $d(\boldsymbol{t}, \Lambda) := \min\{\|\boldsymbol{t} - \boldsymbol{x}\|_2 : \boldsymbol{x} \in \Lambda\}$ denotes the distance of $\boldsymbol{t}$ to the lattice. We will consider the following problem:

GAPCVP$_{\alpha,\beta}$
Input: A lattice $\Lambda := \Lambda(\boldsymbol{B})$ and a target vector $\boldsymbol{t} \in \mathbb{R}^n$
Goal: Distinguish the following cases:
- **YES**. One has $d(\boldsymbol{t}, \Lambda) \leq \alpha$
- **NO**. One has $d(\boldsymbol{t}, \Lambda) > \beta$

We will work towards the following claim:

Theorem 6.2. *For a small enough constant* $c > 0$, GAPCVP$_{c,\sqrt{n}} \in$ **NP** ∩ **coNP**.

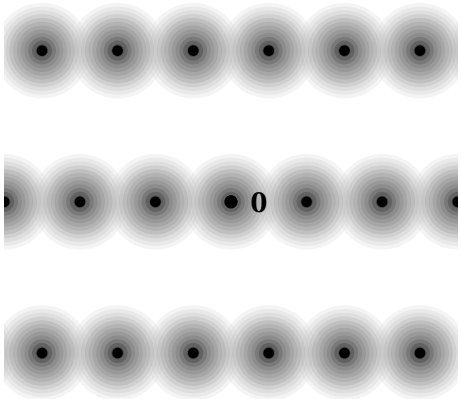
Of course one can scale the lattice and both parameters without changing the complexity, but the particular parameters of $\alpha = c$ and $\beta = \sqrt{n}$ will be convenient for the Fourier view that we will use.

6.2.1 The shifted discrete Gaussian

Given a full rank lattice $\Lambda \subseteq \mathbb{R}^n$, we define the function $F : \mathbb{R}^n \to \mathbb{R}$ with

$$F(\boldsymbol{x}) := \frac{\rho_1(\boldsymbol{x} + \Lambda)}{\rho_1(\Lambda)} \quad \forall \boldsymbol{x} \in \mathbb{R}^n,$$

where $\rho_1(\boldsymbol{x}) = e^{-\pi \|\boldsymbol{x}\|_2^2}$ is the discrete Gaussian from Section 4.2. Note that F is Λ-periodic with $0 < F(\boldsymbol{x}) \leq 1$ for all $\boldsymbol{x} \in \mathbb{R}^n$ (see Lemma 4.22). Intuitively it is not hard to imagine that for points $\boldsymbol{x}$ that are close to the lattice, $F(\boldsymbol{x})$ is large and for points that are far from the lattice, $F(\boldsymbol{x})$ is tiny. A visualization for F in a 2-dimensional lattice can be found below.



We want to formalize the intuition that F can be used to distinguish points that are close to the lattice from points that are far from the lattice:

Lemma 6.3. *Let* $\Lambda \subseteq \mathbb{R}^n$ *be a full rank lattice. For* $\boldsymbol{x} \in \mathbb{R}^n$*, the following holds:*

(i) If $d(\boldsymbol{x},\Lambda) > \sqrt{n}$ then $F(\boldsymbol{x}) \le 2^{-n}$.

(ii) One has $F(\boldsymbol{x}) \ge e^{-\pi d(\boldsymbol{x},\Lambda)^2}$.

Proof. Claim (i) follows from Lemma 4.24 because

$$\rho_1(\Lambda+\boldsymbol{x}) \overset{(\Lambda+\boldsymbol{x})\setminus\sqrt{n}B_2^n=\emptyset}{=} \rho_1\big((\Lambda+\boldsymbol{x})\setminus\sqrt{n}B_2^n\big) \overset{\text{Lem 4.24}}{\le} 2^{-n}\rho_1(\Lambda)$$

Next, consider (ii). Since F is Λ-periodic we may assume that the closest lattice point to $\boldsymbol{x}$ is the origin and so $\|\boldsymbol{x}\|_2 = d(\boldsymbol{x},\Lambda)$. Then it suffices to prove:
Claim. *For any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any $\boldsymbol{x} \in \mathbb{R}^n$ one has $\rho_1(\Lambda+\boldsymbol{x}) \ge e^{-\pi\|\boldsymbol{x}\|_2^2}\rho_1(\Lambda)$.*
Proof of Claim. The proof uses a trick that appears elsewhere when dealing with Gaussian densities. Naively one might be tempted to show that *pointwise* for every $\boldsymbol{y}$ one has $\rho_1(\boldsymbol{x}+\boldsymbol{y}) \ge e^{-\pi\|\boldsymbol{x}\|_2^2}\rho_1(\boldsymbol{y})$. However, this is false for $\langle \boldsymbol{x},\boldsymbol{y}\rangle \gg 0$! But one can argue that this is true if we *average* over the two choices of $\pm\boldsymbol{y}$. Formally, we write

$$\begin{aligned}
\sum_{\boldsymbol{y}\in\Lambda}\rho_1(\boldsymbol{x}+\boldsymbol{y}) &= \frac{1}{2}\sum_{\boldsymbol{y}\in\Lambda}\left(e^{-\pi\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}+e^{-\pi\|\boldsymbol{x}+\boldsymbol{y}\|_2^2}\right)\\
&= e^{-\pi\|\boldsymbol{x}\|_2^2}\sum_{\boldsymbol{y}\in\Lambda}e^{-\pi\|\boldsymbol{y}\|_2^2}\cdot\underbrace{\frac{1}{2}\left(e^{-2\pi\langle\boldsymbol{x},\boldsymbol{y}\rangle}+e^{2\pi\langle\boldsymbol{x},\boldsymbol{y}\rangle}\right)}_{\ge 1}\\
&\ge e^{-\pi\|\boldsymbol{x}\|_2^2}\cdot\rho_1(\Lambda)
\end{aligned}$$

Here we use that $\frac{1}{2}(z+\frac{1}{z}) \ge 1$ for all $z>0$. □

6.2.2 Approximating the function F

Algorithmically we have the problem how to compactly represent the function F. For that purpose we consider the *Fourier series representation* of the function F. Recall that given a Λ-periodic function f and $\boldsymbol{w} \in \Lambda^*$ we defined $\tilde{f}(\boldsymbol{w}) = \mathbb{E}_{\boldsymbol{x}\sim\mathcal{P}(\boldsymbol{B})}[f(\boldsymbol{x})\cdot e^{-2\pi i\langle\boldsymbol{w},\boldsymbol{x}\rangle}]$, see Section 4.1.2.

Lemma 6.4. *For any full-rank lattice $\Lambda \subseteq \mathbb{R}^n$, the function $F(\boldsymbol{x}) := \frac{\rho_1(\boldsymbol{x}+\Lambda)}{\rho_1(\Lambda)}$ has Fourier series coefficients*

$$\tilde{F}(\boldsymbol{w}) = \frac{\rho_1(\boldsymbol{w})}{\rho_1(\Lambda^*)} \quad \forall\,\boldsymbol{w}\in\Lambda^*$$

Proof. Let us abbreviate $G(\boldsymbol{x}) := \rho_1(\boldsymbol{x}+\Lambda)$ so that $F(\boldsymbol{x}) = \frac{G(\boldsymbol{x})}{G(\boldsymbol{0})}$. In Lemma 4.15 we have proven that for $\boldsymbol{w} \in \Lambda^*$ one has

$$\tilde{G}(\boldsymbol{w}) \overset{\text{Lem } 4.15}{=} \det(\Lambda^*) \cdot \hat{\rho}_1(\boldsymbol{w}) \overset{\text{Lem } 4.18}{=} \det(\Lambda^*) \cdot \rho_1(\boldsymbol{w}) \quad (*)$$

Note that indeed $\tilde{G}(\boldsymbol{w})$ is a *Fourier series coefficient* and $\hat{\rho}_1(\boldsymbol{w})$ is a *Fourier coefficient* where $\hat{\rho}_1(\boldsymbol{w}) = \rho_1(\boldsymbol{w})$ by Lemma 4.18. Next, the Fourier series representation (Lemma 4.6) of the function G gives

$$G(\boldsymbol{0}) = \sum_{\boldsymbol{w}\in\Lambda^*} \tilde{G}(\boldsymbol{w}) \cdot \underbrace{e^{2\pi i \langle \boldsymbol{w}, \boldsymbol{0}\rangle}}_{=1} \overset{(*)}{=} \det(\Lambda^*) \cdot \rho_1(\Lambda^*) \quad (**)$$

Then by linearity of the Fourier series coefficient we have

$$\tilde{F}(\boldsymbol{w}) = \frac{\tilde{G}(\boldsymbol{w})}{G(\boldsymbol{0})} \overset{(*)+(**)}{=} \frac{\det(\Lambda^*)\cdot\rho_1(\boldsymbol{w})}{\det(\Lambda^*)\cdot\rho_1(\Lambda^*)} = \frac{\rho_1(\boldsymbol{w})}{\rho_1(\Lambda^*)}$$

as claimed. □

It is worth noting that $\tilde{F}(\boldsymbol{w}) \in \mathbb{R}_{\geq 0}$ for all $\boldsymbol{w} \in \Lambda^*$ and $\sum_{\boldsymbol{w}\in\Lambda^*} \tilde{F}(\boldsymbol{w}) = 1$. In other words the coefficients $\tilde{F}(\boldsymbol{w})$ define a *probability distribution* on the dual lattice Λ^* and that distribution is precisely the discrete Gaussian on the dual lattice, i.e. $\mathcal{D}_1(\Lambda^*)$ as defined in Section 4.4.2. Then the Fourier series representation of F is

$$\begin{aligned} F(\boldsymbol{x}) &= \sum_{\boldsymbol{w}\in\Lambda^*} \tilde{F}(\boldsymbol{w}) \cdot e^{2\pi i\langle \boldsymbol{w},\boldsymbol{x}\rangle} \\ &= \sum_{\boldsymbol{w}\in\Lambda^*} \tilde{F}(\boldsymbol{w}) \cdot \cos(2\pi\langle \boldsymbol{w},\boldsymbol{x}\rangle) \\ &= \mathop{\mathbb{E}}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda^*)} \left[\cos(2\pi\langle \boldsymbol{w},\boldsymbol{x}\rangle)\right] \end{aligned} \tag{6.1}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. Here we used in the 2nd step that $F(\boldsymbol{x})$ and $\tilde{F}(\boldsymbol{w})$ are real, so the imaginary parts must cancel out and $\mathrm{Re}(e^{2\pi i\langle \boldsymbol{w},\boldsymbol{x}\rangle}) = \cos(2\pi\langle \boldsymbol{w},\boldsymbol{x}\rangle) \in [-1,1]$[1]. The representation of F in Eq 6.1 gives rise to a natural idea to represent F in a compact way: simply sample a polynomial number of dual vectors $\boldsymbol{w} \sim \mathcal{D}_1(\Lambda^*)$ and use a "sparse Fourier series representation".

We will also use the following standard fact:

Theorem 6.5 (Chernov-Hoeffding)**.** *Let $X_1,\ldots,X_N \in [a,b]$ be independent random variables. Then for any $\varepsilon > 0$, the sum $Y := X_1 + \ldots + X_N$ satisfies*

$$\Pr\left[|Y - \mathbb{E}[Y]| \geq N\varepsilon\right] \leq 2\exp\left(-\frac{N\varepsilon^2}{(b-a)^2}\right)$$

[1] One can also see explicitly where the cancellation occurs: observe that $\tilde{F}(-\boldsymbol{w}) = \tilde{F}(\boldsymbol{w})$ and $\mathbb{E}_{\sigma\sim\{-1,1\}}[e^{2\pi i\langle \sigma\boldsymbol{w},\boldsymbol{x}\rangle}] = \cos(2\pi\langle \boldsymbol{w},\boldsymbol{x}\rangle)$.

Recall that from Lemma 4.24 we already know that for every lattice Λ one has $\Pr_{\boldsymbol{x}\sim\mathcal{D}_1(\Lambda)}[\|\boldsymbol{x}\|_2 > \sqrt{n}] \le 2^{-n}$. We can also cover a more general regime (note that the following inequality is only non-trivial if $\lambda \ge \Theta(\sqrt{n})$ with a large enough implicit constant):

Lemma 6.6. *There is a universal constant $c > 0$ so that for any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ and any $\lambda \ge 0$ one has $\Pr_{\boldsymbol{x}\sim\mathcal{D}_1(\Lambda)}[\|\boldsymbol{x}\|_2 \ge \lambda\sqrt{n}] \le 8^n \exp(-c\lambda^2 n)$.*

Proof. Take an ε-net $\mathcal{S} \subseteq S^{n-1}$ of size[2] $|\mathcal{S}| \le (\frac{4}{\varepsilon})^n$. Note that for each $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ there is a $\boldsymbol{y} \in \mathcal{S}$ with $\|\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2} - \boldsymbol{y}\|_2 \le \varepsilon$ and so $\|\boldsymbol{x}\|_2 = \langle \boldsymbol{x}, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\rangle = \langle \boldsymbol{x}, \boldsymbol{y} - (\boldsymbol{y} - \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2})\rangle \le \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \varepsilon\|\boldsymbol{x}\|_2$. Then setting $\varepsilon := \frac{1}{2}$, we obtain $\|\boldsymbol{x}\|_2 \le 2\max\{\langle \boldsymbol{x}, \boldsymbol{y}\rangle : \boldsymbol{y} \in \mathcal{S}\}$ while $|\mathcal{S}| \le 8^n$. Hence

$$\Pr_{\boldsymbol{x}\sim\mathcal{D}_1(\Lambda)}[\|\boldsymbol{x}\|_2 \ge \lambda\sqrt{n}] \le \Pr_{\boldsymbol{x}\sim\mathcal{D}_1(\Lambda)}\left[\max_{\boldsymbol{y}\in\mathcal{S}} \langle \boldsymbol{x}, \boldsymbol{y}\rangle \ge \frac{\lambda}{2}\sqrt{n}\right] \le 8^n \cdot \exp(-c\lambda^2 n)$$

using the subgaussianity of the discrete Gaussian (Lemma 4.38). □

Lemma 6.7 (Pointwise approximation Lemma)**.** *Let $\Lambda = \Lambda(\boldsymbol{B}) \subseteq \mathbb{R}^n$ be a full-rank lattice with $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ and let $F : \mathbb{R}^n \to \mathbb{R}$ be the function with $F(\boldsymbol{x}) := \frac{\rho_1(\boldsymbol{x}+\Lambda)}{\rho_1(\Lambda)}$. Set $N := \Theta(\frac{n^2}{\delta^3}\log(2+L))$ where $L := \max_{j=1,\dots,n}\|\boldsymbol{B}^j\|_2$ and sample $\boldsymbol{w}_1, \dots, \boldsymbol{w}_N \sim \mathcal{D}_1(\Lambda^*)$ independently and set $W := (\boldsymbol{w}_1, \dots, \boldsymbol{w}_N)$. Then with probability at least $1 - 2^{-n}$, the function*

$$F_W(\boldsymbol{x}) := \frac{1}{N}\sum_{i=1}^{N} \cos(2\pi\,\langle \boldsymbol{x}, \boldsymbol{w}_i\rangle)$$

satisfies $|F_W(\boldsymbol{x}) - F(\boldsymbol{x})| \le \delta$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

We first prove the statement for a *single* vector $\boldsymbol{x}$:

Claim I. *For a fixed point $\boldsymbol{x} \in \mathbb{R}^n$ one has $\Pr[|F_W(\boldsymbol{x}) - F(\boldsymbol{x})| \ge \frac{\delta}{2}] \le 2\exp(-\frac{N\delta^2}{16})$.*

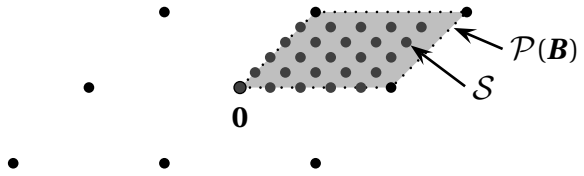
Proof of Claim I. As we draw $\boldsymbol{w}_1, \dots, \boldsymbol{w}_N$ independently, we know that the random variables $X_i := \cos(2\pi\,\langle \boldsymbol{w}_i, \boldsymbol{x}\rangle)$ are independent for $i = 1, \dots, N$. Moreover $-1 \le X_i \le 1$ and $\mathbb{E}[X_i] = F(\boldsymbol{x})$ as we know from (6.1). Then by the Chernov-Hoeffding bound (Theorem 6.5) we obtain

$$\Pr\left[\left|\sum_{i=1}^{N} X_i - N\cdot F(\boldsymbol{x})\right| \ge N\frac{\delta}{2}\right] \le 2\exp\left(-\frac{N\delta^2}{16}\right)$$

Then dividing the expression inside Pr[..] by N gives the claim. □

Next, as the functions F and F_W are Λ-periodic, so it suffices show the claim for all $\boldsymbol{x} \in \mathcal{P}(\boldsymbol{B})$. For a parameter $T \in \mathbb{N}$ that we determine later we consider the set $\mathcal{S} := \frac{1}{T}\cdot\Lambda \cap \mathcal{P}(\boldsymbol{B})$ which is a fine grid inside the fundamental parallelepiped.

[2] Recall that an ε-net is a set $\mathcal{S} \subseteq S^{n-1}$ so that for each $\boldsymbol{x} \in S^{n-1}$, there is a $y \in \mathcal{S}$ with $\|\boldsymbol{x}-\boldsymbol{y}\|_2 \le \varepsilon$.



We can see that $|\mathcal{S}| = T^n$. Using a loose estimate, we note that every point in $\mathcal{P}(\boldsymbol{B})$ is at distance at most $\frac{nL}{T}$ to a point in $\mathcal{S}$. Then with a union bound over all points in the grid $\mathcal{S}$ we have

$$\Pr\left[\forall \boldsymbol{x} \in \mathcal{S} : |F_W(\boldsymbol{x}) - F(\boldsymbol{x})| \leq \frac{\delta}{2}\right] \overset{\text{Claim I}}{\geq} 1 - T^n \cdot 2\exp\left(-\frac{N\delta^2}{16}\right) \tag{6.2}$$

Moreover it will be useful to have a guaranteed upper bound on the length of the vectors $\boldsymbol{w}_i$. In fact, using Lemma 6.6 we know for example that

$$\Pr\left[\forall i \in [N] : \|\boldsymbol{w}_i\|_2 \leq C'\sqrt{n\log(N)}\right] \geq 1 - \frac{1}{2\cdot 2^n N} \tag{6.3}$$

for some constant $C' > 0$. The next step is to show that the functions F and F_W are sufficiently smooth so that controlling their value at the points in the grid $\mathcal{S}$ suffices. Here it will be helpful to know that the vectors in W are not too long. Recall that a function $G : \mathbb{R}^n \to \mathbb{R}$ is called *s-Lipschitz*, if $|G(\boldsymbol{x}) - G(\boldsymbol{y})| \leq s \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_2$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

Claim II. *For any outcome of W with $\|\boldsymbol{w}_i\|_2 \leq R$ for all $i \in [N]$, the function F_W is $2\pi R$-Lipschitz.*

Proof of Claim II. Let $\boldsymbol{x}, \boldsymbol{d} \in \mathbb{R}^n$. We can bound

$$\begin{aligned} |F_W(\boldsymbol{x}+\boldsymbol{d}) - F_W(\boldsymbol{x})| &\leq \frac{1}{N}\sum_{i=1}^{N} \left|\cos(2\pi\langle \boldsymbol{w}_i, \boldsymbol{x}\rangle + 2\pi\langle \boldsymbol{w}_i, \boldsymbol{d}\rangle) - \cos(2\pi\langle \boldsymbol{w}_i, \boldsymbol{x}\rangle)\right| \\ &\leq \frac{1}{N}\sum_{i=1}^{N} 2\pi \cdot |\langle \boldsymbol{w}_i, \boldsymbol{d}\rangle| \leq 2\pi R \cdot \|\boldsymbol{d}\|_2 \end{aligned}$$

using that the derivative of $\cos(\cdot)$ is at most 1 in absolute value.

Claim III. *The function F is $O(1)$-Lipschitz.*

Proof of Claim III. Let $\boldsymbol{x}, \boldsymbol{d} \in \mathbb{R}^n$. Similar to the proof of Claim II we bound

$$\begin{aligned} |F(\boldsymbol{x}+\boldsymbol{d}) - F(\boldsymbol{x})| &\leq \mathop{\mathbb{E}}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda^*)}\left[\left|\cos(2\pi\langle \boldsymbol{w}, \boldsymbol{x}+\boldsymbol{d}\rangle) - \cos(2\pi\langle \boldsymbol{w}, \boldsymbol{x}\rangle)\right|\right] \\ &\leq 2\pi \mathop{\mathbb{E}}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda^*)}\left[|\langle \boldsymbol{w}, \boldsymbol{d}\rangle|\right] \leq O(\|\boldsymbol{d}\|_2) \end{aligned}$$

Here we use in the last step the fact that $\|\langle \boldsymbol{w}, \boldsymbol{d}\rangle\|_{\psi_2} \leq O(\|\boldsymbol{d}\|_2)$ by Lemma 4.38. Also for any mean-zero random variable X one has $\mathbb{E}[|X|] \leq O(\|X\|_{\psi_2})$ by Lemma 4.35. □

Now we can put everything together. We will need to choose N large enough so that both events in (6.2) and (6.3) happen together with probability at least $1-2^{-n}$. In order to determine the parameters, suppose the events in (6.2) and (6.3) indeed are both true. Then fix any $\boldsymbol{x} \in \mathcal{P}(\boldsymbol{B})$ and let $\boldsymbol{y} \in \mathcal{S}$ be the closest point to $\boldsymbol{x}$, which means that $\|\boldsymbol{x}-\boldsymbol{y}\|_2 \leq \frac{nL}{T}$. Then we can see that

$$\begin{aligned} |F_W(\boldsymbol{x}) - F(\boldsymbol{x})| &\leq |F_W(\boldsymbol{y}) + F(\boldsymbol{y})| + |F_W(\boldsymbol{x}) - F_W(\boldsymbol{y})| + |F(\boldsymbol{x}) - F(\boldsymbol{y})| \\ &\overset{\text{Claim II+III}}{\leq} \frac{\delta}{2} + O\left(\sqrt{n\log N}\cdot\frac{n}{T}L\right) \overset{!}{\leq} \delta \end{aligned}$$

We can see that it suffices to choose $T := \Theta(\frac{n^2}{\delta}\sqrt{\log N}\cdot\lceil L\rceil)$. We also see that we require $N \geq \Theta(\frac{n}{\delta^2}\log(2T))$ to satisfy (6.2). A generous choice satisfying this would be $N := \Theta(\frac{n^2}{\delta^3}\log(L+2))$ as claimed.

6.2.3 The verifier

There is an additional technical lemma dealing with the concentration of discrete Gaussian samples. One could have phrased the lemma in more generality and state that for any distribution $\mathcal{D}$ that is $O(1)$-subgaussian, sampling $\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N \sim \mathcal{D}$ independently will with high probability result in a matrix with $\sum_{i=1}^N \boldsymbol{w}_i\boldsymbol{w}_i^T \preceq O(N)$, assuming $N \geq \Theta(n^2)$. Also we should point out that there are matrix concentration techniques that give much finer bounds than the one we obtain here. However, this one suffices for our purpose and we refer the reader to Vershynin's textbook [Ver19] for an extensive and extremely readable account on matrix concentration. We use the following fact:

Lemma 6.8. *For any lattice $\Lambda \subseteq \mathbb{R}^n$ and any $\boldsymbol{y} \in S^{n-1}$ one has $\mathbb{E}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda)}[\langle \boldsymbol{w},\boldsymbol{y}\rangle^2] \leq \frac{1}{2\pi}$.*

We will not give a proof of Lemma 6.8 here. However note that we do know that $\|\langle \boldsymbol{w},\boldsymbol{y}\rangle\|_{\psi_2} \leq O(1)$ by Lemma 4.38 and so $\mathbb{E}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda)}[\langle \boldsymbol{w},\boldsymbol{y}\rangle^2] \leq O(1)$ for some unspecified constant by Lemma 4.35. It is also worth noting that the constant of $\frac{1}{2\pi}$ is tight and matches the value of the continuous Gaussian with density function ρ_1. Now to the actual matrix concentration claim (which we will apply to the dual lattice Λ^* but we keep it general):

Lemma 6.9. *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice. Sample $\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N \sim \mathcal{D}_1(\Lambda)$ where $N \geq C'n^2$ and $C' > 0$ is large enough. Then $\sum_{i=1}^N \boldsymbol{w}_i\boldsymbol{w}_i^T \preceq 3N\boldsymbol{I}_n$ with probability at least $1-2^{-2n}$.*

Proof. Consider an ε-net $\mathcal{S} \subseteq S^{n-1}$ of size $|\mathcal{S}| \leq (\frac{4}{\varepsilon})^n$. For any symmetric matrix

$\boldsymbol{M} \in \mathbb{R}^{n\times n}$, the *maximum singular value* is of the form $\|\boldsymbol{M}\|_{\mathrm{op}} = \max\{\langle \boldsymbol{M}, \boldsymbol{x}\boldsymbol{x}^T\rangle : \boldsymbol{x} \in S^{n-1}\}$. Suppose $\boldsymbol{x} \in S^{n-1}$ attains that maximum and let $\boldsymbol{y} := \boldsymbol{x} + \boldsymbol{d} \in \mathcal{S}$ with $\|\boldsymbol{d}\|_2 \le \varepsilon$ be the net point closest to $\boldsymbol{x}$, then

$$\begin{aligned}\|\boldsymbol{M}\|_{\mathrm{op}} &= \langle \boldsymbol{M}, \boldsymbol{x}\boldsymbol{x}^T\rangle = \langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle - \langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T - \boldsymbol{x}\boldsymbol{x}^T\rangle \\ &\le \langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle + 2|\langle \boldsymbol{M}, \boldsymbol{x}\boldsymbol{d}^T\rangle| + |\langle \boldsymbol{M}, \boldsymbol{d}\boldsymbol{d}^T\rangle| \le |\langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle| + 3\varepsilon\|\boldsymbol{M}\|_{\mathrm{op}}\end{aligned}$$

and so $\|\boldsymbol{M}\|_{\mathrm{op}} \le \frac{1}{1-3\varepsilon}|\langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle|$. Setting $\varepsilon := \frac{1}{9}$ we know that $\|\boldsymbol{M}\|_{\mathrm{op}} \le \frac{3}{2}\max\{|\langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle| : \boldsymbol{y} \in \mathcal{S}\}$ where $|\mathcal{S}| \le 36^n$.

Next, we would like to apply the Chernov Hoeffding bound to control the error in some direction $\boldsymbol{y} \in \mathcal{S}$. However there is a problem: we do not have a guaranteed upper bound on $\|\boldsymbol{w}_i\|_2$ and Chernov Hoeffding requires that the random variables to be in a bounded interval. So we imagine a 2-stage random experiment: for a parameter $\lambda \ge 0$ that we choose later, we sample $\boldsymbol{w}_1, \dots, \boldsymbol{w}_N \sim \mathcal{D}_1(\Lambda)$, but we define *truncated* vectors

$$\tilde{\boldsymbol{w}}_i := \begin{cases} \boldsymbol{w}_i & \text{if } \|\boldsymbol{w}_i\|_2 \le \lambda\sqrt{n} \\ \boldsymbol{0} & \text{otherwise.}\end{cases}$$

By Lemma 6.6 we know that for each $i \in [N]$,

$$\Pr\left[\boldsymbol{w}_i = \tilde{\boldsymbol{w}}_i\right] \ge 1 - 8^n \exp(-c\lambda^2 n)$$

Next, we prove a concentration bound for a single direction with respect to the truncated vectors:

Claim I. *For any* $\boldsymbol{y} \in S^{n-1}$ *one has* $\Pr[\langle \sum_{i=1}^N \tilde{\boldsymbol{w}}_i \tilde{\boldsymbol{w}}_i^T, \boldsymbol{y}\boldsymbol{y}^T\rangle \ge 2N] \le 2\exp\left(-\frac{N}{\lambda^2 n}\right)$.

Proof of Claim I. Consider the random variables $X_i := \langle \tilde{\boldsymbol{w}}_i \tilde{\boldsymbol{w}}_i^T, \boldsymbol{y}\boldsymbol{y}^T\rangle$ with sum $X := \sum_{i=1}^N X_i$. The truncated random vectors $\tilde{\boldsymbol{w}}_1, \dots, \tilde{\boldsymbol{w}}_N$ are still independent meaning that also the random variables $X_1, \dots, X_N$ are independent and moreover $0 \le X_i \le \lambda^2 n$. We also note that

$$\mathbb{E}[X] = N \cdot \mathop{\mathbb{E}}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda)}[\langle \tilde{\boldsymbol{w}}, \boldsymbol{y}\rangle^2] \le N \cdot \mathop{\mathbb{E}}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda)}[\langle \boldsymbol{w}, \boldsymbol{y}\rangle^2] \overset{\text{Lem 6.8}}{\le} \frac{N}{2\pi} \le N.$$

Hence by the Chernov Hoeffding bound (Theorem 6.5) we obtain

$$\Pr[X \ge \mathbb{E}[X] + N] \le 2\exp\left(-\frac{N}{(\lambda\sqrt{n})^2}\right) \qquad \square$$

Now we can finish the main argument. For $\boldsymbol{M} := \sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{w}_i^T$ we have

$$\begin{aligned}\Pr\left[\|\boldsymbol{M}\|_{\mathrm{op}} \geq 3N\right] &\leq \Pr\left[\exists \boldsymbol{y} \in \mathcal{S} : |\langle \boldsymbol{M}, \boldsymbol{y}\boldsymbol{y}^T\rangle| \geq 2N\right] \\ &\leq \sum_{i=1}^{N} \Pr[\tilde{\boldsymbol{w}}_i \neq \boldsymbol{w}_i] + \sum_{\boldsymbol{y}\in\mathcal{S}} \Pr\Big[\sum_{i=1}^{N} \langle \tilde{\boldsymbol{w}}_i \tilde{\boldsymbol{w}}_i^T, \boldsymbol{y}\boldsymbol{y}^T\rangle \geq 2N\Big] \\ &= N \cdot 8^n \exp(-c\lambda^2 n) + 36^n \cdot 2\exp\Big(-\frac{N}{\lambda^2 n}\Big) \leq 2^{-2n}\end{aligned}$$

if we choose $\lambda := C'$ and $N = (C')^3 n^2$ with a large enough constant $C' > 0$. □

We restate and prove the main result of this chapter:

Theorem (Theorem 6.2)**.** *For a small enough constant $c > 0$,* $\mathrm{GAPCVP}_{c,\sqrt{n}} \in$ **NP** ∩ **coNP**.

Proof. Again, it is easy to see that $\mathrm{GAPCVP}_{c,\sqrt{n}} \in$ **NP** — the certificate that we are in the YES case is the lattice vector $\boldsymbol{x} \in \Lambda$ with $\|\boldsymbol{x} - \boldsymbol{t}\|_2 \leq c$. Again we use the fact that one can verify whether one has $\boldsymbol{x} \in \Lambda$ in polynomial time.

We use the following verifier to show that $\mathrm{GAPCVP}_{c,\sqrt{n}} \in$ **coNP**:

Input: Lattice $\Lambda(\boldsymbol{B})$, target vector $\boldsymbol{t} \in \mathbb{R}^n$
Certificate: Vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N \in \mathbb{R}^n$ for $N := \Theta(\frac{n^2}{\delta^3}\log(2+L))$ where $\delta := \frac{1}{n}$
Verifier: Accept if all of the following conditions are satisfied:
(A) One has $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N \in \Lambda^*$
(B) One has $\sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{w}_i^T \preceq 3N$
(C) One has $F_W(\boldsymbol{t}) < \frac{1}{2}$

Claim I. *If $d(\boldsymbol{t}, \Lambda) \leq c$ then the verifier rejects any certificate.*
Proof of Claim I. Assume that (A)+(B) are satisfied; we prove that then (C) is false and so the verifier indeed rejects. Let $\boldsymbol{x} \in \Lambda$ be the lattice point with $\|\boldsymbol{t} - \boldsymbol{x}\|_2 \leq c$. Using that $\cos(z) \geq 1 - \frac{1}{2}z^2$ for all $z \in \mathbb{R}$, we estimate that

$$\begin{aligned}F_W(\boldsymbol{t}) &\overset{\Lambda\text{-periodic}}{=} F_W(\boldsymbol{t}-\boldsymbol{x}) = \frac{1}{N}\sum_{j=1}^{N} \cos(2\pi\langle \boldsymbol{t}-\boldsymbol{x}, \boldsymbol{w}_j\rangle) \geq 1 - \frac{4\pi^2}{2N}\sum_{j=1}^{N}\langle \boldsymbol{t}-\boldsymbol{x}, \boldsymbol{w}_j\rangle^2 \\ &= 1 - \frac{4\pi^2}{2N}\underbrace{\langle \sum_{j=1}^{n} \boldsymbol{w}_j \boldsymbol{w}_j^T, (\boldsymbol{t}-\boldsymbol{x})(\boldsymbol{t}-\boldsymbol{x})^T\rangle}_{\leq 3N\cdot d(\boldsymbol{t},\Lambda)^2} \geq 1 - 60\cdot d(\boldsymbol{t},\Lambda)^2 \overset{c \text{ small}}{>} \frac{9}{10}\end{aligned}$$

Hence (C) is indeed not true. □

Claim II. *If* $d(\boldsymbol{t},\Lambda) > \sqrt{n}$ *then there exists a certificate that the verifier accepts.*
Proof of Claim II. We draw $\boldsymbol{w}_1,\dots,\boldsymbol{w}_N \sim \mathcal{D}_1(\Lambda^*)$ independently at random and consider the two events

(a) $\sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{w}_i^T \preceq 3N$

(b) $|F_W(\boldsymbol{x}) - F(\boldsymbol{x})| \le \delta \ \forall \boldsymbol{x} \in \mathbb{R}^n$

We know that for $N := \Theta(\frac{n^2}{\delta^3}\log(2+L)) = \Theta(n^5\log(2+L))$, (a) happens with overwhelming probability by Lemma 6.9 and (b) happens with overwhelming probability by the Pointwise Approximation Lemma (Lemma 6.7). We fix any outcome satisfying both (a) and (b). Then

$$F_W(\boldsymbol{t}) \le F(\boldsymbol{t}) + \delta \overset{\text{Lem } 6.3}{\le} 2^{-n} + \delta \overset{\delta=\frac{1}{n}}{\le} \frac{2}{n}$$

as $d(\boldsymbol{t},\Lambda) > \sqrt{n}$. Then $(A),(B),(C)$ are clearly satisfied and the verifier accepts. □

Chapter 7

Learning With Errors

In this chapter, we want to discuss a *public key crypto system* that is based on a lattice problem. Recall that the goal for public key cryptography is that two players Alice and Bob can exchange encypted messages while all their communication may be public. In particular there will be two types of keys: the *private key*, known only to say Alice and the *public key* known to everyone.

We may want to revisit the Section 1.6 and wonder what actually went wrong with the Knapsack crypto system. The crypto system was based on (assumed) hardness of a rather particular type of a very sparse random Knapsack problem — which turned out to be solvable in polynomial time. This already show cases one problem: presumably any crypto system will need to somehow generate *random keys* — how would one ever be confident that the particular distribution is not again easy? Another issue is that we cannot even prove **NP** $\neq$ **P** (which is a claim about worst case instances), so we cannot realistically hope to *unconditionally* prove security of a public key crypto system. Then the next best option would be to prove an implication of the form

$$\begin{pmatrix}\text{crypto system } A \text{ using}\\ \text{distribution } \mathcal{D} \text{ can be broken}\end{pmatrix} \quad \Longrightarrow \quad \begin{pmatrix}\text{worst case instances}\\ \text{of presumably hard problem } B\\ \text{can be solved efficiently}\end{pmatrix}$$

Then security of the crypto system would be based on the *worst case hardness* of problem B. In fact, such a crypto system exists and it is called *Learning with Errors*. While the literature on this topic is vast, we will focus on reproducing the seminal result of Regev [Reg09b] where we will in some details from other sources. In particular we recommend the very readable survey of Regev [Reg]. In order to not loose the overview over the various reductions, the reader may consult Figure 7.1.

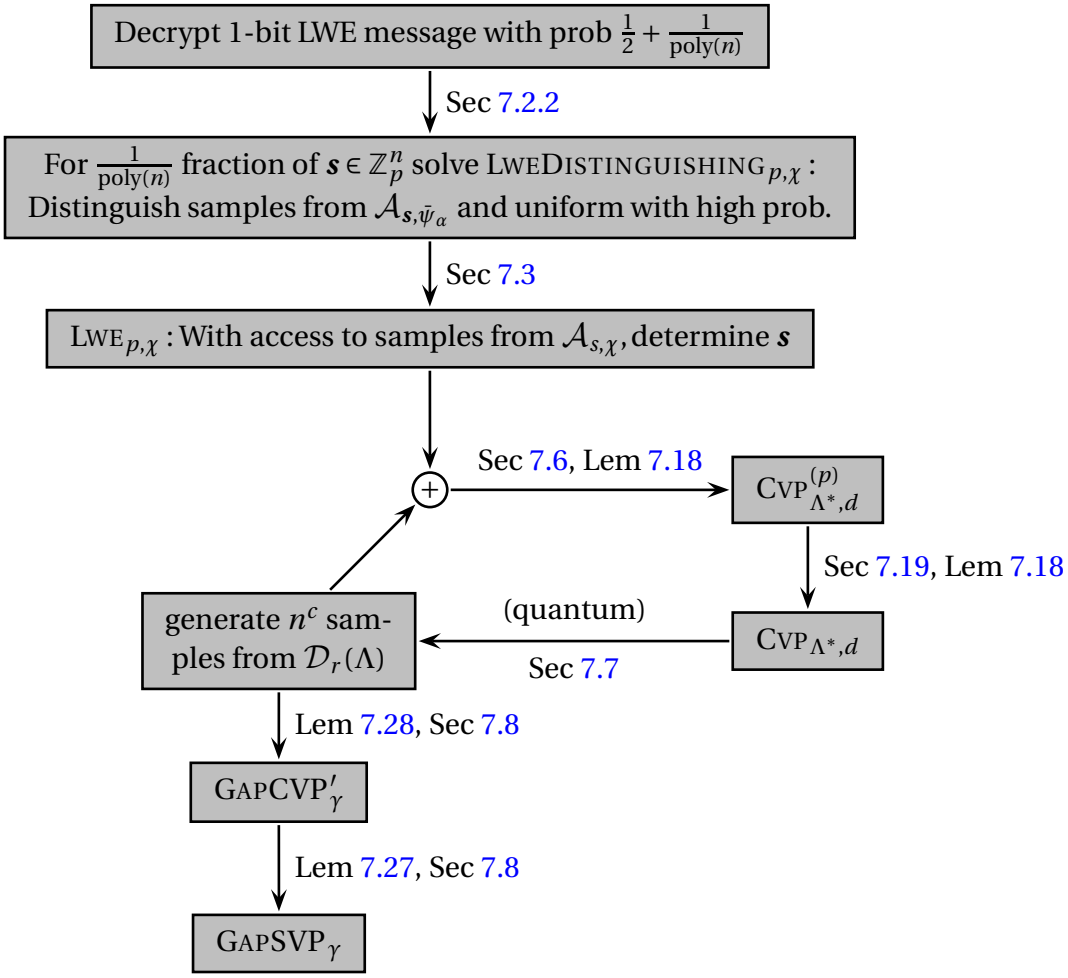


Figure 7.1: Overview over reductions.

7.1 The LWE crypto system

We will introduce a few basics that we need to describe the crypto system.

7.1.1 Preliminaries I

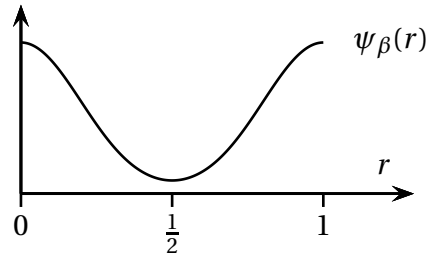
We will frequently use the *cyclic group* $\mathbb{Z}_p = \{0,\dots,p-1\}$ with addition modulo p, where $p \geq 1$ is an integer. For a set $A \subseteq \mathbb{R}$ and $x \in \mathbb{R}$, we will write $d(x,A) := \min\{|x-y| : y \in A\}$. Recall that for $s > 0$, the *discrete Gaussian* is the function

$$\rho_s : \mathbb{R}^n \to \mathbb{R}_{\geq 0} \quad \text{with} \quad \rho_s(\boldsymbol{x}) := e^{-\pi\|\boldsymbol{x}/s\|_2^2} \quad \forall \boldsymbol{x} \in \mathbb{R}^n,$$

as studied in Chapter 4. We abbreviate $\nu_s(\boldsymbol{x}) := \frac{\rho_s(\boldsymbol{x})}{\rho_s(\mathbb{R}^d)}$, where for a continuous set A, we write $\rho_s(A) = \int_A \rho_s(\boldsymbol{x})d\boldsymbol{x}$. Note that ν_s is scaled to be a continuous probability distribution (in fact it corresponds to the Gaussian distribution $N(\boldsymbol{0}, \frac{s^2}{2\pi}\boldsymbol{I}_n)$) and as $\rho_s(\mathbb{R}^n) = s^n$, the density function of that distribution is $\nu_s(\boldsymbol{x}) = s^{-n}e^{-\pi\|\boldsymbol{x}/s\|_2^2}$.

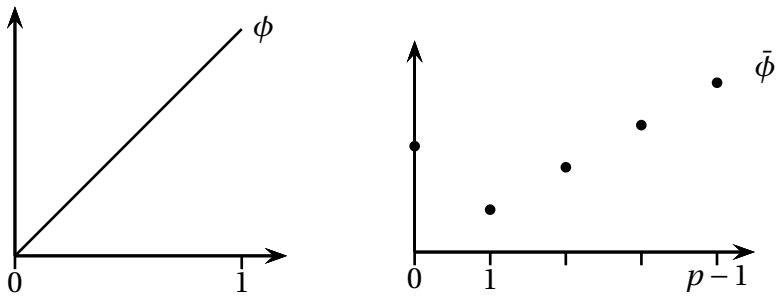
Gaussians modulo 1. For $x \in \mathbb{R}$, we write $x \mod 1$ as the quantity $x - \lfloor x \rfloor \in [0,1)$. We denote $\mathbb{T} := \mathbb{R}\setminus\mathbb{Z}$ as the cyclic group corresponding to $[0,1)$ with the addition modulo 1. For a parameter $\beta > 0$ we denote the distribution of $N(0, \frac{\beta^2}{2\pi}) \mod 1$ as ψ_β. The scaling by 2π might seem odd, but it is consistent with Chapter 4. By summing up the contribution one can easily see that the density of Ψ_β is precisely

$$\Psi_\beta(r) = \sum_{k\in\mathbb{Z}} \frac{1}{\beta}\exp\left(-\pi\left(\frac{r-k}{\beta}\right)^2\right) \quad \forall r \in [0,1)$$



Discretizations. Consider a distribution with density function $\phi : \mathbb{T} \to \mathbb{R}_{\geq 0}$ and let $X \sim \phi$. Then for a parameter p, the *discretization of* ϕ is the discrete probability distribution $\bar{\phi}$ corresponding to pX rounded to the nearest integer modulo p. Formally the probabilities of $\bar{\phi}$ are

$$\bar{\phi}(i) := \int_{(i-1/2)/p}^{(i+1/2)/p} \phi(x \bmod 1)\,dx \quad \forall i \in \{0,\dots,p-1\}$$

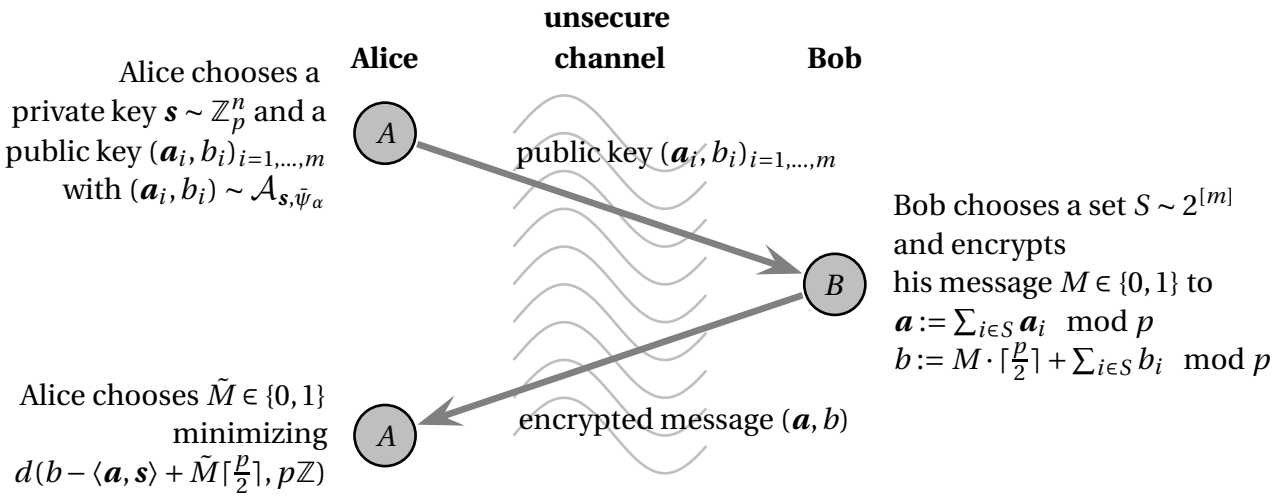


The LWE distribution. Fix a distribution $\chi : \mathbb{Z}_p \to \mathbb{R}_{\geq 0}$ and a vector $\boldsymbol{s} \in \mathbb{Z}_p^n$. Then we denote $\mathcal{A}_{\boldsymbol{s},\chi}$ as the distribution on tuples $\mathbb{Z}_p^n \times \mathbb{Z}_p$ that returns

$$\big(\boldsymbol{a}, \langle \boldsymbol{a}, \boldsymbol{s}\rangle + e \bmod p\big) \sim A_{\boldsymbol{s},\chi} \quad \text{where } \boldsymbol{a} \sim \mathbb{Z}_p^n,\ e \sim \chi \text{ independently}$$

7.1.2 The crypto system

Now we can finally describe the *LWE public key cryptoscheme* in which Bob wants to send a single bit $M \in \{0,1\}$ to Alice. The system depends on parameters $m, n, p \in \mathbb{N}$ and α for which we will make a choice later. Also note that in order to send more than one bit, one can reuse secret and public key.



Before we come to the formal analysis we want to give some intuition behind the crypto system. The public key consists of the m pairs $(\boldsymbol{a}_i, b_i)_{i\in[m]}$ with $b_i = \langle \boldsymbol{a}_i, \boldsymbol{s}\rangle + e_i$ where we should think of $e_i \sim \bar{\psi}_\alpha$ as an error/noise term. Hence if an adversary was intented to compute the secret key he/she would have to solve a

system of equations

$$\begin{aligned} \langle \boldsymbol{a}_1, \boldsymbol{s} \rangle &\approx b_1 \mod p \\ &\vdots \\ \langle \boldsymbol{a}_m, \boldsymbol{s} \rangle &\approx b_m \mod p \end{aligned}$$

where the ith equation holds up to an additive Gaussian error. One can prove that with high probability there will be a unique good approximate solution $\boldsymbol{s} \in \mathbb{Z}_p^n$. But while one can solve systems of exact equations modulo p using Gaussian elimination, this does not obviously extend to approximate equations which gives us hope that an adversary cannot determine the secret key $\boldsymbol{s}$.

Now consider the message $(\boldsymbol{a}, b)$ with $\boldsymbol{a} \equiv_p \sum_{i \in S} \boldsymbol{a}_i$ and $b \equiv_p \sum_{i \in S} \langle \boldsymbol{a}_i, s \rangle + \sum_{i \in S} e_i + M\lceil \frac{p}{2} \rceil$. Note that $\boldsymbol{a} \sim \mathbb{Z}_p^n$ is distributed uniformly and (approximately) one has

$$b \equiv_p (\text{uniform from } \mathbb{Z}_p) + \bar{\psi}_{\sqrt{\frac{m}{2}} \cdot \alpha} + M \left\lceil \frac{p}{2} \right\rceil$$

Of course, $\boldsymbol{a}$ and b are not chosen independently, but they are correlated. However, in order to understand the correlation it appears one needs to know the secret key $\boldsymbol{s}$. So without the secret key it seems impossible to isolate the term $M\lceil \frac{p}{2} \rceil$ if some uniform random choice from $\mathbb{Z}_p$ is added.

Now, consider the situation from Alice's point of view. Alice of course knows the secret key $\boldsymbol{s}$ and the message $(\boldsymbol{a}, b)$ and so she knows in particular the "correlation" $\langle \boldsymbol{a}, \boldsymbol{s} \rangle$. Hence she can compute

$$b - \langle \boldsymbol{a}, \boldsymbol{s} \rangle \equiv_p \bar{\psi}_{\sqrt{\frac{m}{2}} \cdot \alpha} + M \left\lceil \frac{p}{2} \right\rceil$$

As long as the standard deviation of the noise term $\bar{\psi}_{\sqrt{\frac{m}{2}} \cdot \alpha}$ (which is of the order $\sqrt{m} \cdot \alpha$) is $\ll \frac{p}{4}$, Alice will be able to isolate the value of $M \in \{0, 1\}$.

7.2 Correctness and Security of the LWE Crypto System

7.2.1 Correctness of LWE

First we prove that with a suitable choice of parameters, the encryption scheme is correct, meaning that Alice will be able to decode the bit that Bob is sending. Of course, these won't be the only parameters that work, but these are parameters that also make sense from an efficiency perspective as the size of the public key is $\Theta(mn \log(p))$ and each message has $\Theta(n \log(p))$ bits.

Theorem 7.1 (Correctness of LWE)**.** *Set* $\alpha := \frac{1}{\sqrt{n}\log^2(n)}$, $m := 2n\log(p)$, p *prime with* $n^2 \le p \le 2n^2$ *and* n *large enough. Then* $\Pr[M \neq \tilde{M}] \le n^{-\Theta(\log(n))}$.

Proof. Note that

$$b - \langle \boldsymbol{a}, \boldsymbol{s} \rangle \equiv_p \sum_{i \in S} (\underbrace{\langle \boldsymbol{a}_i, \boldsymbol{s} \rangle + e_i}_{=b_i} - \langle \boldsymbol{a}_i, \boldsymbol{s} \rangle) \equiv_p \sum_{i \in S} e_i$$

Hence in order for Alice to not make mistake during encryption it suffices if $|\sum_{i \in S} e_i| < \lfloor \frac{p}{4} \rfloor$. Then fixing any choice of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{Z}_p^n$, $M \in \{0,1\}$ and $S \subseteq [m]$ and just with the randomness over the choice of $e_i \sim \bar{\psi}_\alpha$ we have

$$\begin{aligned}
\Pr[M \neq \tilde{M}] &\le \Pr_{e_i \sim \bar{\psi}_\alpha} \left[d\Big(\sum_{i \in S} e_i, p\mathbb{Z} \Big) \ge \left\lfloor \frac{p}{4} \right\rfloor \right] \\
&\le \Pr_{x_i \sim N(0, \frac{\alpha^2}{2\pi})} \left[p \left| \sum_{i \in S} x_i \right| + |S| \ge \frac{p}{8} \right] \\
&\overset{m \le \frac{p}{16}}{\le} \Pr_{X \sim N(0, \frac{1}{2\pi}|S|\alpha^2)} \left[|X| > \frac{1}{16} \right] \le 2\exp\Big(-\Theta\Big(\frac{1}{m\alpha^2} \Big) \Big) \le \exp(-\Theta(\log^2(n)))
\end{aligned}$$

□

Basically, the proof says that as long as $\sqrt{m} \cdot \alpha \ll 1$, then a single choice of (M, S) will have a good chance to be decoded correctly.

7.2.2 Security of LWE

Next, we come to what is often called the "proof of security of LWE", though we will (for now) really just prove that an efficient algorithm to decrypt messages implies an algorithm to distinguish samples from $\mathcal{A}_{\boldsymbol{s}, \bar{\psi}_\alpha}$ from the uniform distribution on tuples $\mathbb{Z}_p^n \times \mathbb{Z}_p$. It will take a lot more effort to later derive further consequences from the existence of such a distinguisher. We introduce a statistical tool first.

Statistical distance and the left over hash lemma

If ϕ_1 and ϕ_2 are two continuous distributions on $\mathbb{R}^n$, then we define their *statistical distance* as

$$\Delta(\phi_1, \phi_2) := \int_{\mathbb{R}^n} |\phi_1(\boldsymbol{x}) - \phi_2(\boldsymbol{x})| d\boldsymbol{x}$$

Note that $0 \le \Delta(\phi_1, \phi_2) \le 2$. Similarly, if ϕ_1, ϕ_2 are distributions over some discrete set Q, then one defines $\Delta(\phi_1, \phi_2) := \sum_{\boldsymbol{x} \in Q} |\phi_1(\boldsymbol{x}) - \phi_2(\boldsymbol{x})|$. In either case,

this quantity is useful for the following fact: If we think of $A : \mathbb{R}^n \to \{0,1\}$ as the output behaviour of an algorithm, then $|\Pr_{\boldsymbol{x}\sim\phi_1}[A(\boldsymbol{x}) = 1] - \Pr_{\boldsymbol{x}\sim\phi_2}[A(\boldsymbol{x}) = 1]| \leq \frac{1}{2}\Delta(\phi_1, \phi_2)$.

We need an auxiliary result which will imply that the bit b in Bob's message is distributed approximately uniformly over $\mathbb{Z}_p$. The lemma that we are about to prove is called the "Left over Hash Lemma" and applies in more generality. We should point out that the Abelian group G in the lemma will simply be $\mathbb{Z}_p^n \times \mathbb{Z}_p$ in our application.

Lemma 7.2 (Left over Hash Lemma)**.** *Let $g_1, \ldots, g_m \in G$ elements of an Abelian group $(G, \oplus)$. Define the random variables*

$$\begin{aligned} X &\sim G \text{ uniformly} \\ Y &= \bigoplus_{i\in S} g_i \text{ where } S \subseteq \{1,\ldots,m\} \text{ uniformly} \end{aligned}$$

Then $\mathbb{E}_{g_1,\ldots,g_m\sim G}[\Delta(X,Y)] \leq \sqrt{|G|/2^m}$.

Proof. First, let us fix any choice for $g_1, \ldots, g_m \in G$. Then

$$\begin{aligned} \sum_{g\in G} \Pr[Y = g]^2 &\overset{\text{indep.}}{=} \Pr_{S_1,S_2\subseteq[m]}\Big[\bigoplus_{i\in S_1} g_i = \bigoplus_{i\in S_2} g_i\Big] \qquad (*) \\ &\leq \underbrace{\Pr_{S_1,S_2\subseteq[m]}[S_1 = S_2]}_{=(1/2)^m} + \Pr_{S_1,S_2\subseteq[m]:S_1\neq S_2}\Big[\bigoplus_{i\in S_1} g_i = \bigoplus_{i\in S_2} g_i\Big] \end{aligned}$$

Now, taking the expectation over the choice of $g_1, \ldots, g_m$ we see that[1] $\Pr_{g_1,\ldots,g_m\sim G}[\bigoplus_{i\in S_1} g_i = \bigoplus_{i\in S_2} g_i] = \frac{1}{|G|}$. Hence the expectation of $(*)$ will be

$$\mathop{\mathbb{E}}_{g_1,\ldots,g_m\sim G}\Big[\sum_{g\in G}\Pr[Y = g]^2\Big] \leq \Big(\frac{1}{2}\Big)^m + \frac{1}{|G|} \qquad (**)$$

Note that this shows that the L_2-norm of the distribution Y is not much bigger than the L_2-norm of the uniform distribution. The statistical distance on the other hand is a claim on the L_1-distance, hence we need the following claim:

Claim I. *For any $\boldsymbol{a} \in \mathbb{R}^n_{\geq 0}$ with $\|\boldsymbol{a}\|_1 = 1$ one has $\|\boldsymbol{a} - \frac{\mathbf{1}}{n}\|_1 \leq (\|\boldsymbol{a}\|_2^2 - \frac{1}{n})^{1/2}$.*

Proof of Claim I. We write

$$\Big\|\boldsymbol{a} - \frac{\mathbf{1}}{n}\Big\|_1 \leq \sqrt{n}\cdot\Big\|\boldsymbol{a} - \frac{\mathbf{1}}{n}\Big\|_2 = \sqrt{n}\cdot\Big(\|\boldsymbol{a}\|_2^2 \underbrace{- \frac{2}{n}\underbrace{\langle \boldsymbol{a}, \mathbf{1}\rangle}_{=1} + \underbrace{\Big\|\frac{\mathbf{1}}{n}\Big\|_2^2}_{=1/n}}_{=-1/n}\Big)^{1/2} = \sqrt{n}\cdot\Big(\|\boldsymbol{a}\|_2^2 - \frac{1}{n}\Big)^{1/2}$$

[1] This is where we use that $(G, \oplus)$ is an Abelian group. To be more precise we use that in an Abelian group, any equation $a + x = b$ has a unique solution $x \in G$.

Then we can finally bound the statistical distance as

$$\begin{aligned}
\mathop{\mathbb{E}}_{g_1,\dots,g_m\sim G}[\Delta(X,Y)] &\overset{\text{Def }\Delta}{=} \mathop{\mathbb{E}}_{g_1,\dots,g_m\sim G}\Big[\sum_{g\in G}\Big|\Pr[Y=g]-\frac{1}{|G|}\Big|\Big] \\
&\overset{\text{Claim I}}{\leq} \sqrt{|G|}\cdot\mathop{\mathbb{E}}_{g_1,\dots,g_m\sim G}\Big[\Big(\sum_{g\in G}\Pr[Y=g]^2-\frac{1}{|G|}\Big)^{1/2}\Big] \\
&\overset{\text{Jensen Ineq.}}{\leq} \sqrt{|G|}\cdot\Big(\mathop{\mathbb{E}}_{g_1,\dots,g_m\sim G}\Big[\sum_{g\in G}\Pr[Y=g]^2\Big]-\frac{1}{|G|}\Big)^{1/2} \overset{(**)}{\leq} \sqrt{|G|}\cdot\sqrt{\frac{1}{2^m}}
\end{aligned}$$

In the second-to-last inequality we use Jensen's Inequality with the concavity of $x\mapsto\sqrt{x}$. □

Security of LWE

In the following, let $\mathcal{U}$ denote the uniform distribution on tuples $\mathbb{Z}_q^n\times\mathbb{Z}$. Our goal is to show that if one was able to decode messages from the LWE crypto system then one could also solve the following problem for $\chi=\bar{\psi}_\alpha$:

LWEDISTINGUISHING$_{p,\chi}$.
Input: Access to a distribution $\mathcal{R}$ with $\mathcal{R}\in\{\mathcal{U},\mathcal{A}_{\boldsymbol{s},\chi}\}$ with $\chi:\{0,\dots,p-1\}$ (where the choice of $\boldsymbol{s}$ is not known to the algorithm)
Output: Determine whether $\mathcal{R}=\mathcal{U}$ or $\mathcal{R}=\mathcal{A}_{\boldsymbol{s},\chi}$.

The exact statement that we prove is as follows:

Theorem 7.3 (Security of LWE)**.** *Suppose there exists a polynomial time algorithm W that without access to the secret key, can correctly decode at least a $\frac{1}{2}+\frac{1}{poly(n)}$ fraction of messages. Then there exists a distinguisher Z that for a $\frac{1}{poly(n)}$-fraction of $\boldsymbol{s}\in\mathbb{Z}_p^n$, can distinguish between $\mathcal{A}_{\boldsymbol{s},\bar{\psi}_\alpha}$ and $\mathcal{U}$ with high probability.*

Proof. For a bit $M\in\{0,1\}$, and a public key $(\boldsymbol{a}_i,b_i)_{i\in[m]}$, let

$$\mathrm{Bob}((\boldsymbol{a}_i,b_i)_{i\in[m]},M)=\Big(\sum_{i\in S}\boldsymbol{a}_i,\sum_{i\in S}b_i+M\cdot\Big\lceil\frac{p}{2}\Big\rceil\Big)$$

denote the message that Bob sends. The algorithm W knows the public key $(\boldsymbol{a}_i,b_i)_{i\in[m]}$ and the message $(\boldsymbol{a},b)$ send by Bob and W either accepts (i.e. W believes 1 is the encrypted message) or it rejects (i.e. W believes 0 was the encrypted message). We abbreviate $\chi:=\bar{\psi}_\alpha$ and define the probabilities

$$\begin{aligned}
p_M(\boldsymbol{s}) &:= \Pr_{(a_i,\boldsymbol{b}_i)\sim\mathcal{A}_{\boldsymbol{s},\chi}}\big[W\text{ accepts }\big((\boldsymbol{a}_i,b_i)_i,\mathrm{Bob}((\boldsymbol{a}_i,b_i)_i,M)\big)\big]\quad\text{for } M\in\{0,1\} \\
p_{\mathcal{U}}(\boldsymbol{s}) &:= \Pr_{(a_i,\boldsymbol{b}_i)\sim\mathcal{U},(\boldsymbol{a},b)\sim\mathcal{U}}\big[W\text{ accepts }\big((\boldsymbol{a}_i,b_i)_i,(\boldsymbol{a},b)\big)\big]
\end{aligned}$$

By assumption, there has to be a set $Y \subseteq \mathbb{Z}_q^n$ of size $\frac{|Y|}{|\mathbb{Z}_p^n|} \geq \frac{1}{n^c}$ so that $|p_1(\boldsymbol{s}) - p_0(\boldsymbol{s})| \geq \frac{1}{n^c}$ for all $\boldsymbol{s} \in Y$. Now, consider a distribution $\mathcal{R} \in \{\mathcal{U}, \mathcal{A}_{\boldsymbol{s},\chi}\}$ where $\boldsymbol{s} \in Y$ — our goal will be to use the distinguisher W to tell which distribution $\mathcal{R}$ is. Consider the quantity

$$\rho := \max_{M\in\{0,1\}} \Big| \underbrace{\Pr_{(\boldsymbol{a}_i,b_i)\sim\mathcal{R}} \big[W \text{ accepts } \big((\boldsymbol{a}_i,b_i), \text{Bob}((\boldsymbol{a}_i,b_i),M)\big)\big]}_{(*)} - \underbrace{\Pr_{(\boldsymbol{a}_i,b_i)\sim\mathcal{R},(\boldsymbol{a},b)\sim\mathcal{U}} \big[W \text{ accepts } \big((\boldsymbol{a}_i,b_i),(\boldsymbol{a},b)\big)\big]}_{(**)} \Big|$$

By taking a large enough number of samples from $\mathcal{R}$ and using our distinguisher W, we can estimate ρ up to say $\frac{1}{16n^c}$ error. Now we distinguish two cases.

- *Case* $\mathcal{R} = \mathcal{U}$. By the left over hash lemma we know that the statistical distance of Bob's message to the uniform distribution is bounded by

$$\mathop{\mathbb{E}}_{(\boldsymbol{a}_i,b_i)\sim\mathcal{U}} \Big[\Delta\big(\text{Bob}((\boldsymbol{a}_i,b_i)_i,0),\mathcal{U}\big)\Big] \leq \sqrt{\frac{|\mathbb{Z}_p^n \times \mathbb{Z}_p|}{2^m}} = \sqrt{\frac{p^{n+1}}{2^m}} \leq 2^{-\Theta(n)}$$

  Hence the two probabilities $(*)$ and $(**)$ differ by at most twice that statistical distance and so $\rho \leq 2 \cdot 2^{-\Theta(n)}$.

- *Case* $\mathcal{R} = \mathcal{A}_{\boldsymbol{s},\chi}$. Then

$$\rho = \max_{M\in\{0,1\}} \big|p_M(\boldsymbol{s}) - (**)\big| \overset{\text{triangle ineq}}{\geq} \frac{1}{2}|p_1(\boldsymbol{s}) - p_0(\boldsymbol{s})| \geq \frac{1}{2n^c}$$

  since the probability in $(**)$ does not depend on M.

□

7.3 From an LWE Distinguisher for a fraction of keys to solving LWE for all keys

So far we have proven in Theorem 7.3 that if one could decrypt a $\frac{1}{2} + \frac{1}{\text{poly}(n)}$ fraction of messages from the LWE crypto system, then one could distinguish the distributions $\mathcal{U}$ and $\mathcal{A}_{\boldsymbol{s},\bar{\psi}_\alpha}$ — but possibly only for a $\frac{1}{\text{poly}(n)}$ fraction of keys $\boldsymbol{s}$. This does not seem particularly useful. In this section we will show that via a sequence of two additional reductions, this translates into an algorithm that solves the following much more powerful problem on *all* inputs $\boldsymbol{s}$ (and again with the distribution $\chi := \bar{\psi}_\alpha$):

LWE$_{p,\chi}$
Input: Access to distribution $\mathcal{A}_{\boldsymbol{s},\chi}$ where $\chi : \{0,\ldots,p-1\} \to \mathbb{R}_{\geq 0}$, $\boldsymbol{s} \in \mathbb{Z}_p^n$
Output: Determine $\boldsymbol{s} \in \mathbb{Z}_p^n$

Intuitively, LWE$_{p,\chi}$ corresponds to the problem of solving systems of approximate equations modulo p that we mentioned earlier. The algorithms for LWEDISTINGUISHING$_{p,\chi}$ and LWE$_{p,\chi}$ will be randomized and so they will come with some success probability. We can define that success probability to be at least say $\frac{2}{3}$, but one can repeat the algorithms polynomially often and boost the success probability to any quantity of the form $1 - 2^{-\text{poly}(n)}$ if desired. So we do not need to explicitly mention the probability threshold.

Theorem 7.4 (Reduction LWEDISTINGUISHING to LWE)**.** *Let p be a prime with $p \leq poly(n)$ and let $\chi : \{0,\ldots,p-1\} \to \mathbb{R}_{\geq 0}$ be any distribution. Assume there is a randomized polynomial time algorithm that solves* LWEDISTINGUISHING$_{p,\chi}$ *for keys $Y \subseteq \mathbb{Z}_p^n$ with $|Y| \geq \frac{|\mathbb{Z}_p^n|}{poly(n)}$. Then there is a randomized polynomial time algorithm that solves* LWE$_{p,\chi}$ *for all $\boldsymbol{s} \in \mathbb{Z}_p^n$.*

Proof. We show the main claim divided into two reductions that we denote as Claim I and Claim II.

Claim I. *Poly-time algo for* LWEDISTINGUISHING$_{p,\chi}$ *for key $Y \subseteq \mathbb{Z}_p^n$ with $\frac{|Y|}{|\mathbb{Z}_p^n|} \geq \frac{1}{poly(n)} \Longrightarrow$ Poly-time algo for* LWEDISTINGUISHING$_{p,\chi}$ *for all $\boldsymbol{s} \in \mathbb{Z}_p^n$.*

Proof of Claim I. Let W be the distinguisher that works for keys from Y where we assume that $|Y| \geq n^{-c}|\mathbb{Z}_p^n|$ for some constant $c > 0$. Let $\mathcal{R} \in \{\mathcal{U}, \mathcal{A}_{\boldsymbol{s},\chi}\}$ be the given distribution. Our goal is to distinguish the cases $\mathcal{R} = \mathcal{U}$ from $\mathcal{R} = \mathcal{A}_{\boldsymbol{s},\chi}$ whether or not $\boldsymbol{s}$ is in Y. Consider the following procedure:

(1) Set $\rho_{\mathcal{U}} := \Pr_{(\boldsymbol{a},b)\sim\mathcal{U}}[W \text{ accepts } (\boldsymbol{a},b)]$

(2) Sample $T \subseteq \mathbb{Z}_p^n$ with $|T| = n^{1+c}$

(3) Set

$$\rho := \max_{\boldsymbol{t}\in T} \Big\{ \Big| \underbrace{\Pr\Big[W \text{ accepts distr. } (\boldsymbol{a}, b + \langle \boldsymbol{a}, \boldsymbol{t}\rangle \bmod p) \text{ where } (\boldsymbol{a},b) \sim \mathcal{R}\Big]}_{(*)} - p_{\mathcal{U}} \Big| \Big\}$$

Note that we can estimate the probabilities in (1) and (3) up to any polynomially small error by sampling repeatedly from $\mathcal{U}$ or $\mathcal{R}$ and feeding inputs to W. So, for the sake of simplicity, we work with the exact value of ρ but there will be enough slack in the argument to distinguish the cases also in the presence of sampling error.

Claim I.A. *If $\mathcal{R} = \mathcal{U}$, then $\rho = 0$.*
Proof of Claim I.A. In this case, for any fixed $\boldsymbol{t}$ and $(\boldsymbol{a}, b) \sim \mathcal{U}$, the distribution of $(\boldsymbol{a}, b + \langle \boldsymbol{a}, \boldsymbol{t}\rangle) \sim \mathcal{U}$ is still uniform. Then ρ is the maximum difference between identical probabilities and so $\rho = 0$. □
Claim II.B. If $\mathcal{R} = \mathcal{A}_{\boldsymbol{s},\chi}$ for some $\boldsymbol{s} \in \mathbb{Z}_p^n$, then with high probability $\rho \geq \frac{1}{3}$.
Proof of Claim II.B. With high probability over the choice of T, there is a $\boldsymbol{t} \in T$ so that $\boldsymbol{s}+\boldsymbol{t} \in Y$. We condition on this event and fix $\boldsymbol{t}$. For $(\boldsymbol{a}, b) = (\boldsymbol{a}, \langle \boldsymbol{a}, \boldsymbol{s}\rangle + e) \sim \mathcal{A}_{\boldsymbol{s},\chi}$ we know that $(\boldsymbol{a}, b + \langle \boldsymbol{a}, \boldsymbol{t}\rangle) = (\boldsymbol{a}, \langle \boldsymbol{a}, \boldsymbol{s}+\boldsymbol{t}\rangle + e) \sim \mathcal{A}_{\boldsymbol{s}+\boldsymbol{t},\chi}$. Then as $\boldsymbol{s}+\boldsymbol{t} \in Y$, the distinguisher will be successful in the sense that $(*) \geq 2/3$ while $p_{\mathcal{U}} \leq 1/3$ and so $\rho \geq \frac{1}{3}$. □
This concludes the first reduction.
Claim II. *Poly-time algo for* $\text{LWEDISTINGUISHING}_{p,\chi}$ *for all* $\boldsymbol{s} \in \mathbb{Z}_p^n \Longrightarrow$ *Poly-time algorithm for* $LWE_{p,\chi}$ *for all* $\boldsymbol{s} \in \mathbb{Z}_p^n$.
Proof of Claim II. Let W be the LWE distinguisher and suppose we have access to samples of distribution $\mathcal{A}_{\boldsymbol{s},\chi}$. The goal is determine $\boldsymbol{s}$. Fix an index $i \in [n]$ and for $q \in \{0, \ldots, p-1\}$ define the following distribution:

> *Distribution* $\mathcal{R}_{i,q}$. Return $(\boldsymbol{a} + \ell \cdot \boldsymbol{e}_i, b + \ell \cdot q)$ where $(\boldsymbol{a}, b) \sim \mathcal{A}_{\boldsymbol{s},\chi}$ and $\ell \sim \mathbb{Z}_p$ uniformly.

We claim that by feeding $\mathcal{R}_{i,q}$ into the distinguisher W we can determine whether $q = s_i$ — then repeating the procedure for all indices $i \in [n]$ and $q \in \mathbb{Z}_p$ will provide $\boldsymbol{s}$. We prove the following two subclaims.
Claim II.A. *If $q = s_i$ then $\mathcal{R}_{i,q} = \mathcal{A}_{\boldsymbol{s},\chi}$.*
Proof of Claim II.A. Write $(\boldsymbol{a}, b) = (\boldsymbol{a}, \langle \boldsymbol{a}, \boldsymbol{s}\rangle + e) \sim \mathcal{A}_{\boldsymbol{s},\chi}$. Let $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{Z}_p^n$ be the ith unit vector. Then $b + \ell \cdot q \equiv_p \langle \boldsymbol{a}, \boldsymbol{s}\rangle + \ell s_i \equiv_p \langle \boldsymbol{a} + \ell e_i, \boldsymbol{s}\rangle$ and so $\mathcal{R}_{i,q}$ returns $(\boldsymbol{a} + \ell \boldsymbol{e}_i, \langle \boldsymbol{a} + \ell \boldsymbol{e}_i, \boldsymbol{s}\rangle + e) \sim \mathcal{A}_{\boldsymbol{s},\chi}$.
Claim II.B. *If $q \neq s_i$ then $\mathcal{R}_{i,q} = \mathcal{U}$.* Again write $(\boldsymbol{a}, b) = (\boldsymbol{a}, \langle \boldsymbol{a}, \boldsymbol{s}\rangle + e) \sim \mathcal{A}_{\boldsymbol{s},e}$. Then

$$\begin{aligned}(\boldsymbol{a} + \ell \boldsymbol{e}_i, b + \ell q) &= (\boldsymbol{a} + \ell e_i, \langle \boldsymbol{a} + \ell \boldsymbol{e}_i, \boldsymbol{s}\rangle + \ell(q - s_i)) \\ &\sim (\boldsymbol{a}, \langle \boldsymbol{a}, \boldsymbol{s}\rangle + \ell \underbrace{(q - s_i)}_{\neq 0}) \sim \mathcal{U}\end{aligned}$$

where we write $\sim$ for "having the same distribution". Note that here we need that p is a prime so that $\ell(q - s_i) \sim \mathbb{Z}_p$ uniformly whenever $\ell \sim \mathbb{Z}_p$. □

Combining Theorem 7.3 and Theorem 7.4 we obtain the following:

Corollary 7.5. Suppose there exists a polynomial time algorithm W that without access to the secret key, can correctly decode at least a $\frac{1}{2} + \frac{1}{\text{poly}(n)}$ fraction of messages. Then for all $\boldsymbol{s}$, given samples from $A_{\boldsymbol{s},\bar{\psi}_\alpha}$ one can determine $\boldsymbol{s} \in \mathbb{Z}_p^n$ in polynomial time.

7.4 Discrete Gaussian Sampling and the Smoothing Parameter

In this section, we will build up some technical tools that will be useful for the reductions later, but are also of general interest.

7.4.1 Sampling from a wide enough discrete Gaussian

We begin by proving a simple lemma about the change of Gaussian density:

Lemma 7.6. *Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $s, t, \ell > 0$ with $\|\boldsymbol{x}\|_2 \leq t$ and $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \ell$. Then $\rho_s(\boldsymbol{y}) \geq (1 - \pi(\frac{2\ell t + \ell^2}{s^2})) \cdot \rho_s(\boldsymbol{x})$.*

Proof. After scaling we may assume that $s = 1$. Then

$$\begin{aligned} \frac{\rho_1(\boldsymbol{y})}{\rho_1(\boldsymbol{x})} &= \exp\left(-\pi \cdot (\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{x}\|_2^2)\right) \\ &\overset{\text{triangle ineq.}}{\geq} \exp\left(-\pi((\|\boldsymbol{x}\|_2 - \ell)^2 - \|\boldsymbol{x}\|_2^2)\right) = \exp\left(-\pi(-2\ell \underbrace{\|\boldsymbol{x}\|_2}_{\leq t} + \ell^2)\right) \end{aligned}$$

□

In general, generating samples from the discrete Gaussian $\mathcal{D}_r(\Lambda)$ is a hard problem — but it becomes tractable if the parameter r is very large compared to the length of the basis vectors.

Lemma 7.7. *There is a polynomial time algorithm that for a lattice $\Lambda \subseteq \mathbb{R}^n$ and any parameter $r > 2^{2n}\lambda_n(\Lambda)$ computes a sample from a distribution $\tilde{\mathcal{D}}$ with $\Delta(\tilde{\mathcal{D}}, \mathcal{D}_r(\Lambda)) \leq 2^{-\Omega(n)}$.*

Proof. After scaling we may assume that $r = 1$ and $\lambda_n(\Lambda) \leq 2^{-2n}$. We use the LLL algorithm to compute an LLL-reduced basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ of the lattice. By Lemma 1.23, the vectors satisfy $\|\boldsymbol{b}_i\|_2 \leq 2^n \cdot \lambda_n(\Lambda) \leq 2^{-n}$ for all $i = 1, \ldots, n$. Then we construct the following distribution:

> **Distribution $\tilde{\mathcal{D}}$.** Sample $\boldsymbol{y} \sim \nu_1$. Return $\sum_{i=1}^n \boldsymbol{b}_i \lfloor y_i \rfloor \in \Lambda$.

Note that ν_1 is our notation for the Gaussian distribution $N(\boldsymbol{0}, \frac{1}{\sqrt{2\pi}}\boldsymbol{I}_n)$. Clearly $\tilde{\mathcal{D}}$ returns only lattice vectors in Λ. Recall that by Lemma 4.24, one has $\Pr_{\boldsymbol{x} \sim \mathcal{D}_1(\Lambda)}[\|\boldsymbol{x}\|_2 >$

$\sqrt{n}] \leq 2 \cdot 2^{-n}$. Hence it suffices to fix a lattice point $\boldsymbol{x} \in \Lambda$ with $\|\boldsymbol{x}\|_2 \leq \sqrt{n}$ and show the one sided bound of $\frac{\Pr[\mathcal{D}_1(\Lambda) \text{ returns } \boldsymbol{x}]}{\Pr[\tilde{\mathcal{D}} \text{ returns } \boldsymbol{x}]} \leq 1 + 2^{-\Omega(n)}$. First we estimate

$$\Pr[\mathcal{D}_1(\Lambda) \text{ returns } \boldsymbol{x}] = \frac{\rho_1(\boldsymbol{x})}{\rho_1(\Lambda)} \overset{\text{Poisson summation formula,Cor 4.19}}{=} \frac{\rho_1(\boldsymbol{x})}{\det(\Lambda^*)\rho_1(\Lambda^*)} \overset{\substack{\rho_1(\Lambda^*)\geq 1,\\ \det(\Lambda^*)\det(\Lambda)=1}}{\leq} \rho_1(\boldsymbol{x}) \cdot \det(\Lambda)$$

Next, let us abbreviate $R := \operatorname{diam}(\mathcal{P}(\boldsymbol{B})) \leq \sum_{i=1}^{n} \|\boldsymbol{b}_i\|_2 \leq n 2^n \lambda_n(\Lambda) \leq n 2^{-n}$. Then using the estimate from Lemma 7.6 we obtain

$$\begin{aligned} \Pr[\tilde{\mathcal{D}} \text{ returns } \boldsymbol{x}] &= \int_{\boldsymbol{x}+\mathcal{P}(\boldsymbol{B})} \nu_1(\boldsymbol{y}) d\boldsymbol{y} \overset{\text{Lem 7.6}}{\geq} (1 - \underbrace{6(\sqrt{n}R + R^2)}_{\leq 2^{-\Omega(n)}}) \cdot \underbrace{\nu_1(\boldsymbol{x})}_{=\rho_1(\boldsymbol{x})} \cdot \underbrace{\operatorname{Vol}_n(\mathcal{P}(\boldsymbol{B}))}_{=\det(\Lambda)} \\ &\geq (1 - 2^{-\Omega(n)}) \cdot \rho_1(\boldsymbol{x}) \cdot \det(\Lambda) \end{aligned}$$

□

7.4.2 The Smoothing Parameter

We introduce the following definition:

Definition 7.8. Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $\varepsilon > 0$. The smallest $s > 0$ with $\rho_{1/s}(\Lambda^* \setminus \{\boldsymbol{0}\}) \leq \varepsilon$ is called the *smoothing parameter* $\eta_\varepsilon(\Lambda)$.

That means $\eta_\varepsilon(\Lambda)$ gives the threshold beyond which the dual lattice has almost all the Gaussian weight on the origin. But admittedly it takes a while to understand why this quantity is so important for handling Gaussians on lattices. For example, beyond the smoothing threshold, shifting a lattice does not affect the value $\rho_r(\Lambda + \boldsymbol{c})$ by much. Note that this is closely related to Lemma 4.24 where we show essentially the same claim for $r = 1$ under the assumption that $\lambda_1(\Lambda^*) > \sqrt{n}$.

Lemma 7.9. *For any lattice $\Lambda \subseteq \mathbb{R}^n$, $\boldsymbol{c} \in \mathbb{R}^n$, $\varepsilon > 0$ and $r \geq \eta_\varepsilon(\Lambda)$ one has*

$$\rho_r(\Lambda + \boldsymbol{c}) \in (1 \pm \varepsilon) \cdot r^n \det(\Lambda^*)$$

Proof. By assumption $\rho_{1/r}(\Lambda^* \setminus \{\boldsymbol{0}\}) \leq \varepsilon$. Then by Lemma 4.20 we have

$$\begin{aligned} \rho_r(\Lambda + \boldsymbol{c}) &\overset{\text{Lem 4.20}}{=} \det(\Lambda^*) \cdot r^n \sum_{\boldsymbol{y} \in \Lambda^*} \rho_{1/r}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{y}, \boldsymbol{c} \rangle} \\ &= \det(\Lambda^*) \cdot r^n \cdot \Big(e^0 \cdot \underbrace{\rho_{1/r}(\boldsymbol{0})}_{=1} + \underbrace{\sum_{\boldsymbol{y} \in \Lambda^* \setminus \{\boldsymbol{0}\}} \rho_{1/r}(\boldsymbol{y}) \cdot e^{2\pi i \langle \boldsymbol{y}, \boldsymbol{c} \rangle}}_{|\cdot| \leq \varepsilon} \Big) \\ &= \det(\Lambda^*) \cdot r^n \cdot (1 \pm \varepsilon) \end{aligned}$$

□

Next, we want to relate the value of $\eta_\varepsilon(\Lambda)$ to known lattice parameters. For example one can prove that for $\varepsilon = \frac{1}{\text{poly}(n)}$, one has $\Theta(\sqrt{\log(n)}/n) \cdot \lambda_n(\Lambda) \leq \eta_\varepsilon(\Lambda) \leq \Theta(\sqrt{\log(n)}) \cdot \lambda_n(\Lambda)$ meaning that up to a factor of n, the smoothing parameter is proportional to $\lambda_n(\Lambda)$.

Parts of the following argument are taken from Regev and Micciancio [MR07a].

Lemma 7.10. *For a lattice $\Lambda \subseteq \mathbb{R}^n$ and $0 < \varepsilon \leq \frac{1}{2}$ one has*

$$\frac{1}{n}\sqrt{\frac{1}{\pi}\ln\left(\frac{1}{\varepsilon}\right)} \cdot \lambda_n(\Lambda) \leq \eta_\varepsilon(\Lambda) \leq \sqrt{\frac{1}{\pi}\ln\left(\frac{4n}{\varepsilon}\right)} \cdot \lambda_n(\Lambda)$$

Moreover $\eta_\varepsilon(\Lambda) \geq \sqrt{\frac{1}{\pi}\ln\left(\frac{1}{\varepsilon}\right)} \cdot \frac{1}{\lambda_1(\Lambda^*)}$.

Proof. For the sake of a simpler exposition we will ignore the exact constants and rescale the lattice as convinient. We split the bounds into two main claims.

Claim I. *If $\eta_\varepsilon(\Lambda) = 1$, then $\lambda_1(\Lambda^*) \geq \Theta(\sqrt{\ln(\frac{1}{\varepsilon})})$ and $\lambda_n(\Lambda) \leq \Theta(n/\sqrt{\ln(\frac{1}{\varepsilon})})$.*

Proof of Claim I. Let $\boldsymbol{y} \in \Lambda^*$ be a vector with $\|\boldsymbol{y}\|_2 = \lambda_1(\Lambda^*)$. Then $\varepsilon = \rho_1(\Lambda^* \setminus \{\boldsymbol{0}\}) \geq \exp(-\pi\|\boldsymbol{y}\|_2^2)$ and so $\|\boldsymbol{y}\|_2 \geq \Theta(\sqrt{\ln(\frac{1}{\varepsilon})})$. The 2nd part of the claim follows from the fact that $\lambda_1(\Lambda^*) \cdot \lambda_n(\Lambda) \leq 2n$ by Banaszczyk's Theorem (Cor 4.2). □

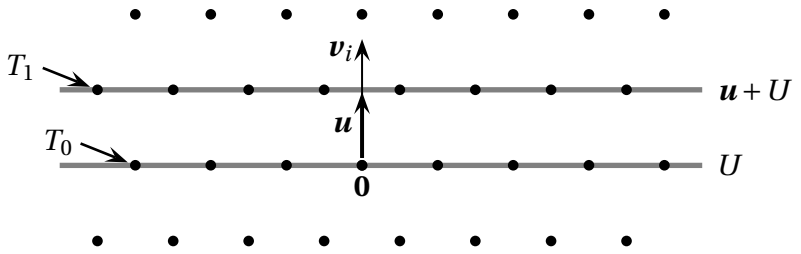
This gives the "moreover" part as well as the lower bound. It remains to prove the upper bound of $\eta_\varepsilon(\Lambda) \leq O(\sqrt{\log(\frac{4n}{\varepsilon})}) \cdot \lambda_n(\Lambda)$ which we do as follows:

Claim II. *If $\eta_\varepsilon(\Lambda) = 1$, then $\lambda_n(\Lambda) \geq \Theta(\frac{1}{\sqrt{\ln(\frac{4n}{\varepsilon})}})$.*

Proof of Claim II. By assumption $\rho_1(\Lambda^* \setminus \{\boldsymbol{0}\}) = \varepsilon$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \Lambda$ be linearly independent vectors of length $\|\boldsymbol{v}_i\|_2 \leq \lambda_n(\Lambda)$. Consider the sets of lattice points $S_i := \{\boldsymbol{y} \in \Lambda^* \mid \langle \boldsymbol{y}, \boldsymbol{v}_i \rangle \neq 0\}$. As $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent, we have $\Lambda^* \setminus \{\boldsymbol{0}\} = \bigcup_{i=1}^n S_i$. By the pigeonhole principle we can fix an index $i \in [n]$ so that $\rho_1(S_i) \geq \frac{\varepsilon}{n}$. Set $\boldsymbol{u} := \frac{\boldsymbol{v}_i}{\|\boldsymbol{v}_i\|_2^2}$ and note that $\|\boldsymbol{u}\|_2 = \frac{1}{\|\boldsymbol{v}_i\|_2} \geq \frac{1}{\lambda_n(\Lambda)}$; hence it suffices to prove that $\|\boldsymbol{u}\|_2 \leq \Theta(\sqrt{\ln(\frac{4n}{\varepsilon})})$. We consider the $(n-1)$-dimensional subspace $U := \{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{y} \perp \boldsymbol{u}\}$. Consider the "slices"

$$T_j := \left\{\boldsymbol{y} \in \Lambda^* \mid \langle \boldsymbol{y}, \boldsymbol{v}_i \rangle = j\right\} = \Lambda^* \cap (j\boldsymbol{u} + U)$$

for $j \in \mathbb{Z}$.



Then $\rho_1(T_j) = e^{-\pi\|j\boldsymbol{u}\|_2^2}\rho_1(T_j - j\boldsymbol{u}) \le e^{-\pi\|j\boldsymbol{u}\|_2^2}\rho_1(T_0)$ using first orthogonality and then the fact from Lemma 4.21 that shifts through the origin maximize the Gaussian weight. Hence

$$\frac{\varepsilon}{n} \le \sum_{j\neq 0}\rho_1(T_j) \le \underbrace{\rho_1(T_0)}_{\le 1+\varepsilon\le 2}\sum_{j\neq 0} e^{-\pi\|j\boldsymbol{u}\|_2^2} \le 2\sum_{j\ge 1}\left(e^{-\pi\|\boldsymbol{u}\|_2^2}\right)^j = \frac{2}{e^{\pi\|\boldsymbol{u}\|_2^2}-1}$$

Rearranging gives $\|\boldsymbol{u}\|_2 \le \sqrt{\frac{1}{\pi}\ln(1+\frac{2n}{\varepsilon})}$ which is of the desired form. □

A simple variant is the following:

Lemma 7.11. *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice. Then $\eta_{2^{-n}}(\Lambda) \le \frac{\sqrt{n}}{\lambda_1(\Lambda^*)}$.*

Proof. For convinience we prove an upper bound on $\eta_{2\cdot 2^{-n}}(\Lambda)$ instead. After scaling assume $\lambda_1(\Lambda^*) = \sqrt{n}$. Then by Lemma 4.26 we have $\rho_1(\Lambda^*\setminus\{\mathbf{0}\}) \le 2\cdot 2^{-n}$ which gives the claim. □

7.4.3 Statistical distance of Gaussian to Discrete Gaussian

Another property of the smoothing parameter is that it gives us a threshold so that the sum of discrete Gaussian and continuous Gaussian is statistically close to a "wider" Gaussian.

Lemma 7.12. *For any lattice $\Lambda \subseteq \mathbb{R}^n$, $\boldsymbol{u} \in \mathbb{R}^n$, $0 < \varepsilon \le \frac{1}{2}$ and $r, s > 0$ with $\frac{rs}{\sqrt{r^2+s^2}} \ge \eta_\varepsilon(\Lambda)$ one has*

$$\Delta\big(\mathcal{D}_r(\boldsymbol{u}+\Lambda) + \nu_s, \nu_{\sqrt{r^2+s^2}}\big) \le 4\varepsilon$$

Proof. In order to keep the exposition simple, we only consider the case $r = s = 1$. We ensure the reader that conceptially nothing is being hidden — but the calculations simplify. Then the claim reads that for a lattice with $\eta_\varepsilon(\Lambda) \le \frac{1}{\sqrt{2}}$ one has $\Delta(\mathcal{D}_1(\boldsymbol{u}+\Lambda)+\nu_1, \nu_{\sqrt{2}}) \le 4\varepsilon$ for any $\boldsymbol{u} \in \mathbb{R}^n$. Let us consider a random variable $\boldsymbol{Y} \sim \big(\mathcal{D}_1(\boldsymbol{u}+\Lambda)+\nu_1\big)$. Note that due to the added Gaussian, $\boldsymbol{Y}$ is a continuous

random variable. For convinience we denote the density at a point $\boldsymbol{x} \in \mathbb{R}^n$ as $\boldsymbol{Y}(\boldsymbol{x})$. First we determine an explicit expression for that density; note that we hope that this density is close to $\nu_{\sqrt{2}}(\boldsymbol{x}) = 2^{-n/2} e^{-\frac{\pi}{2}\|\boldsymbol{x}\|_2^2}$. We have

$$\begin{aligned}
\boldsymbol{Y}(\boldsymbol{x}) &= \sum_{\boldsymbol{y}\in\boldsymbol{u}+\Lambda} \frac{\rho_1(\boldsymbol{y})}{\rho_1(\boldsymbol{u}+\Lambda)} \cdot \frac{\rho_1(\boldsymbol{x}-\boldsymbol{y})}{\rho_1(\mathbb{R}^n)} \\
&\overset{\rho_1(\mathbb{R}^n)=1}{=} \frac{1}{\rho_1(\boldsymbol{u}+\Lambda)} \sum_{\boldsymbol{y}\in\boldsymbol{u}+\Lambda} \exp\left(-\pi\left(\|\boldsymbol{y}\|_2^2 + \|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right)\right) \\
&= \frac{\exp(-\frac{\pi}{2}\|\boldsymbol{x}\|_2^2)}{\rho_1(\boldsymbol{u}+\Lambda)} \sum_{\boldsymbol{y}\in\boldsymbol{u}+\Lambda} \exp\left(-2\pi\left\|\boldsymbol{y}-\frac{1}{2}\boldsymbol{x}\right\|_2^2\right) \\
&= \rho_{\sqrt{2}}(\boldsymbol{x}) \cdot \frac{\rho_{1/\sqrt{2}}(\Lambda+\boldsymbol{u}-\frac{\boldsymbol{x}}{2})}{\rho_1(\boldsymbol{u}+\Lambda)} \\
&\overset{\text{Lem 7.9 twice}}{=} \rho_{\sqrt{2}}(\boldsymbol{x}) \cdot \frac{(1\pm\varepsilon)\cdot(1/\sqrt{2})^n \det(\Lambda^*)}{(1\pm\varepsilon)\cdot\det(\Lambda^*)} = (1\pm 4\varepsilon)\cdot 2^{-n/2}\rho_{\sqrt{2}}(\boldsymbol{x})
\end{aligned}$$

The claim follows as the densities are within a $1\pm 4\varepsilon$ factor. □

7.5 Overview over reduction

So the remaining question for this chapter is what would be the implication of a polynomial time algorithm for $\text{LWE}_{p,\bar{\psi}_\alpha}$. Recall that $\mathcal{D}_r(\Lambda)$ is the distribution that samples each lattice point $\boldsymbol{x} \in \Lambda$ with probability $\frac{\rho_r(\boldsymbol{x})}{\rho_r(\Lambda)}$. We can formulate this sampling as a computational problem called *Discrete Gaussian Sampler (DGS)*:

DGS_f
Input: Lattice $\Lambda \subseteq \mathbb{R}^n$ and parameter r with $r > f(\Lambda)$.
Output: Compute a sample $\boldsymbol{x} \sim \mathcal{D}_r(\Lambda)$

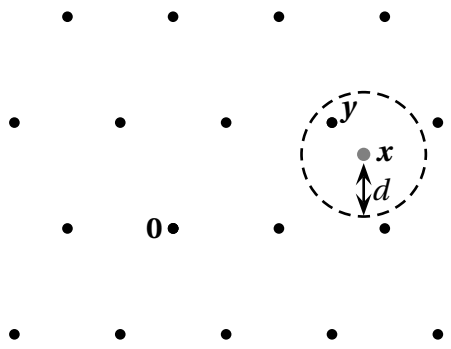
Here we think of f as a threshold that gives a lower bound on the admissible values of r. We allow that f is a function depending on the lattice, for example f could be in terms of the smoothing radius $\eta_\varepsilon(\Lambda)$ or in terms of the shortest vector in the dual lattice $\lambda_1(\Lambda^*)$. Also note that the larger the threshold f is, the easier is the problem of sampling from $\mathcal{D}_r(\Lambda)$. For example we have already proven that the problem $\text{DGS}_{2^{2n}\lambda_n(\Lambda)}$ is solvable in polynomial time[2].

The 2nd problem that we use in the reductions is as follows:

[2]Technically speaking there is an exponentially small statistical error — we will ignore that issue for the sake of a cleaner representation.

CVP$_{\Lambda,d}$
Input: Point $\boldsymbol{x} \in \mathbb{R}^n$.
Output: Compute $\boldsymbol{y} \in \Lambda$ with $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq d$ (if there is any)

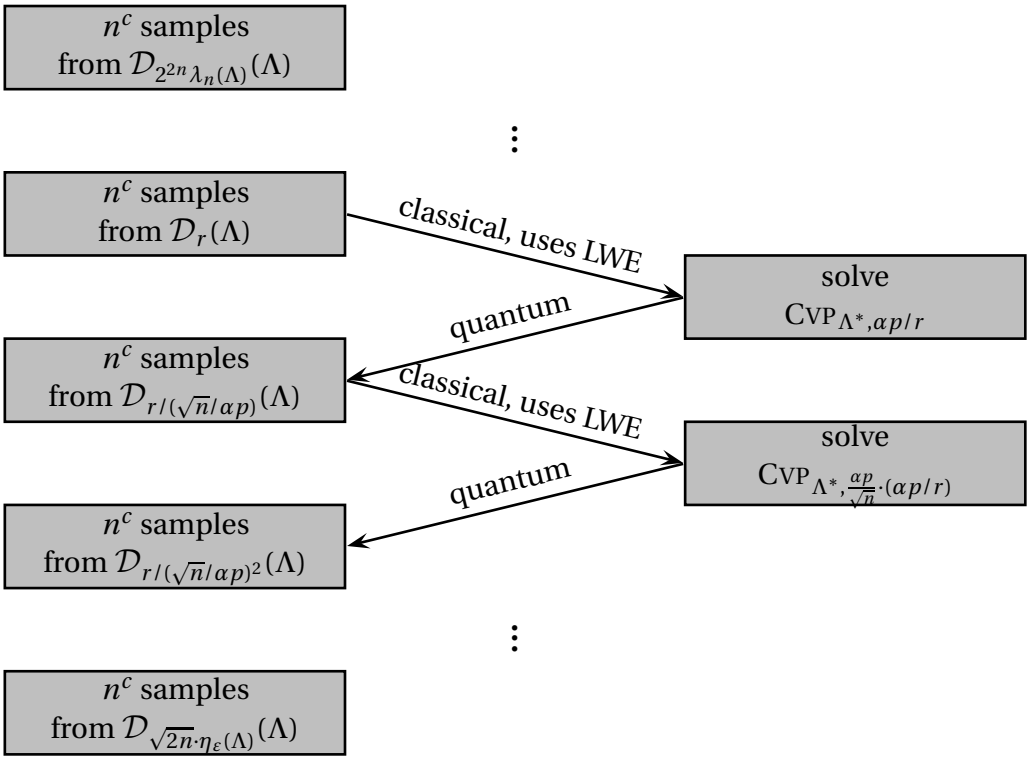
In our application we will always choose the distance as $d < \frac{1}{2}\lambda_1(\Lambda)$ so that the choice of $\boldsymbol{y}$ will be unique (if there is any). We also denote $\kappa_\Lambda(\boldsymbol{x}) := \operatorname{argmin}\{\|\boldsymbol{x} - \boldsymbol{y}\|_2 : \boldsymbol{y} \in \Lambda\}$ as the closest lattice point.



We assume that we can solve $\text{LWE}_{p,\bar{\psi}_\alpha}$ in polynomial time and the goal is to construct a DGS sampling algorithm with a parameter r that is so tight that it also implies an algorithm for GAPSVP with only a polynomial gap.

Roughly speaking, we will start with a large parameter $R := 2^{2n}\lambda_n(\Lambda)$ and generate n^c many samples from $\mathcal{D}_R(\Lambda)$. Then we can use the LWE oracle to solve the Closest vector problem in the dual lattice (to be precise we solve $\text{CVP}_{\Lambda^*,\alpha p/R}$). Then assuming to be able to solve $\text{CVP}_{\Lambda^*,\alpha p/R}$ we can design a *quantum algorithm* that generates n^c samples from $\mathcal{D}_{R\cdot\frac{\sqrt{n}}{\alpha p}}(\Lambda)$. So if $\alpha p < \sqrt{n}$, then this means we are able to sample from a narrower discrete Gaussian. We can iterate this argument until it breaks down when the parameter reaches $\sqrt{2n}\cdot\eta_\varepsilon(\Lambda)/\alpha$. In other words, the argument stops working when we are just polynomially above the smoothing threshold.

We reproduce a helpful figure from Regev depicting the sequence of reductions.



Each iterative step can be broken down into two disjoint parts:

- **Part I:** Use samples from $\mathcal{D}_r(\Lambda)$ and the LWE oracle to solve $\text{CVP}_{\Lambda^*\alpha p/(\sqrt{2}r)}$
- **Part II:** Use $\text{CVP}_{\Lambda^*,\alpha p/(\sqrt{2}r)}$ to sample from $\mathcal{D}_{r\cdot\sqrt{n}/(\alpha p)}(\Lambda)$

We state both parts formally here and will defer their proof to a later section.

Lemma 7.13 (Part I of iterative step)**.** *Let $\varepsilon \leq n^{-\omega(1)}$ and $p \in \mathbb{Z}_{\geq 2}$, $0 < \alpha < 1$ and assume that there is a polynomial time algorithm for $\text{LWE}_{p,\bar{\psi}_\alpha}$. Then there is a constant $c > 0$ so that there is an efficient algorithm that has the following behavior:*

- *Input: $\Lambda \subseteq \mathbb{R}^n$, $r > \sqrt{2}p \cdot \eta_\varepsilon(\Lambda)$ and n^c samples from $\mathcal{D}_r(\Lambda)$*
- *Output: Solves the problem $\text{CVP}_{\Lambda^*,\alpha p/(\sqrt{2}r)}$.*

Lemma 7.14 (Part II of iterative step)**.** *There is an efficient quantum algorithm that has the following behaviour:*

- *Input: Lattice $\Lambda \subseteq \mathbb{R}^n$, parameter $d < \frac{1}{2}\lambda_1(\Lambda^*)$, oracle to $\text{CVP}_{\Lambda^*,d}$*
- *Output: A sample to $\mathcal{D}_{\sqrt{n}/(2d)}(\Lambda)$.*

Combining both parts gives us a full iterative step:

Lemma 7.15 (Iterative Step)**.** *Let $\varepsilon \leq n^{-\omega(1)}$, $p \in \mathbb{Z}_{\geq 1}$, $0 < \alpha < 1$ and assume that we can solve $\mathrm{LWE}_{p,\bar{\psi}_\alpha}$ in polynomial time. Then for some constant $c > 0$, there is an efficient quantum algorithm that has the following behavior:*

- Input: $\Lambda \in \mathbb{R}^n$, $r > \sqrt{2}p \cdot \eta_\varepsilon(\Lambda)$, n^c *samples from* $\mathcal{D}_r(\Lambda)$
- Output: *A sample from* $\mathcal{D}_{r\sqrt{n}/(\alpha p)}(\Lambda)$

Proof. Combine Lemma 7.13 and Lemma 7.13. Also use $\frac{\alpha p}{\sqrt{2}r} \leq \frac{1}{\eta_\varepsilon(\Lambda)} \leq \frac{1}{2}\lambda_1(\Lambda^*)$. □

We can now conclude the main theorem of Regev, proving that if we could solve $\mathrm{LWE}_{p,\bar{\psi}_\alpha}$ in polynomial time, then there would be a polynomial time quantum algorithm that could sample from the discrete Gaussian just polynomially above the smoothing threshold:

Theorem 7.16 (Main Theorem I)**.** *Let $\varepsilon \leq n^{-\omega(1)}$, $p \in \mathbb{Z}_{\geq 2}$, $0 < \alpha < 1$ so that $\alpha p > 2\sqrt{n}$ and assume $\mathrm{LWE}_{p,\bar{\psi}_\alpha}$ can be solved in polynomial time. Then there is an efficient quantum algorithm for $\mathrm{DGS}_{\sqrt{2n}\cdot\eta_\varepsilon(\Lambda)/\alpha}$.*

Proof. Fix a lattice $\Lambda \subseteq \mathbb{R}^n$ and parameter $r > \sqrt{2n} \cdot \eta_\varepsilon(\Lambda)$. The goal is to generate a sample from $\mathcal{D}_r(\Lambda)$. We set $r_i := r \cdot (\alpha p/\sqrt{n})^i$. Note that $\frac{\alpha p}{\sqrt{n}} \geq 2$ and so $r_i \geq 2^i r$. Then $r_{3n} \geq 2^{3n} r \geq 2^{2n}\lambda_n(\Lambda)$. In other words, r_{3n} is so large that we can efficiently generate samples from $\mathcal{D}_{r_{3n}}(\Lambda)$. Then for $i = 3n, 3n-1, \ldots, 0$ we may assume that we are able to generate n^c samples from $\mathcal{D}_{r_i}(\Lambda)$. Then feeding the same n^c samples into Lemma 7.15 we can generate n^c samples of r_{i-1}. The samples from $\mathcal{D}_{r_0}(\Lambda)$ satisfy the claim. □

We can also relate the hardness of LWE to the more standard problem of Shortest Vector. We state the main conclusion and defer the details to Section 7.8.

Theorem 7.17 (Main Theorem II)**.** *Assume that for some $0 < \alpha < 1$ and $p \in \mathbb{N}$ with $\alpha p \geq 2\sqrt{n}$ we can solve $\mathrm{LWE}_{p,\bar{\psi}_\alpha}$ in polynomial time. Then there is an efficient quantum algorithm for $\mathrm{GAPSVP}_{\Theta(n^{3/2}\log(n)/\alpha)}$.*

Proof. Set $\varepsilon := n^{-\log(n)}$. Under the assumption of the claim we can conclude from Theorem 7.16 that there is an efficient quantum algorithm for $\mathrm{DGS}_{\sqrt{2n}\cdot\eta_\varepsilon(\Lambda)/\alpha}$. Then by Lemma 7.10 and Cor 4.2 we have $\eta_\varepsilon(\Lambda) \leq \Theta(\sqrt{\ln(\frac{4n}{\varepsilon})})\lambda_n(\Lambda) \leq \Theta(n\ln(n))/\lambda_1(\Lambda^*)$. Hence we also have an efficient quantum algorithm for $\mathrm{DGS}_{\Theta(n^{3/2}\log(n))/\alpha}$. Then

applying Theorem 7.29, this implies an efficient quantum algorithm also for $\text{GAPSVP}_{\Theta(n^{3/2}\log(n))/\alpha}$. □

7.6 Using samples and LWE to solve CVP

In this section we will show the part of the reduction in which we use samples for $\mathcal{D}_r(\Lambda)$ and an oracle for $\text{LWE}_{p,\psi_\alpha}$ to solve the CVP problem. We break this part of the reduction into two steps; in the first step we show how to solve a variant of CVP where instead of the closest lattice vector $\kappa_\Lambda(\boldsymbol{x}) \in \Lambda$ we only need to compute a vector in $\kappa_\Lambda(\boldsymbol{x}) + p\Lambda$. The reason is that an $\text{LWE}_{p,\psi_\alpha}$ oracle naturally only finds answers in $\mathbb{Z}_p^n$. Formally, we will first solve the following problem:

$\text{CVP}^{(p)}_{\Lambda,d}$
Input: Point $\boldsymbol{x} \in \mathbb{R}^n$.
Output: Return any vector in $\kappa_\Lambda(\boldsymbol{x}) + p\Lambda$.

Equivalently one could also fix a basis $\boldsymbol{B}$ of the lattice Λ and then ask for the coefficient vector $\boldsymbol{B}^{-1}\kappa_\Lambda(\boldsymbol{x}) \mod p$. Since this part of the reduction is the key ingredient to the hardness of LWE, we want to give an informal overview first. So, suppose we can generate samples from $\mathcal{D}_r(\Lambda)$ and we can solve $\text{LWE}_{p,\psi_\alpha}$. Now we are given a target vector $\boldsymbol{x}$ as input for $\text{CVP}^{(p)}_{\Lambda^*,\alpha p/(\sqrt{2}r)}$, meaning that we are supposed to find a dual lattice vector close to $\boldsymbol{x}$ (modulo p). Consider the following distribution

> Sample $\boldsymbol{v} \sim \mathcal{D}_r(\Lambda)$ and return $(\boldsymbol{a}, b) := (\boldsymbol{B}^{-1}\boldsymbol{v} \bmod p, \langle \boldsymbol{x}, \boldsymbol{v}\rangle \bmod p) \in \mathbb{Z}_p^n \times \mathbb{Z}_p$.

Since r is larger than the smoothing radius of $p\Lambda$ we will be able to argue that $\boldsymbol{a}$ is approximately uniform from $\mathbb{Z}_p^n$. Next, we fix $\boldsymbol{a}$ and consider the conditional distribution of $b \equiv_p \langle \boldsymbol{x}, \boldsymbol{v}\rangle = \langle \boldsymbol{x} - \kappa_{\Lambda^*}(\boldsymbol{x}), \boldsymbol{v}\rangle + \langle \kappa_{\Lambda^*}(\boldsymbol{x}), \boldsymbol{v}\rangle$. By assumption, the vector $\boldsymbol{x} - \kappa_{\Lambda^*}(\boldsymbol{x})$ is short and even fixing $\boldsymbol{a} = \boldsymbol{B}^{-1}\boldsymbol{v} \bmod p$, the vector $\boldsymbol{v}$ still behaves enough like a Gaussian so that $\langle \boldsymbol{x} - \kappa_{\Lambda^*}(\boldsymbol{x}), \boldsymbol{v}\rangle$ is approximately Gaussian with small standard deviation. For the 2nd term we write $\langle \kappa_{\Lambda^*}(\boldsymbol{x}), \boldsymbol{v}\rangle = \langle (\boldsymbol{B}^*)^{-1}\kappa_{\Lambda^*}(\boldsymbol{v}), \boldsymbol{B}^{-1}\boldsymbol{v}\rangle \equiv_p \langle \boldsymbol{s}, \boldsymbol{a}\rangle$ where we set $\boldsymbol{s} := (\boldsymbol{B}^*)^{-1}\kappa_{\Lambda^*}(\boldsymbol{x}) \bmod p \in \mathbb{Z}_p^n$. We realize that this vector $\boldsymbol{s}$ is exactly the target coefficient vector that we'd like to find as $\boldsymbol{B}^*\boldsymbol{s} \in \kappa_{\Lambda^*}(\boldsymbol{x}) + p\Lambda$. And we have concluded that the distribution of $(\boldsymbol{a}, b)$ is indeed approximately of the form $(\boldsymbol{a}, \text{Gaussian} + \langle \boldsymbol{s}, \boldsymbol{a}\rangle)$ with $\boldsymbol{a} \sim \mathbb{Z}_p^n$ approximately uniform. Then we can use an LWE oracle to extract $\boldsymbol{s}$. We will now work to make this formal; for example we will need to add some extra Gaussian noise to b.

Lemma 7.18. *Let $\varepsilon \le n^{-\omega(1)}$, $p \in \mathbb{Z}_{\ge 2}$ and $0 < \alpha \le \frac{1}{2}$. Suppose we have a polynomial time algorithm for $\mathrm{LWE}_{p,\psi_\alpha}$ and access to samples from $\mathcal{D}_r(\Lambda)$. Then there is a polynomial time algorithm for $\mathrm{CVP}^{(p)}_{\Lambda^*,\alpha p/(\sqrt{2}r)}$ for any $r > \sqrt{2}p \cdot \eta_\varepsilon(\Lambda)$.*

Proof. Let $\boldsymbol{x} \in \mathbb{R}^n$ be the input to $\mathrm{CVP}^{(p)}_{\Lambda^*,\alpha p/(\sqrt{2}r)}$. Let $\boldsymbol{B}^* := (\boldsymbol{B}^{-1})^T$ be the basis of Λ^*. We define a vector $\boldsymbol{s} \in \mathbb{Z}_p^n$ by $\boldsymbol{s} := (\boldsymbol{B}^*)^{-1}\kappa_{\Lambda^*}(\boldsymbol{x}) \mod p$. Note that $\boldsymbol{s}$ is exactly the vector that we want to find (since $\boldsymbol{B}^*\boldsymbol{s} \in \kappa_{\Lambda^*}(\boldsymbol{x}) + p\Lambda^*$). Consider the following probability distribution (that we can efficiently sample from by assumption):

Distribution $\mathcal{D}^*$. Sample $\boldsymbol{v} \sim \mathcal{D}_r(\Lambda)$ and $e \sim \frac{\alpha}{\sqrt{2}}N(0,\frac{1}{2\pi})$. Return

$$(\boldsymbol{a}, b) := \left(\boldsymbol{B}^{-1}\boldsymbol{v} \bmod p, \frac{\langle \boldsymbol{x}, \boldsymbol{v}\rangle}{p} + e \bmod 1\right) \in \mathbb{Z}_p^n \times [0,1).$$

Intuitively, $\boldsymbol{a}$ is the coefficient vector modulo p of a random lattice point and b is a sum of the correlation of that random lattice vector plus a random Gaussian noise modulo 1. We claim that the pair $(\boldsymbol{a}, b)$ is indistinguishably close to the LWE distribution for the target vector $\boldsymbol{s}$:

Claim I. *For some $0 < \beta \le \alpha$ one has $\Delta((\boldsymbol{a}, b), \mathcal{A}_{\boldsymbol{s},\psi_\beta}) \le n^{-\omega(1)}$ where $(\boldsymbol{a}, b) \sim \mathcal{D}^*$.*

Proof of Claim I. First we discuss the distribution of $\boldsymbol{a}$ alone when $(\boldsymbol{a}, \beta) \sim \mathcal{D}^*$. Indeed, note that the probability for an individual outcome of $\boldsymbol{a}$ is

$$\Pr[\boldsymbol{a}] = \frac{\rho_r(p\Lambda + \boldsymbol{B}\boldsymbol{a})}{\rho_r(\Lambda)} \overset{\text{Lem 7.9}}{=} \frac{(1\pm\varepsilon)r^n \det((p\Lambda)^*)}{r^n \det(\Lambda^*)} = (1\pm\varepsilon)\cdot p^{-n}$$

Here we use Lemma 7.9 and the fact that $\eta_\varepsilon(p\Lambda) = p \cdot \eta_\varepsilon(\Lambda) < r$ by assumption. Next, fix an outcome of $\boldsymbol{a} \in \mathbb{Z}_p^n$ and consider the conditional distribution of b. The goal will be to show that this conditional distribution of b is (approximately) of the form $\langle \boldsymbol{s}, \boldsymbol{a}\rangle + \beta \cdot N(0,\frac{1}{2\pi}) \bmod 1$. We abbreviating $\boldsymbol{x}' := \frac{1}{p}(\boldsymbol{x} - \kappa_{\Lambda^*}(\boldsymbol{x}))$ and note that by assumption $\boldsymbol{x}'$ is short with length $\|\boldsymbol{x}'\|_2 \le \frac{\alpha p/(\sqrt{2}r)}{p} = \frac{\alpha}{\sqrt{2}r}$. Then we can write

$$\begin{aligned}
b \equiv_1 \frac{\langle \boldsymbol{x}, \boldsymbol{v}\rangle}{p} + e \quad &\equiv_1 \quad \langle \frac{\boldsymbol{x} - \kappa_{\Lambda^*}(\boldsymbol{x})}{p}, \boldsymbol{v}\rangle + e + \frac{1}{p}\langle \kappa_{\Lambda^*}(\boldsymbol{x}), \boldsymbol{v}\rangle \\
&\overset{(\boldsymbol{B}^*)^T\boldsymbol{B}=\boldsymbol{I}_n}{\equiv_1} \quad \langle \boldsymbol{x}', \boldsymbol{v}\rangle + e + \frac{1}{p}\langle \underbrace{(\boldsymbol{B}^*)^{-1}\kappa_{\Lambda^*}(\boldsymbol{x})}_{\equiv_p \boldsymbol{s}}, \underbrace{\boldsymbol{B}^{-1}\boldsymbol{v}}_{\equiv_p \boldsymbol{a}}\rangle \\
&\equiv_1 \quad \langle \boldsymbol{x}', \boldsymbol{v}\rangle + e + \frac{1}{p}\langle \boldsymbol{s}, \boldsymbol{a}\rangle
\end{aligned}$$

Note that as we fixed $\boldsymbol{a}$, the distribution of $\boldsymbol{v}$ is $\mathcal{D}_r(p\Lambda + \boldsymbol{B}\boldsymbol{a})$. Then by Lemma 7.12 we know that $\langle \boldsymbol{x}', \mathcal{D}_r(p\Lambda + \boldsymbol{B}\boldsymbol{a})\rangle + \frac{\alpha}{\sqrt{2}}N(0,\frac{1}{2\pi}) \bmod 1$ is (up to a statistical error of

$n^{-\omega(1)}$) distributed like ψ_β where $\beta := \sqrt{(r\|\boldsymbol{x}'\|_2)^2 + (\frac{\alpha}{\sqrt{2}})^2} \leq \alpha$. Here we have to verify that indeed[3] $1/\sqrt{\frac{1}{r^2} + (\frac{\sqrt{2}\|\boldsymbol{x}'\|_2}{\alpha})^2} \geq \frac{r}{\sqrt{2}} > \eta_\varepsilon(\Lambda)$. The claim follows. □

Next, we would like to feed samples from $(\boldsymbol{a}, b)$ into our oracle for $\text{LWE}_{p,\psi_\alpha}$ and extract $\boldsymbol{s}$. There is a small issue since $(\boldsymbol{a}, b)$ is approximately distributed as $\mathcal{A}_{\boldsymbol{s},\psi_\beta}$ for some (unknown) parameter $0 < \beta \leq \alpha$ and our oracle comes with parameter α and not β. But there is a simple solution. We guess the value of β to enough accuracy and add some Gaussian noise so that we indeed have samples from $\mathcal{A}_{\boldsymbol{s},\psi_\alpha}$. Then our oracle can find $\boldsymbol{s}$. We leave it as an exercise we can verify which of the computed candidates for $\boldsymbol{s}$ is the correct one (which is necessary as we had to try out all guesses for β). □

Next, we show that an efficient algorithm for $\text{CVP}^{(p)}_{\Lambda,d}$ indeed implies an efficient algorithm for $\text{CVP}_{\Lambda,d}$.

Lemma 7.19. *Let $p \in \mathbb{Z}_{\geq 2}$ and let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $d < \frac{\lambda_1(\Lambda)}{2}$. If there is an efficient algorithm for $\text{CVP}^{(p)}_{\Lambda,d}$ then there is an efficient algorithm for $\text{CVP}_{\Lambda,d}$.*

Proof. We say that Λ' is a *t-scaled affine sublattice* of Λ, if $\Lambda' = \boldsymbol{u} + t\Lambda$ where $\boldsymbol{u} \in \Lambda$ and $t \in \mathbb{Z}_{\geq 1}$. Note that in particular $\Lambda' \subseteq \Lambda$. We assume access to an oracle for the problem $\text{CVP}^{(p)}_{\Lambda,d}$ where Λ, d and p are fixed s.t. $d < \frac{\lambda_1(\Lambda)}{2}$. We observe that by shifting and scaling the input to the oracle, the same oracle (with the same lattice, d and p) can be used to solve the following problem:

> $\widetilde{\text{CVP}}^{(p)}_{\Lambda',td}$: Given a t-scaled affine sublattice Λ' of Λ where $t \in \mathbb{Z}_{\geq 1}$. Then for any input $\boldsymbol{x}$ with $\text{dist}(\boldsymbol{x}, \Lambda') \leq td$, return a tp-scaled affine sublattice Λ'' of Λ so that $\Lambda'' \subseteq \Lambda'$ and $d(\boldsymbol{x}, \Lambda'') \leq td$.

Now, let $\boldsymbol{x} \in \mathbb{R}^n$ be our target vector with $\text{dist}(\boldsymbol{x}, \Lambda) \leq d$. We iteratively compute a sequence $\Lambda_0, \Lambda_1, \ldots, \Lambda_N$ where $\Lambda_0 := \Lambda$ and $\Lambda_{i+1} := \widetilde{\text{CVP}}^{(p)}_{\Lambda_i, p^i d}(\boldsymbol{x})$ for $i = 0, \ldots, N-1$. By construction, each Λ_i is a p^i-scaled affine sublattice of Λ. We want to prove via induction that $\kappa_\Lambda(\boldsymbol{x}) \in \Lambda_i$ for all i. If this holds true for an index $i \geq 0$, then $\text{dist}(\boldsymbol{x}, \Lambda_i) \leq d \leq p^i d$. Hence the next oracle call will return a sparser affine sublattice Λ_{i+1} with $\text{dist}(\boldsymbol{x}, \Lambda_{i+1}) \leq p^i d$. By assumption $d < \frac{\lambda_1(\Lambda)}{2}$ and hence the minimum distance between points of Λ_i is strictly bigger than $2p^i d$. Hence there is a *unique* point $\boldsymbol{y}_i \in \Lambda_i$ with $\|\boldsymbol{x} - \boldsymbol{y}_i\|_2 \leq p^i d$ and that point must be $\boldsymbol{y}_i = \kappa_\Lambda(\boldsymbol{x})$. Hence also $\kappa_\Lambda(\boldsymbol{x}) \in \Lambda_{i+1}$, concluding the inductive argument.

[3] To be more precise, we can use Lemma 7.12 to first bound the statistical distance of the n-dimensional distribution $\mathcal{D}_r(p\Lambda + \boldsymbol{B}\boldsymbol{a}) + \frac{\alpha}{\sqrt{2}} \frac{1}{\|\boldsymbol{x}'\|_2} N(0, \frac{1}{2\pi}\boldsymbol{I}_n)$ to an n-dimensional Gaussian. Then applying the map $\langle \boldsymbol{x}', \cdot \rangle \bmod 1$ to both sides cannot increase the statistical distance.

After polynomially many iterations N, the lattice Λ_N will be sparse enough so that Lemma 2.8 can solve the problem $\text{CVP}(\Lambda_N, \boldsymbol{x})$ exactly. □

7.7 A quantum algorithm to generate samples with a CVP oracle

In this section, we will sketch the proof of Lemma 7.14 and explain why with an oracle for $\text{CVP}_{\Lambda^*,d}$ one can construct an efficient quantum algorithm that generates a sample from the discrete Gaussian $\mathcal{D}_{\sqrt{n}/(\sqrt{2}d)}(\Lambda)$ (assuming that $d < \frac{\lambda_1(\Lambda^*)}{2}$). Note that this reduction does not even use an LWE oracle and we only describe it for the sake of completeness.

7.7.1 A brief intro to quantum computing

We will give a brief introduction to *quantum computing*. We refer to the popular textbook of Nielsen and Chuang [NC00] for details. For a *complex number* $z \in \mathbb{C}$ we can write $z = a + bi$ with $a, b \in \mathbb{R}$ where a is the *real part* and b is the *imaginary part*. Its *absolute value* is $|z| = \sqrt{a^2 + b^2}$ and the *complex conjugate* is $\bar{z} = a - bi$. A *(normalized) n-bit quantum state* is of the form

$$\sum_{\boldsymbol{x} \in \{0,1\}^n} \alpha_{\boldsymbol{x}} \cdot |\boldsymbol{x}\rangle$$

where $|\boldsymbol{x}\rangle$ denotes a vector of Euclidean length 1 in the 2^n-dimensional vector space $\mathbb{C}^{2^n}$, vectors $|\boldsymbol{x}\rangle$ and $|\boldsymbol{y}\rangle$ with $\boldsymbol{x} \neq \boldsymbol{y}$ are orthogonal, one has $\alpha_{\boldsymbol{x}} \in \mathbb{C}$ for all $\boldsymbol{x} \in \{0,1\}^n$ and $\|\boldsymbol{\alpha}\|_2^2 = \sum_{\boldsymbol{x} \in \{0,1\}^n} |\alpha_{\boldsymbol{x}}|^2 = 1$. Moreover, $|\boldsymbol{x}\rangle$ is equal to the *tensor product* $|x_1\rangle \otimes \ldots \otimes |x_n\rangle$ where each $|x_i\rangle$ is a 2-dimensional vector. The vectors $|\boldsymbol{x}\rangle$ (with $\boldsymbol{x} \in \{0,1\}^n$) are also called *computational basis states*, the complex number $\alpha_{\boldsymbol{x}}$ is called *amplitude* and the linear combination $\sum_{\boldsymbol{x} \in \{0,1\}^n} \alpha_{\boldsymbol{x}} \cdot |\boldsymbol{x}\rangle$ of those states is called a *superposition*.

Recall that for a matrix $\boldsymbol{U} \in \mathbb{C}^{2^n \times 2^n}$, the *conjugate transpose* is the matrix $\boldsymbol{U}^*$ with entries $(\boldsymbol{U}^*)_{ij} = \bar{U}_{ji}$. In the physics literature one may also find the notation $\boldsymbol{U}^\dagger$ instead of $\boldsymbol{U}^*$. A matrix $\boldsymbol{U} \in \mathbb{C}^{2^n \times 2^n}$ is *unitary* if $\boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I}$ where $\boldsymbol{I}$ is the identity in $\mathbb{C}^{2^n}$. A *quantum algorithm* is a sequence of unitary matrices that are iteratively applied to the current quantum state. More precisely for some initial n-bit quantum state $\psi = \sum_{\boldsymbol{x} \in \{0,1\}^n} \alpha_{\boldsymbol{x}} \cdot |\boldsymbol{x}\rangle$ and unitary matrices $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T$ the quantum algorithm computes the state

$$\boldsymbol{U}_T \cdots \boldsymbol{U}_2 \boldsymbol{U}_1 \, \psi$$

Each such unitary matrix $\boldsymbol{U}_t$ is also called a *quantum gate*. In principle, there is no restriction to which unitary matrices may be used as quantum gates. But an arbitrary unitary matrix could have description length of the order 2^n, so for efficiency reasons we need to ask that quantum gates only operate on few coordinates and act as the identity elsewhere.

For example, the *NOT gate for coordinate 1* is the unique unitary matrix $\boldsymbol{X}$ so that $\boldsymbol{X}|x_1, x_2, \ldots, x_n\rangle = |(1-x_1), x_2, \ldots, x_n\rangle$ for all $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$. We will say that a quantum algorithm is *polynomial time* if T is bounded by a polynomial in the input length and each intermediate quantum gate has bounded description length (such as the NOT gate). Note that a unitary matrix $\boldsymbol{U}$ corresponds to an *invertible linear map* that is *length preserving* and in particular a normalized quantum state is always mapped to another normalized quantum state.

At the end of its computation, a quantum algorithm can perform a *measurement*[4]. If the algorithm is in the quantum state $\sum_{\boldsymbol{x}\in\{0,1\}^n} \alpha_{\boldsymbol{x}} \cdot |\boldsymbol{x}\rangle$ then the measurement will produce a random variable that returns $\boldsymbol{x} \in \{0,1\}^n$ with probability $|\alpha_{\boldsymbol{x}}|^2$.

Note that any quantum computer could in principle be implemented on a classical computer. But the intermediate quantum states would require description length 2^n. In contrast a quantum computer[5] would only require n quantum bits and could perform the task efficiently. Not surprisingly, a quantum computer can also perform computations of a "classical" Turing machine, though stating that precisely takes some care. By replacing AND, OR, NOT gates with corresponding matrix operations (that can be made unitary by adding extra bits) one can obtain:

Theorem 7.20 (Quantum Simulation of Classical Circuit)**.** *Let $f : \{0,1\}^n \to \{0,1\}^m$ be a function that can be computed with a classical circuit of size t. Then there is a quantum algorithm with poly(n,m,t) iterations that for all $\boldsymbol{x} \in \{0,1\}^n$ maps the state $|\boldsymbol{x}, \boldsymbol{0}^{m+s}\rangle$ to $|\boldsymbol{x}, f(\boldsymbol{x}), \boldsymbol{0}^s\rangle$ (here s is also bounded by a polynomial in n,m,t).*

We also need an operation that has the somewhat mystical name of "uncomputing". Behind it is the simple observation that Theorem 7.20 produces a unitary matrix $\boldsymbol{U}$ and its inverse $\boldsymbol{U}^{-1}$ is again unitary and corresponds again to a quantum algorithm (with the same number of quantum gates, just in reverse order).

Theorem 7.21 (Uncomputing)**.** *Let $f : \{0,1\}^n \to \{0,1\}^m$ be a function that can be*

[4] Actually the quantum computing model allows to sample some quantum bits in the middle of its run but this simplest of measurements will suffice for us.

[5] If it were to would exist..

computed with a classical circuit of size t. Then there is a quantum algorithm with poly(n,m,t) iterations that for all $\boldsymbol{x} \in \{0,1\}^n$ maps the state $|\boldsymbol{x}, f(\boldsymbol{x}), \mathbf{0}^s\rangle$ to $|\boldsymbol{x}, \mathbf{0}^{m+s}\rangle$ (here s is also bounded by a polynomial in n,m,t).

7.7.2 Lattices and Quantum Computing

The quantum states that we will construct for our lattice problem will be in the form

$$\psi = \sum_{\boldsymbol{x}\in\Lambda} \alpha_{\boldsymbol{x}} \cdot |\boldsymbol{x}\rangle \tag{7.1}$$

with $\alpha_{\boldsymbol{x}} \in \mathbb{C}$. Note that in principle one could replace each number $x_i \in \mathbb{Q}$ by polynomially many bits to go back to the bit model. But this would be a notational nightmare and we will rather use the "lattice quantum state" notation. The $\|\cdot\|_2$-length of such a quantum state is $\|\psi\|_2 = \sqrt{\sum_{\boldsymbol{x}\in\Lambda} |\alpha_{\boldsymbol{x}}|^2}$. Also it will be notationally convinient to not require the states to be normalized. We will rather keep in mind that any state of the form (7.1) corresponds to the normalized quantum state

$$\frac{\psi}{\|\psi\|_2} = \frac{1}{\sqrt{\sum_{\boldsymbol{x}\in\{0,1\}^n} |\alpha_{\boldsymbol{x}}|^2}} \sum_{\boldsymbol{x}\in\Lambda} \alpha_{\boldsymbol{x}} \cdot |\boldsymbol{x}\rangle$$

We will use the following well known algorithm:

Theorem 7.22 (Quantum Fourier Transform)**.** *Let $R \in \mathbb{N}$ and let $\Lambda \subseteq \mathbb{R}^n$ be a lattice. Then there is a polynomial time quantum algorithm that maps a quantum state*

$$\psi = \sum_{\boldsymbol{s}\in\mathbb{Z}_R^n} \alpha_{\boldsymbol{s}} \cdot |\boldsymbol{s}\rangle$$

to

$$\mathrm{QFT}(\psi) = \frac{1}{(\sqrt{R})^n} \sum_{\boldsymbol{t}\in\mathbb{Z}_R^n} \beta_{\boldsymbol{t}} \cdot |\boldsymbol{t}\rangle \quad \textit{where} \quad \beta_{\boldsymbol{t}} = \sum_{\boldsymbol{s}\in\mathbb{Z}_R^n} \alpha_{\boldsymbol{s}} \cdot \exp\left(2\pi i \frac{\langle \boldsymbol{s}, \boldsymbol{t}\rangle}{R}\right)$$

7.7.3 From CVP to sampling from the discrete Gaussian

We will need the following statement which says that we can compute a quantum state that is close to the continuous Gaussian.

Lemma 7.23. *There is a polynomial time quantum algorithm that takes as input a lattice $\Lambda := \Lambda(\boldsymbol{B})$ and $r > 2^{2n} \cdot \lambda_n(\Lambda)$ and generates a quantum state ψ that is within $\|\cdot\|_2$ distance at most $2^{-\Omega(n)}$ to the state*

$$\sum_{\boldsymbol{x}\in\Lambda} \rho_{\sqrt{2}r}(\boldsymbol{x}) \cdot |\boldsymbol{x}\rangle$$

We will skip the proof as it is very similar to Lemma 7.7 and refer to [Reg09b] for details. As a sanity check, suppose we would perform a quantum measurement on the state $\sum_{\boldsymbol{x}\in\Lambda}\rho_{\sqrt{2}r}(\boldsymbol{x})\cdot|\boldsymbol{x}\rangle$ produced by Lemma 7.23. This would produce a lattice vector $\boldsymbol{x}\in\Lambda$ with probability proportional to

$$\rho_{\sqrt{2}r}(\boldsymbol{x})^2 = \exp(-\pi\|\boldsymbol{x}/(\sqrt{2}r)\|_2^2)^2 = \exp(-\pi\|\boldsymbol{x}/r\|_2^2) = \rho_r(\boldsymbol{x}). \tag{7.2}$$

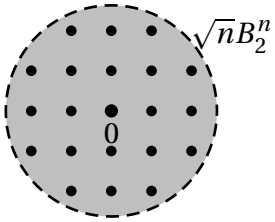
Hence we recover the statement from Lemma 7.7.

Recall that for a matrix $\boldsymbol{B}\in\mathbb{R}^{n\times n}$, $\mathcal{P}(\boldsymbol{B}) = \{\boldsymbol{B}\boldsymbol{y} \mid \boldsymbol{y}\in[0,1)^n\}$ denotes the *fundamental parallelepiped* of the lattice $\Lambda(\boldsymbol{B})$. Also recall that for any vector $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}$, we write $\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B}) = \sum_{i=1}^n \boldsymbol{b}_i(y_i - \lfloor y_i\rfloor)\in\mathcal{P}(\boldsymbol{B})$ as the translate by a lattice vector that lies in the parallelepiped $\mathcal{P}(\boldsymbol{B})$.

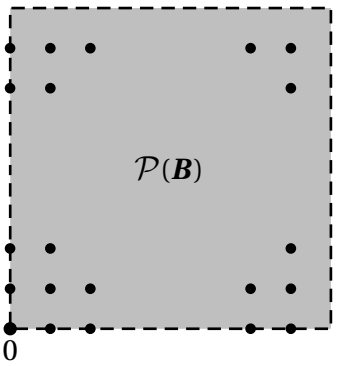
Lemma 7.24. *Let $\Lambda\subseteq\mathbb{R}^n$ be a full rank lattice with $\lambda_1(\Lambda) > 2\sqrt{n}$ where $\Lambda := \Lambda(\boldsymbol{B})$. Consider the quantum states*

$$\psi_1 = \sum_{\boldsymbol{x}\in\frac{\Lambda}{R}:\|\boldsymbol{x}\|_2<\sqrt{n}} \rho_1(\boldsymbol{x})\cdot|\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B})\rangle \quad \text{and} \quad \psi_2 = \sum_{\boldsymbol{x}\in\frac{\Lambda}{R}\cap\mathcal{P}(\boldsymbol{B})} \rho_1(\boldsymbol{x}-\Lambda)\cdot|\boldsymbol{x}\rangle$$

Then $\|\frac{\psi_1}{\|\psi_1\|_2} - \frac{\psi_2}{\|\psi_2\|_2}\|_2 \le 2^{-\Omega(n)}$.



points $\frac{\Lambda}{R}\cap\sqrt{n}B_2^n$



points $\left(\frac{\Lambda}{R}\cap\sqrt{n}B_2^n\right) \bmod \mathcal{P}(\boldsymbol{B})$

Proof. First it will be useful to note that indeed

$$\psi_2 = \sum_{\boldsymbol{x}\in\frac{\Lambda}{R}\cap\mathcal{P}(\boldsymbol{B})}\sum_{\boldsymbol{y}\in\Lambda}\rho_1(\boldsymbol{x}-\boldsymbol{y})\cdot|\boldsymbol{x}\rangle = \sum_{\boldsymbol{x}\in\frac{\Lambda}{R}}\rho_1(\boldsymbol{x})\cdot|\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B})\rangle \tag{7.3}$$

which uses that $R\in\mathbb{N}$.

Next, we argue that each ket $|\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B})\rangle$ appears only once in the definition of ψ_1.

Claim I. *For $\boldsymbol{x},\boldsymbol{y}\in\frac{\Lambda}{R}$ with $\|\boldsymbol{x}\|_2,\|\boldsymbol{y}\|_2 < \sqrt{n}$ one has $\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B}) \neq \boldsymbol{y} \bmod \mathcal{P}(\boldsymbol{B})$.*

Proof of Claim. Suppose for the sake of contradiction that $\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B}) = \boldsymbol{y} \bmod \mathcal{P}(\boldsymbol{B})$. Then $\boldsymbol{x}-\boldsymbol{y} \in \Lambda$ and so $\|\boldsymbol{x}-\boldsymbol{y}\|_2 \geq \lambda_1(\Lambda) > 2\sqrt{n}$ which is a contradiction. □

Note that $\rho_1(\boldsymbol{x})^2 = \rho_{1/\sqrt{2}}(\boldsymbol{x})$ (see Eq (7.2)) which we can use to estimate

$$\|\psi_1\|_2 = \sum_{\boldsymbol{x}\in\frac{\Lambda}{R}:\|\boldsymbol{x}\|_2<\sqrt{n}} \rho_1(\boldsymbol{x})^2 = \rho_{1/\sqrt{2}}\Big(\frac{\Lambda}{R} \cap \sqrt{n}\cdot \mathrm{int}(B_2^n)\Big) \overset{\text{Cor } 4.25}{\geq} (1-2^{-n})\cdot \rho_{1/\sqrt{2}}\Big(\frac{\Lambda}{R}\Big) \tag{7.4}$$

where we Cor 4.25 says that $\rho_r(\Lambda \setminus \sqrt{n}B_2^n) \leq 2^{-n}\rho_r(\Lambda)$ for any $r \leq 1$. Then the difference between the states is

$$\begin{aligned}
\|\psi_1-\psi_2\|_2 &\overset{}{\leq} \|\psi_1-\psi_2\|_1 \\
&\overset{\text{Claim I+(7.3)}}{=} \sum_{\boldsymbol{x}\in\frac{\Lambda}{R}:\|\boldsymbol{x}\|_2\geq\sqrt{n}} \rho_1(\boldsymbol{x}) \\
&\overset{\text{Cor } 4.25}{\leq} 2^{-n}\cdot\rho_1\Big(\frac{\Lambda}{R}\Big) \overset{\text{Cor } 4.23}{\leq} 2^{-n}\cdot 2^{n/2}\cdot\rho_{1/\sqrt{2}}\Big(\frac{\Lambda}{R}\Big) \overset{(7.4)}{\leq} 2^{-\Omega(n)}\cdot\|\psi_1\|_2
\end{aligned}$$

Recall that Cor 4.23 states that for $r \geq 1$ and any $s > 0$ one has $\rho_{rs}(\Lambda) \leq r^n \rho_s(\Lambda)$. □

Lemma 7.25. *For any lattice Λ and $R \in \mathbb{N}$, the quantum fourier transform maps the state*

$$\psi = \sum_{\boldsymbol{x}\in\frac{\Lambda^*}{R}\cap\mathcal{P}(\boldsymbol{B}^*)} \rho_1(\boldsymbol{x}-\Lambda^*)\cdot|\boldsymbol{x}\rangle$$

to

$$\mathrm{QFT}(\psi) = R^{n/2}\det(\Lambda) \sum_{\boldsymbol{x}\in\mathcal{P}(R\boldsymbol{B})\cap\Lambda} \rho_1(R\Lambda-\boldsymbol{x})\cdot|\boldsymbol{x}\rangle$$

where $\boldsymbol{B}$ is a basis for Λ and $\boldsymbol{B}^* = (\boldsymbol{B}^{-1})^T$ the corresponding basis for Λ^*.

We will apply Lemma 7.25 with $R \to \infty$, but as a sanity check one might want to verify that for $R = 1$ we obtain that $\psi = \rho_1(\Lambda^*)\cdot|\boldsymbol{0}\rangle$ is mapped to $\mathrm{QFT}(\psi) = \det(\Lambda)\cdot\rho_1(\Lambda)\cdot|\boldsymbol{0}\rangle$. The norm needs to be preserved, i.e. $\rho_1(\Lambda^*) = \det(\Lambda)\cdot\rho_1(\Lambda)$, but this is indeed true by Cor 4.19.

Proof of Lemma 7.25. We will apply the quantum fourier transform in the "coefficient space" rather than the "lattice point space" (i.e. instead of $\boldsymbol{x} = \boldsymbol{B}^*\boldsymbol{y}$ we work with $\boldsymbol{y}$) and so changing the basis we can rewrite ψ to

$$\psi' = \sum_{\boldsymbol{s}\in\mathbb{Z}_R^n} \underbrace{\sum_{\boldsymbol{r}\in\mathbb{Z}^n} \rho_1\Big(\frac{\boldsymbol{B}^*\boldsymbol{s}}{R} - \boldsymbol{B}^*\boldsymbol{r}\Big)}_{\alpha_{\boldsymbol{s}}}\cdot|\boldsymbol{s}\rangle$$

Then $\mathrm{QFT}(\psi') = R^{-n/2}\sum_{\boldsymbol{t}\in\mathbb{Z}_R^n}\beta_{\boldsymbol{t}}\cdot|\boldsymbol{t}\rangle$ where by Theorem 7.22 for $\boldsymbol{t}\in\mathbb{Z}_R^n$ one has

$$\begin{aligned}
\beta_{\boldsymbol{t}} &= \sum_{\boldsymbol{s}\in\mathbb{Z}_R^n}\overbrace{\sum_{\boldsymbol{r}\in\mathbb{Z}^n}\rho_1\Big(\frac{\boldsymbol{B}^*\boldsymbol{s}}{R}-\boldsymbol{B}^*\boldsymbol{r}\Big)}^{\alpha_{\boldsymbol{s}}}\cdot\exp\Big(2\pi i\frac{\langle\boldsymbol{s},\boldsymbol{t}\rangle}{R}\Big)\\
&\overset{\text{cancellation}}{=} \sum_{\boldsymbol{y}\in\mathbb{Z}^n}\rho_1\Big(\frac{\boldsymbol{B}^*\boldsymbol{y}}{R}\Big)\cdot\exp\Big(2\pi i\frac{\langle\boldsymbol{y},\boldsymbol{t}\rangle}{R}\Big)\\
&\overset{\boldsymbol{x}=\boldsymbol{B}^*\boldsymbol{y}/R}{=} \sum_{\boldsymbol{x}\in\Lambda^*/R}\rho_1(\boldsymbol{x})\cdot\exp\big(2\pi i\underbrace{\langle(\boldsymbol{B}^*)^{-1}\boldsymbol{x},\boldsymbol{t}\rangle}_{=\langle\boldsymbol{x},\boldsymbol{B}\boldsymbol{t}\rangle}\big)\\
&= \sum_{\boldsymbol{x}\in\Lambda^*/R}\rho_1(\boldsymbol{x})\cdot\exp(2\pi i\,\langle\boldsymbol{x},\boldsymbol{B}\boldsymbol{t}\rangle)\\
&\overset{\text{Cor 4.17},\hat{\rho}_1(\boldsymbol{x})=\rho_1(\boldsymbol{x})}{=} \det(R\Lambda)\cdot\rho_1(R\Lambda-\boldsymbol{B}\boldsymbol{t})
\end{aligned}$$

Going back the lattice space, $\mathrm{QFT}(\psi')$ corresponds to the state

$$R^{n/2}\det(\Lambda)\sum_{\boldsymbol{x}\in\mathcal{P}(R\boldsymbol{B})\cap\Lambda}\rho_1(R\Lambda-\boldsymbol{x})\cdot|\boldsymbol{x}\rangle$$

as claimed. □

Finally we proof Lemma 7.14 which describes part II of the iterative step. We may scale the lattice and the parameters so that $d=\sqrt{n}$. Then the claim simplifies to:

Lemma 7.26 (Part II of iterative step — restated)**.** *There is an efficient quantum algorithm that has the following behaviour:*

- *Input: Lattice $\Lambda\subseteq\mathbb{R}^n$ with $\lambda_1(\Lambda^*)>2\sqrt{n}$, oracle to $\mathrm{CVP}_{\Lambda^*,\sqrt{n}}$*
- *Output: A sample to $\mathcal{D}_{1/2}(\Lambda)$.*

Proof. Let $\boldsymbol{B}$ be a basis of Λ and let $\boldsymbol{B}^*$ be a basis for Λ^*. For two (not necessarily normalized) quantum states we write $\psi_1\approx\psi_2$ if the Euclidean distance of the corresponding normalized states is less than $2^{-\Omega(n)}$. Then the claim is satisfied by the following quantum algorithm:

(1) Choose an integer R with $R\geq 2^{3n}\lambda_n(\Lambda^*)$.

(2) Use Lemma 7.23 to compute the quantum state

$$\sum_{\boldsymbol{x}\in\frac{\Lambda^*}{R}}\rho_1(\boldsymbol{x})\cdot|\boldsymbol{x}\rangle\approx\sum_{\boldsymbol{x}\in\frac{\Lambda^*}{R}:\|\boldsymbol{x}\|_2<\sqrt{n}}\rho_1(\boldsymbol{x})\cdot|\boldsymbol{x}\rangle$$

(3) We use that the map $\boldsymbol{x} \mapsto \boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B}^*)$ is computable in polynomial time and so by Theorem 7.20 we can add a new register to (2) and obtain the state

$$\sum_{\boldsymbol{x} \in \frac{\Lambda^*}{R} : \|\boldsymbol{x}\|_2 < \sqrt{n}} \rho_1(\boldsymbol{x}) \cdot |\boldsymbol{x}, \boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B}^*)\rangle$$

(4) Since the $\text{CVP}_{\Lambda^*, \sqrt{n}}$ oracle gives a polynomial time map $(\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B}^*)) \to \boldsymbol{x}$, we can use Theorem 7.21 uncompute the first register in (3) and obtain the state

$$\sum_{\boldsymbol{x} \in \frac{\Lambda^*}{R} : \|\boldsymbol{x}\|_2 < \sqrt{n}} \rho_1(\boldsymbol{x}) \cdot |\boldsymbol{x} \bmod \mathcal{P}(\boldsymbol{B}^*)\rangle \overset{\text{Lem } 7.24}{\approx} \sum_{\boldsymbol{x} \in \frac{\Lambda^*}{R} \cap \mathcal{P}(\boldsymbol{B}^*)} \rho_1(\boldsymbol{x} - \Delta^*) \cdot |\boldsymbol{x}\rangle$$

(5) We apply the quantum fourier transform from Lemma 7.25 and obtain a state proportional to

$$\sum_{\boldsymbol{x} \in \mathcal{P}(R\boldsymbol{B}) \cap \Lambda} \rho_1(R\Delta - \boldsymbol{x}) \cdot |\boldsymbol{x}\rangle \approx \sum_{\boldsymbol{x} \in \Lambda : \|\boldsymbol{x}\|_2 < \sqrt{n}} \rho_1(\boldsymbol{x}) \cdot |\boldsymbol{x} \bmod \mathcal{P}(R\boldsymbol{B})\rangle$$

(6) Measure the state in (5) and with probability proportional to $\rho_1(\boldsymbol{x})^2 = \rho_{1/\sqrt{2}}(\boldsymbol{x})$ we draw the vector $\boldsymbol{x} - \boldsymbol{y} \in \mathcal{P}(R\boldsymbol{B})$ where $\boldsymbol{y} \in R\Lambda$. Then $d(\boldsymbol{x} - \boldsymbol{y}, R\Lambda) \leq \|(\boldsymbol{x} - \boldsymbol{y}) - (-\boldsymbol{y})\|_2 = \|\boldsymbol{x}\|_2 \leq \sqrt{n}$ and $\lambda_1(R\Lambda) \geq 2^{3n}$, hence we can use Babai's algorithm (see Lemma 2.8) to compute $\boldsymbol{x}$ in polynomial time.

□

7.8 Reduction from GAPSVP to DGS

So far we managed to reduce the hardness of LWE to assumed hardness of sampling from the discrete Gaussian polynomially above the smoothing threshold. But we can actually continue the reduction to the better known problem of GAPSVP that we have seen earlier in Chapter 6. For convinience we restate the problem (with an additional parameter d — but by a scaling argument this does not change the problem).

GAPSVP$_\gamma$
Input: Lattice $\Lambda \subseteq \mathbb{R}^n$ and a parameter $d > 0$.
Goal: Distinguish the cases
- **Yes.** One has $\lambda_1(\Lambda) \leq d$.
- **No.** One has $\lambda_1(\Lambda) > \gamma \cdot d$

Recall that $d(\boldsymbol{t}, \Lambda) := \min\{\|\boldsymbol{x} - \boldsymbol{t}\|_2 : \boldsymbol{x} \in \Lambda\}$ denotes the Euclidean distance of a point $\boldsymbol{t}$ to the lattice. Our reduction will go via an intermediate problem which is as follows

GAPCVP$'_\gamma$
Input: Lattice $\Lambda \subseteq \mathbb{R}^n$, a target $\boldsymbol{t} \in \mathbb{R}^n$ and a parameter $d > 0$.
Goal: Distinguish the cases
- **Yes.** One has $d(\boldsymbol{t}, \Lambda) \leq d$
- **No.** One has $d(\boldsymbol{t}, \Lambda) > \gamma \cdot d$ and $\lambda_1(\Lambda) > \gamma \cdot d$

Note that the NO case of GAPCVP$'_\gamma$ has an additional condition of $\lambda_1(\Lambda) > \gamma \cdot d$ which potentially might make it easier to distinguish the YES case and the NO case. Still an oracle for GAPCVP$'_\gamma$ can be used to solve GAPSVP$_\gamma$ as was shown by Micciancio and Regev [MR07b].

Lemma 7.27 ([MR07b])**.** *For any $\gamma \geq 1$, there is a polynomial time reduction from* GAPSVP$_\gamma$ *to* GAPCVP$'_\gamma$*.*

Proof. Let (Λ, d) be the input for GAPSVP$_\gamma$. Our reduction returns the following answer:

$$A := \bigvee_{i=1}^{n} \text{GAPCVP}'_\gamma\big((\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}, 2\boldsymbol{b}_i, \boldsymbol{b}_{i+1}, \ldots, \boldsymbol{b}_n), \boldsymbol{b}_i, d\big).$$

For convinience we abbreviate $\Lambda_i := \Lambda(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}, 2\boldsymbol{b}_i, \boldsymbol{b}_{i+1}, \ldots, \boldsymbol{b}_n)$. Note that Λ_i is a strict sublattice of Λ with exactly half the density. Also note that it may be the case that for some indices i, we call GAPCVP$'_\gamma(\Lambda_i, \boldsymbol{b}_i, d)$ while the instance $(\Lambda_i, \boldsymbol{b}_i, d)$ neither satisfies the YES condition nor the NO condition. In that case the oracle may return either answer. Our reduction will still work. We prove two directions.

Claim I. $\lambda_1(\Lambda) \leq d \implies$ *A equals YES.*

Proof of Claim I. Let $\boldsymbol{x} \in \Lambda \setminus \{\boldsymbol{0}\}$ be the shortest vector. Then one can write $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}$ with coefficient vector $\boldsymbol{y} \in \mathbb{Z}^n$. We can now fix a coordinate i so that y_i is odd — this coordinate has to exist since otherwise $\frac{\boldsymbol{x}}{2} \in \Lambda$ would have been a shorter vector. We claim that for this coordinate i, GAPCVP$'_\gamma(\Lambda_i, \boldsymbol{b}_i, d)$ is true. And in fact, we know that $\boldsymbol{x} + \boldsymbol{b}_i \in \Lambda_i$ and so $d(\boldsymbol{b}_i, \Lambda_i) \leq \|(\boldsymbol{x} + \boldsymbol{b}_i) - \boldsymbol{b}_i)\|_2 \leq d$. □

Claim II. $\lambda_1(\Lambda) > \gamma d \implies$ *A equals NO.*

Proof of Claim II. We need to verify that for every index $i \in [n]$, GAPCVP$'_\gamma(\Lambda_i, \boldsymbol{b}_i, d)$ is in the NO case. First, we have $\lambda_1(\Lambda_i) \geq \lambda_1(\Lambda) > \gamma d$. So it remains to verify that $d(\boldsymbol{b}_i, \Lambda_i) > \gamma d$. Suppose this is not true and $d(\boldsymbol{b}_i, \Lambda_i) \leq \gamma d$. Then $\|\boldsymbol{b}_i - \boldsymbol{x}\|_2 \leq \gamma d$ for some $\boldsymbol{x} \in \Lambda_i$. But $(\boldsymbol{b}_i - \boldsymbol{x}) \in \Lambda \setminus \{\boldsymbol{0}\}$ (as $\boldsymbol{b}_i \notin \Lambda_i$) which is a contradiction to $\lambda_1(\Lambda) > \gamma d$. □ □

Now we come to the reduction from DGS to GAPCVP′. The good news is that essentially we have already proven this reduction in Chapter 6. While at that time we focused on the **NP** ∩ **coNP** aspect, the proof really shows the following: if for a given lattice Λ we can sample from the dual lattice Λ^*, then we can approximate the function $F : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ defined as

$$F(\boldsymbol{x}) := \frac{\rho_1(\boldsymbol{x}+\Lambda)}{\rho_1(\Lambda)} = \mathop{\mathbb{E}}_{\boldsymbol{w}\sim\mathcal{D}_1(\Lambda^*)} \left[\cos(2\pi\langle \boldsymbol{w}, \boldsymbol{x}\rangle)\right]$$

up to a small error and then use this function to distinguish whether a given point $\boldsymbol{t}$ is very close to the lattice or rather far. It remains to verify that we can make the argument work with our specific parameters.

Lemma 7.28. *For any $\gamma \geq 1$, there is a polynomial time reduction from* $\text{GAPCVP}'_{100\sqrt{n}\gamma}$ *to* $\text{DGS}_{\sqrt{n}\gamma/\lambda_1(\Lambda^*)}$.

Proof. Let $(\Lambda, \boldsymbol{t}, d)$ be an input instance for $\text{GAPCVP}'_{100\sqrt{n}\gamma}$. We rescale the instance so that $d = \frac{1}{100}$. Then we have to distinguish the cases

- **YES.** One has $d(\boldsymbol{t}, \Lambda) \leq \frac{1}{100}$.
- **NO.** One has $d(\boldsymbol{t}, \Lambda) > \sqrt{n}\gamma$ and $\lambda_1(\Lambda) > \sqrt{n}\gamma$.

We will use the following decision algorithm for that distinction:

(1) Draw $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N \sim \mathcal{D}_1(\Lambda^*)$ using the DGS oracle where N is a large enough polynomial (see Lemma 6.7).

(2) IF $\sum_{i=1}^N \boldsymbol{w}_i \boldsymbol{w}_i^T \not\leq 3N \cdot \boldsymbol{I}_n$ THEN return YES

(3) Define the function

$$F_W(\boldsymbol{x}) := \frac{1}{N}\sum_{i=1}^N \cos\left(2\pi\langle \boldsymbol{x}, \boldsymbol{w}_i\rangle\right)$$

(4) IF $F_W(\boldsymbol{t}) \leq \frac{2}{n}$ THEN return NO ELSE return YES

We will prove that this algorithm correctly decides whether we are in the YES or NO case. First assume that we are in the NO case; we want to discuss what the algorithm returns. Note that the property $\lambda_1(\Lambda) > \sqrt{n}\gamma$ is equivalent to $1 > \sqrt{n}\gamma/\lambda_1(\Lambda)$. Hence parameter $r = 1$ that we use in (1) is above the guaranteed threshold where we know that our DGS oracle works (note that we apply DGS

to the dual lattice and use that $(\Lambda^*)^* = \Lambda$). Then the vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ are indeed independent samples from $\mathcal{D}_1(\Lambda^*)$. Then by Lemma 6.9 and Lemma 6.7 we know that for a suitable choice of N, with overwhelming probability the following events are both true:

(I) $\sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{w}_i^T \preceq 3N \cdot \boldsymbol{I}_n$.

(II) $|F_W(\boldsymbol{x}) - F(\boldsymbol{x})| \leq \frac{1}{n}$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Since we are in the NO case, we know that $d(\boldsymbol{t}, \Lambda) > \sqrt{n}\gamma$ and hence $F_W(\boldsymbol{t}) \leq \frac{2}{n}$ as we have shown in Claim II of the proof of Theorem 6.2. So the algorithm correctly returns NO.

Next, assume that the instance $(\Lambda, \boldsymbol{t}, d = \frac{1}{100})$ is in the YES case. Then we we have no control over the vectors sampled in (1). But if the algorithm terminates in (2) then then it correctly returns YES, so assume that the algorithm did not terminate then and instead $\sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{w}_i^T \preceq 3N \cdot \boldsymbol{I}_n$ holds (still this does not mean that (II) has to be satisfied). Then as $d(\boldsymbol{t}, \Lambda) \leq \frac{1}{100}$ we know from Claim I of the proof of Theorem 6.2 that $F_W(\boldsymbol{t}) > \frac{9}{10}$ and so the algorithm correctly returns YES in (4). □

Combining Lemma 7.27 and Lemma 7.28 we obtain the following:

Theorem 7.29. *For any $\beta \geq \sqrt{n}$, there is a polynomial time reduction from* $\text{GAPSVP}_{100\beta}$ *to* $\text{DGS}_{\beta/\lambda_1(\Lambda^*)}$.

7.9 Exercises

Exercise 7.1.
Prove (without using Lemma 7.10) that for all $\varepsilon \geq 2^{-n}$, one has $\eta_\varepsilon(\Lambda) \leq O(\frac{\sqrt{n}}{\lambda_1(\Lambda^*)})$. You may use the fact proven in Exercise 3.3 on page 69: For any lattice $\Lambda \subseteq \mathbb{R}^n$ and any $t > 0$ one has $|\Lambda \cap t\lambda_1(\Lambda) B_2^n| \leq (2t+1)^n$.

Exercise 7.2.
Let $r, t \geq 1$. Consider a lattice $\Lambda \subseteq \mathbb{R}^n$ with $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \leq t$ (we know that always $t \leq 2n$ but t can be a lot smaller). Prove that $|\Lambda \cap r\lambda_1(\Lambda) B_2^n| \leq 1 + n^{O(t^2 r^2)}$.
Hint. Use Lemma 7.10.
Comment. For a general lattice Λ one always has the bound $|\Lambda \cap O(1) \cdot \lambda_1(\Lambda) B_2^n| \leq 2^{O(n)}$. But for those lattices where the product $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*)$ is a lot smaller than the worst case, this bound improves. In particular if $\lambda_1(\Lambda) \cdot \lambda_n(\Lambda^*) \leq O(1)$, then $|\Lambda \cap O(1) \cdot \lambda_1(\Lambda) B_2^n| \leq n^{O(1)}$.

Chapter 8

The Reverse Minkowski Theorem and an Approximation to the Covering Radius

In this chapter, we will be discussing a recent result by Regev and Stephens-Davidowitz [RS17]. Suppose we have a lattice Λ that does not contain any dense sublattice, say $\det(\Lambda') \geq 1$ for all sublattices $\Lambda' \subseteq \Lambda$. Then Λ still may contain non-zero vectors of length $\Theta(1)$. For example this is the case for the standard lattice $\mathbb{Z}^n$. But remarkably, one can prove that any such lattice contains very few short vectors. Formally the statement is as follows:

Theorem 8.1 (Reverse Minkowski Theorem)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice that satisfies $\det(\Lambda') \geq 1$ for all sublattices $\Lambda' \subseteq \Lambda$. Then for $t = C \log n$ with $C > 0$ large enough one has $\rho_{1/t}(\Lambda) \leq \frac{3}{2}$.*

Note that the condition $\rho_{1/t}(\Lambda) \leq \frac{3}{2}$ implies that $|\Lambda \cap rB_2^n| \leq n^{O(\log n) \cdot r^2}$ for all $r \geq 1$.

A consequence of the Reverse Minkowski Theorem is a surprisingly tight characterization of the covering radius. First, note that for any full rank lattice $\Lambda \subseteq \mathbb{R}^n$ we have a simple determinant-based lower bound of

$$\mu(\Lambda) \geq \frac{\sqrt{n}}{6} \cdot \det(\Lambda)^{1/n},$$

which we have already proven in Lemma 1.43. For the proof we noted that a ball of radius $r := \frac{\sqrt{n}}{6} \cdot \det(\Lambda)^{1/n}$ has a volume of $\mathrm{Vol}_n(rB_2^n) < \det(\Lambda)$ and so a random translate of rB_2^n will in expectation intersect less than one lattice point. But this determinant lower bound can be arbitrarily loose — for example for the

2-dimensional lattice $\Lambda = \Lambda(\boldsymbol{B})$ with $\boldsymbol{B} = \begin{pmatrix} M & 0 \\ 0 & 1/M \end{pmatrix}$ where $\det(\Lambda) = 1$ and $\mu(\Lambda) \approx \frac{M}{2}$ for M large. On the other hand, one can take a subspace $W \subseteq \mathbb{R}^n$ and certainly has $\mu(\Lambda) \geq \mu(\Pi_W(\Lambda))$ since the projection of a covering is again a covering. Hence one can consider the best determinant lower bound that can be obtained for any projections:

$$\mu_{\det}(\Lambda) := \max_{W \subseteq \mathbb{R}^n \text{ subspace}} \sqrt{\dim(W)} \cdot \det(\Pi_W(\Lambda))^{1/\dim(W)}$$

Then one can prove that this approximates the covering radius astonishingly well:

Theorem 8.2 (Covering radius approximation)**.** *For any lattice $\Lambda \subseteq \mathbb{R}^n$ one has*

$$\Theta(1) \cdot \mu_{\det}(\Lambda) \leq \mu(\Lambda) \leq O(\log^{3/2} n) \cdot \mu_{\det}(\Lambda)$$

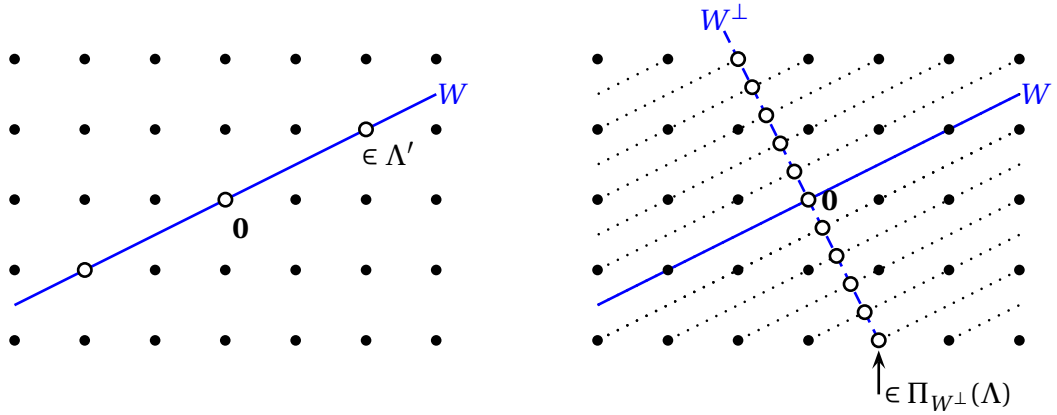
The result of Regev and Stephens-Davidowitz [RS17] is rather involved and requires a substantial amount of tools. We will take the liberty of skipping some proofs that take too much of a detour. We recommend Chapter 2 in the thesis of Stephens-Davidowitz [Ste17] for the omitted details.

8.1 Sublattices and quotient lattices

For most of these notes, we have only studied *full rank* lattices. That will be different for this chapter, where we will need to talk about *sublattices* that naturally do not have full rank. For a matrix $\boldsymbol{B} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_k) \in \mathbb{R}^{n \times k}$ with k linearly independent columns, we also write $\Lambda(\boldsymbol{B}) := \{\sum_{i=1}^k \lambda_i \boldsymbol{b}_i : \lambda_1, \dots, b_k \in \mathbb{Z}\}$. We define $\operatorname{rank}(\Lambda(\boldsymbol{B})) = \dim(\operatorname{span}(\Lambda(\boldsymbol{B}))) = k$. The definition of the determinant can be extended to $\det(\Lambda(\boldsymbol{B})) = \sqrt{\det_k(\boldsymbol{B}^T\boldsymbol{B})} = \operatorname{Vol}_k(\mathcal{P}(\boldsymbol{B}))$. A sublattice $\Lambda' \subseteq \Lambda$ is called a *primitive sublattice* of Λ if there is a subspace $W \subseteq \mathbb{R}^n$ so that $\Lambda' = \Lambda \cap W$. For example $\Lambda(\boldsymbol{e}_1, \dots, \boldsymbol{e}_k)$ is a primitive sublattice of $\mathbb{Z}^n$ while $2\mathbb{Z}^n$ is a non-primitive sublattice. A subspace $W \subseteq \mathbb{R}^n$ with $\operatorname{span}(\Lambda \cap W) = W$ is called a *lattice subspace (w.r.t. lattice Λ)*. So far we have only considered the *dual lattice* of a full rank lattice. For a (not necessarily full rank) lattice $\Lambda \subseteq \mathbb{R}^n$ we extend the definition to

$$\Lambda^* = \{\boldsymbol{x} \in \operatorname{span}(\Lambda) \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z} \ \ \forall \boldsymbol{y} \in \Lambda\}$$

Note that by construction $\operatorname{span}(\Lambda) = \operatorname{span}(\Lambda^*)$. For any primitive sublattice $\Lambda' \subseteq \Lambda$ with corresponding subspace $W := \operatorname{span}(\Lambda')$, we define the *quotient lattice* as $\Lambda/\Lambda' := \Pi_{W^\perp}(\Lambda)$.



We prove a few useful properties:

Lemma 8.3 (Properties of Lattice Subspaces)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $W \subseteq \mathbb{R}^n$ be a lattice subspace. Then*

(a) $(\Pi_W(\Lambda))^* = \Lambda^* \cap W$.

(b) $\det(\Lambda) = \det(\Lambda \cap W) \cdot \det(\Pi_{W^\perp}(\Lambda))$.

(c) $\det(\Lambda^* \cap W) = \det(\Lambda \cap W^\perp) \cdot \det(\Lambda^*)$.

Proof. (a) Clear because $\Lambda^* \cap W = \{\boldsymbol{x} \in W : \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z} \ \ \forall\, \boldsymbol{y} \in \Lambda\} = \{\boldsymbol{x} \in W \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z} \ \ \forall\, \boldsymbol{y} \in \Pi_W(\Lambda)\} = \Pi_W(\Lambda)^*$.

(b) Follows from the fact that shearing does not change the volume of the fundamental parallelepiped.

(c) We have

$$\det(\Lambda^* \cap W) \overset{(b)}{=} \frac{\det(\Lambda^*)}{\det(\Pi_{W^\perp}(\Lambda^*))} \overset{\det(\tilde{\Lambda}) \cdot \det(\tilde{\Lambda}^*) = 1}{=} \det(\Lambda^*) \cdot \det((\Pi_{W^\perp}(\Lambda^*))^*) \overset{(a)}{=} \det(\Lambda^*) \cdot \det(\Lambda \cap W^\perp)$$

□

So far we have only considered the *covering radius* for full rank lattices. Again, for an arbitrary lattice $\Lambda \subseteq \mathbb{R}^n$ we extend the definition to

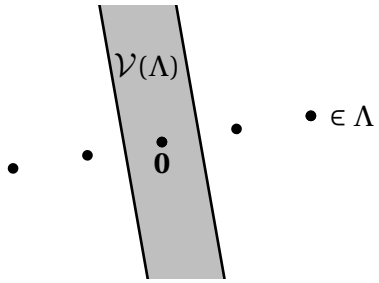
$$\mu(\Lambda) = \max_{\boldsymbol{x} \in \mathrm{span}(\Lambda)} \min_{\boldsymbol{y} \in \Lambda} \|\boldsymbol{x} - \boldsymbol{y}\|_2 = \min\left\{r \geq 0 \mid \Lambda + r B_2^n \supseteq \mathrm{span}(\Lambda)\right\}$$

Recall that the *standard Gaussian distribution* γ_n is the distribution with density function $\gamma_n(\boldsymbol{x}) = (2\pi)^{-n/2} \cdot e^{-\|\boldsymbol{x}\|_2^2/2}$ for all $\boldsymbol{x} \in \mathbb{R}^n$. We denote the *Voronoi cell* of a

lattice $\Lambda \subseteq \mathbb{R}^n$ as

$$\mathcal{V}(\Lambda) = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{y} - \boldsymbol{x}\|_2 \ \forall \boldsymbol{y} \in \Lambda\right\} = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{x}\rangle \leq \frac{\|\boldsymbol{y}\|_2^2}{2} \ \forall \boldsymbol{y} \in \Lambda\right\}$$

Note that now, Λ might not have full rank, in which case $\mathcal{V}(\Lambda)$ is unbounded in directions orthogonal to $\mathrm{span}(\Lambda)$.



For lattices Λ_1, Λ_2 that are orthogonal (i.e. $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$ for all $\boldsymbol{x} \in \Lambda_1, \boldsymbol{y} \in \Lambda_2$) we define the *direct sum* as $\Lambda_1 \oplus \Lambda_2 = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in \Lambda_1, \boldsymbol{y} \in \Lambda_2\}$. In fact, whenever we write $\Lambda_1 \oplus \Lambda_2$ we have the implicit assumption that Λ_1 and Λ_2 are orthogonal.

Lemma 8.4. *For lattices $\Lambda_1, \Lambda_2 \subseteq \mathbb{R}^n$ and $s > 0$ one has $\rho_s(\Lambda_1 \oplus \Lambda_2) = \rho_s(\Lambda_1) \cdot \rho_s(\Lambda_2)$.*

Proof. We have $\rho_s(\Lambda_1 \oplus \Lambda_2) = \sum_{\boldsymbol{x} \in \Lambda_1, \boldsymbol{y} \in \Lambda_2} e^{-\pi \|(\boldsymbol{x}+\boldsymbol{y})/s\|_2^2} = \sum_{\boldsymbol{x} \in \Lambda_1} \sum_{\boldsymbol{y} \in \Lambda_2} e^{-\pi \|\boldsymbol{x}/s\|_2^2} e^{-\pi \|\boldsymbol{y}/s\|_2^2} = \rho_s(\Lambda_1) \cdot \rho_s(\Lambda_2)$ using the orthogonality of Λ_1 and Λ_2. □

Lemma 8.5 (Properties of quotient lattices)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $\Lambda' \subseteq \Lambda$ be a primitive sublattice. Then*

(a) One has $\det(\Lambda) = \det(\Lambda/\Lambda') \cdot \det(\Lambda')$.

(b) One has $(\Lambda/\Lambda')^* = \Lambda^* \cap \mathrm{span}(\Lambda')^\perp$.

(c) For any $s > 0$, $\rho_s(\Lambda) \leq \rho_s(\Lambda') \cdot \rho_s(\Lambda/\Lambda')$.

(d) One has $\mu(\Lambda)^2 = \mu(\Lambda/\Lambda')^2 + \mu(\Lambda')^2$.

(e) One has $\gamma_n(\mathcal{V}(\Lambda)) \geq \gamma_n(\mathcal{V}(\Lambda/\Lambda')) \cdot \gamma_n(\mathcal{V}(\Lambda'))$.

Proof. We prove the items in order where we set $W := \mathrm{span}(\Lambda')$:

(a) This is equivalent to Lemma 8.3.(b).

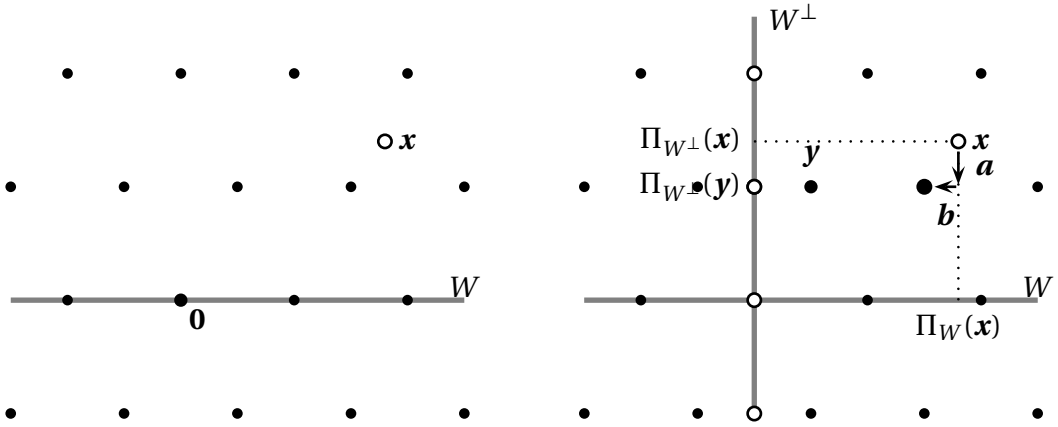
(b) We have $\Lambda^* \cap W^\perp = \Pi_{W^\perp}(\Lambda)^* = (\Lambda/\Lambda')^*$ using Lemma 8.3.(a) and the definition of quotient lattice.

(c) We claim that

$$\rho_s(\Lambda) \overset{(*)}{\leq} \rho_s(\Lambda' \oplus (\Lambda/\Lambda')) \overset{\text{Lem } 8.4}{=} \rho_s(\Lambda') \cdot \rho_s(\Lambda/\Lambda')$$

To see $(*)$, let $\boldsymbol{v} \in \Lambda$. Then the contributions of points in the the slice $\boldsymbol{v} + U$ to the LHS of $(*)$ is $\rho_s(\boldsymbol{v}+\Lambda') \leq \rho_s(\Pi_{W^\perp}(\boldsymbol{v})) \cdot \rho_s(\Lambda')$ which is the contribution to the RHS. Here we use that discrete Gaussian weight is maximized for the central slice, see Lemma 4.21.

(d) Consider any $\boldsymbol{x} \in \operatorname{span}(\Lambda)$. Let $\boldsymbol{y} \in \Lambda$ so that the difference vector $\boldsymbol{a} := \Pi_{W^\perp}(\boldsymbol{y}) - \Pi_{W^\perp}(\boldsymbol{x})$ has length $\|\boldsymbol{a}\|_2 \leq \mu(\Pi_{W^\perp}(\Lambda))$. Then there is a vector $\boldsymbol{b}$ so that $\boldsymbol{x} + \boldsymbol{a} + \boldsymbol{b} \in \boldsymbol{y} + \Lambda' \subseteq \Lambda$ where $\|(\boldsymbol{x} + \boldsymbol{a} + \boldsymbol{b}) - \boldsymbol{y}\|_2 \leq \mu(\Lambda')$ and $\boldsymbol{a} \perp \boldsymbol{b}$. Eventually $\|\boldsymbol{a} + \boldsymbol{b}\|_2^2 = \|\boldsymbol{a}\|_2^2 + \|\boldsymbol{b}\|_2^2 \leq \mu(\Lambda')^2 + \mu(\Lambda/\Lambda')^2$.



(e) We skip the proof here; see [Ste17] for details.

□

It is a useful fact that the determinant has a submodular behavior. For lattices this implies the following:

Lemma 8.6 (Submodularity of lattice determinants)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $\Lambda_1, \Lambda_2 \subseteq \Lambda$ be primitive sublattices. Then $\operatorname{rank}(\Lambda_1) + \operatorname{rank}(\Lambda_2) = \operatorname{rank}(\Lambda_1 \cap \Lambda_2) + \operatorname{rank}(\Lambda_1 + \Lambda_2)$ and moreover $\det(\Lambda_1 \cap \Lambda_2) \cdot \det(\Lambda_1 + \Lambda_2) \leq \det(\Lambda_1)\det(\Lambda_2)$.*

Proof. We only show the notationally easier case of $\Lambda_1 \cap \Lambda_2 = \{\mathbf{0}\}$ (see [Ste17] for the full argument). As both lattices are primitive, we also have $\operatorname{span}(\Lambda_1) \cap \operatorname{span}(\Lambda_2) = \{\mathbf{0}\}$. Then the claim on the rank is clear. For the moreover part we use that $\det(\Lambda_1 + \Lambda_2) = \det(\Lambda_1) \cdot \det(\Pi_{\operatorname{span}(\Lambda_1)^\perp}(\Lambda_2)) \leq \det(\Lambda_1)\det(\Lambda_2)$ since (dimension-preserving) projection can only decrease the determinant. □

8.2 Stable lattices

We come to a crucial definition:

Definition 8.7. A lattice $\Lambda \subseteq \mathbb{R}^n$ is *stable* if $\det(\Lambda) = 1$ and $\det(\Lambda') \geq 1$ for all sublattices $\Lambda' \subseteq \Lambda$.

Intuitively, a stable lattice is a lattice that does not contain a sublattice that is denser than Λ itself. We say that Λ is a *scaling of a stable lattice* if there is some $s > 0$ so that $s \cdot \Lambda$ is stable. It is not hard to see that one must have $s = \det(\Lambda)^{-1/\mathrm{rank}(\Lambda)}$.

Lemma 8.8. *If $\Lambda \subseteq \mathbb{R}^n$ is a stable lattice, then also Λ^* is stable.*

Proof. We have $\det(\Lambda^*) = \frac{1}{\det(\Lambda)} = 1$. Consider a lattice subspace $W \subseteq \mathbb{R}^n$.

$$\det(\Lambda^* \cap W) \overset{\text{Lem 8.3.(c)}}{=} \underbrace{\det(\Lambda \cap W^\perp)}_{\geq 1} \cdot \underbrace{\det(\Lambda^*)}_{=1} \geq 1$$

□

Direct sums of (orthogonal) stable lattices are also stable

Lemma 8.9. *If Λ_1, Λ_2 are stable lattices, then $\Lambda_1 \oplus \Lambda_2$ is stable.*

We skip the proof here.

Lemma 8.10. *For any stable lattice $\Lambda \subseteq \mathbb{R}^n$, there exists a basis $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$, $k = \mathrm{rank}(\Lambda)$ with $\Lambda = \Lambda(\boldsymbol{B})$ so that $\|\boldsymbol{b}_i\|_2 \leq 2k^{1.5}$ for all $i = 1, \ldots, k$. Moreover $\|\boldsymbol{B}\|_F \leq 2n^{2.5}$.*

Proof. It suffices to consider the case where Λ has full rank, i.e. $k = n$. By Lemma 8.8, also Λ^* is stable. Then $\lambda_1(\Lambda^*) \geq 1$. By Banaszczyk's Transference Theorem (Cor 4.2) this means $\lambda_n(\Lambda) \leq 2n$. Then a KZ-reduced basis $\boldsymbol{B}$ has $\|\boldsymbol{b}_i\|_2 \leq \sqrt{n} \cdot \lambda_i(\Lambda) \leq 2n^{3/2}$ for all i (see Lemma 1.40). □

Lemma 8.11. *Let $\Lambda \subseteq \mathbb{R}^n$ be a stable lattice and let $\Lambda' \subseteq \Lambda$ be a primitive sublattice with $\det(\Lambda') = 1$. Then both Λ' and Λ/Λ' are stable.*

Proof. Let $W := \mathrm{span}(\Lambda')$ be the lattice subspace corresponding to Λ'. It is clear that Λ' is stable. Now consider a sublattice of Λ/Λ' which we may write as $\Pi_{W^\perp}(\tilde{\Lambda})$ where $\tilde{\Lambda} \subseteq \Lambda$ is a primitive sublattice. In fact we may assume that $\Lambda' \subseteq \tilde{\Lambda}$ since

$\Pi_{W^\perp}(\Lambda') = \{\mathbf{0}\}$ anyway. Then

$$\det(\Pi_{W^\perp}(\tilde{\Lambda})) \overset{\text{Lem 8.3.(b)}}{=} \frac{\det(\tilde{\Lambda})}{\det(\underbrace{\tilde{\Lambda} \cap W}_{=\Lambda'})} \geq 1$$

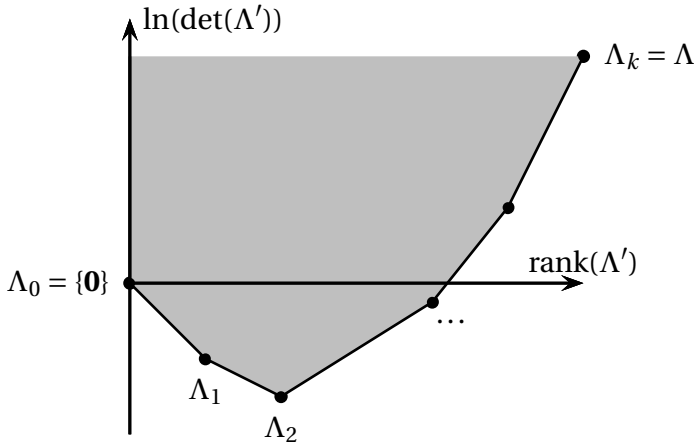
as $\det(\tilde{\Lambda}) \geq 1$ by stability of Λ and $\det(\Lambda') = 1$ by assumption. □

8.3 The canonical filtration of a lattice

Fix any lattice $\Lambda \subseteq \mathbb{R}^n$. Consider the 2-dimensional point set

$$Q := \big\{\big(\text{rank}(\Lambda'), \ln(\det(\Lambda'))\big) \mid \text{sublattice } \Lambda' \subseteq \Lambda\big\}$$

which is called the *canonical plot* of Λ. Here we include the trivial sublattice $\{\mathbf{0}\}$ which has $\det(\{\mathbf{0}\}) = 1$. The lower envelope of $\text{conv}(Q)$ is called the *canonical polygon* of Λ. We will prove that each vertex in the canonical polygon belongs to a *unique* sublattice. Suppose we label the vertices with $0, \ldots, k$ starting with the vertex $(0,0)$ and for the ith vertex, we write Λ_i as the corresponding lattice. Then we will also prove that indeed $\{\mathbf{0}\} = \Lambda_0 \subset \Lambda_1 \subset \ldots \subset \Lambda_k = \Lambda$. This sequence is called the *canonical filtration* of Λ. In particular, every vertex corresponds to a densest sublattice Λ' subject to fixing the rank.



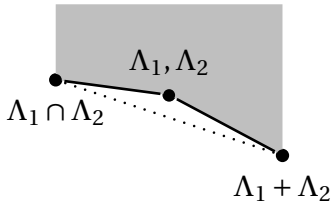
We begin by proving uniqueness:

Lemma 8.12. *Each vertex of the canonical polygon corresponds to a unique sublattice $\Lambda' \subseteq \Lambda$.*

Proof. Suppose for the sake of contradiction that there are two different sublattices $\Lambda_1, \Lambda_2 \subseteq \Lambda$ with $r := \text{rank}(\Lambda_1) = \text{rank}(\Lambda_2)$ and $D := \det(\Lambda_1) = \det(\Lambda_2)$. Then by Lemma 8.6 we have $\frac{1}{2}\text{rank}(\Lambda_1 \cap \Lambda_2) + \frac{1}{2}\text{rank}(\Lambda_1 + \Lambda_2) = r$ and

$$\frac{1}{2}\ln(\det(\Lambda_1 \cap \Lambda_2)) + \frac{1}{2}\ln(\det(\Lambda_1 + \Lambda_2)) \leq \ln(D),$$

which means that the line segment connecting the points of $\Lambda_1 \cap \Lambda_2$ and $\Lambda_1 + \Lambda_2$ goes below (or through) the point of Λ_1, Λ_2. This is a contradiction to the points of Λ_1 and Λ_2 being extreme points of the canonical polygon.

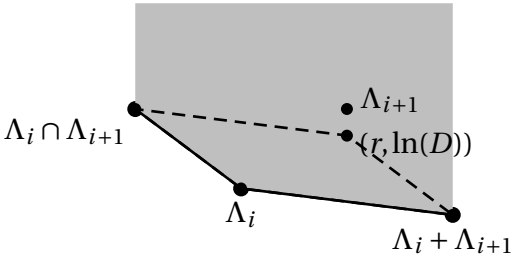


□

Next we prove that the vertices correspond to a *chain* of sublattices:

Lemma 8.13. *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $\Lambda_0, \ldots, \Lambda_k$ be the unique lattices corresponding to vertices of the canonical polygon (in order of increasing rank). Then $\{\mathbf{0}\} = \Lambda_0 \subset \Lambda_1 \subset \ldots \subset \Lambda_k = \Lambda$.*

Proof. Suppose for the sake of contradiction that $\Lambda_i \not\subset \Lambda_{i+1}$ for some index i. Consider the 3 distinct points in the canonical plot belonging to $\Lambda_i \cap \Lambda_{i+1}, \Lambda_i, \Lambda_i + \Lambda_{i+1}$. These 3 points form a parallelogram together with a 4th point whose coordinates we write as $(r, \ln(D))$. Then by Lemma 8.6 we have $r = \text{rank}(\Lambda_{i+1})$ and $D \leq \det(\Lambda_{i+1})$. That is a contradiction to Λ_{i+1} being a vertex of the canonical plot.



□

Theorem 8.14 (Properties of the canonical filtration)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $\{\mathbf{0}\} = \Lambda_0 \subset \Lambda_1 \subset \ldots \subset \Lambda_k = \Lambda$ be the canonical filtration. Then*

(A) Setting $r_i := \det(\Lambda_i / \Lambda_{i-1})^{1/\text{rank}(\Lambda_i \setminus \Lambda_{i-1})}$, one has $r_1 < r_2 < \ldots < r_k$.

(B) *For any sublattice* $\tilde{\Lambda} \subseteq \Lambda$ *and any* $i \in \{1,\dots,k\}$, $\det(\tilde{\Lambda}) \geq \det(\Lambda_i) \cdot r_i^{\mathrm{rank}(\tilde{\Lambda})-\mathrm{rank}(\Lambda_i)}$.

(C) *For any sublattice* $\tilde{\Lambda} \subseteq \Lambda$ *and any* $i \in \{1,\dots,k\}$, $\det(\tilde{\Lambda}) \geq \det(\Lambda_{i-1}) \cdot r_i^{\mathrm{rank}(\tilde{\Lambda})-\mathrm{rank}(\Lambda_{i-1})}$.

(D) *Each quotient lattice* Λ_i/Λ_{i-1} *is a scaling of a stable lattice.*

Proof. **For (A).** Using that $\mathrm{rank}(\Lambda_i/\Lambda_{i-1}) = \mathrm{rank}(\Lambda_i) - \mathrm{rank}(\Lambda_{i-1})$ and $\det(\Lambda_i/\Lambda_{i-1}) = \det(\Lambda_i)/\det(\Lambda_{i-1})$ we know that

$$\ln(r_i) = \frac{\ln(\det(\Lambda_i)) - \ln(\det(\Lambda_{i-1}))}{\mathrm{rank}(\Lambda_i) - \mathrm{rank}(\Lambda_{i-1})}$$

That means the quantity $\ln(r_i)$ is exactly the slope of the canonical polygon between the vertices Λ_{i-1} and Λ_i and that slope must be increasing by convexity!
For (B). By convexity, the point $(\mathrm{rank}(\tilde{\Lambda}), \ln(\det(\tilde{\Lambda})))$ has to lie above the line with slope $\ln(r_i)$ that goes through the point $(\mathrm{rank}(\Lambda_i), \ln(\det(\Lambda_i)))$. That means

$$\begin{aligned} & \ln(\det(\tilde{\Lambda})) \geq \ln(r_i) \cdot (\mathrm{rank}(\tilde{\Lambda}) - \mathrm{rank}(\Lambda_i)) + \ln(\det(\Lambda_i)) \\ \Leftrightarrow \quad & \det(\tilde{\Lambda}) \geq r_i^{\mathrm{rank}(\tilde{\Lambda})-\mathrm{rank}(\Lambda_i)} \cdot \det(\Lambda_i) \end{aligned}$$

For (C). Same argument as (B) with the preceeding point $(\mathrm{rank}(\Lambda_{i-1}), \ln(\det(\Lambda_{i-1})))$.
For (D). We fix an index i. The claim is invariant under scaling Λ, so we assume that $\det(\Lambda_i/\Lambda_{i-1}) = 1$ and so $r_i = 1$. Now consider a primitive sublattice of the quotient lattice Λ_i/Λ_{i-1}; such a sublattice can be written as $\Pi_{\mathrm{span}(\Lambda_{i-1})^\perp}(\Lambda')$ where $\Lambda' \subseteq \Lambda_i$. Then

$$\det\left(\Pi_{\mathrm{span}(\Lambda_{i-1})^\perp}(\Lambda')\right) \overset{\text{Lem 8.5.(a)}}{=} \frac{\det(\Lambda')}{\det(\Lambda_{i-1})} \overset{(C)}{\geq} r_i^{\mathrm{rank}(\Lambda')-\mathrm{rank}(\Lambda_{i-1})} = 1$$

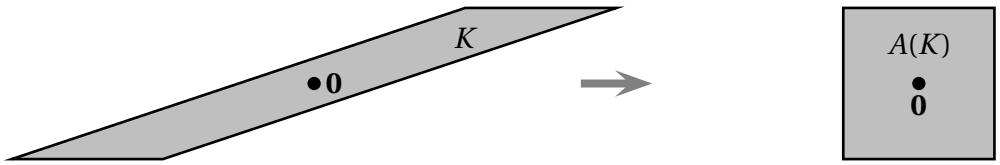
□

8.4 The Gaussian isotropic position

In this section, we discuss how to put an arbitrary symmetric convex body in a suitable normal position. Recall from earlier that the *standard Gaussian distribution* γ_n is the distribution with density function $\gamma_n(\boldsymbol{x}) = (2\pi)^{-n/2} \cdot e^{-\|\boldsymbol{x}\|_2^2/2}$ for all $\boldsymbol{x} \in \mathbb{R}^n$. It will be convenient to also define a distribution $\gamma_{n,s}$ with density function $\gamma_{n,s}(\boldsymbol{x}) = s^{-n} e^{-\pi\|\boldsymbol{x}/s\|_2^2}$ for $\boldsymbol{x} \in \mathbb{R}^n$ where $s > 0$. In this notation one has $\gamma_n = \gamma_{n,\sqrt{2\pi}}$ which seems slightly odd; but this way $\gamma_{n,1}$ has the same density function as the discrete Gaussian ρ_1 that we have frequently used while γ_n is the "standard" distribution studied in convex geometry. For later reference, note that

$$\gamma_{n,s}(U) = \int_{U/s} \rho_1(\boldsymbol{x})\, d\boldsymbol{x} = \gamma_n\left(\frac{\sqrt{2\pi}}{s} U\right) \tag{8.1}$$

for any measurable set $U \subseteq \mathbb{R}^n$. By a slight abuse of notation we will consider a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ also as the linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ with $A(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ (we omit the bold font when we refer to the map). Note that for any measurable set K and any matrix $\boldsymbol{A}$ with $|\det(\boldsymbol{A})| = 1$ one has $\mathrm{Vol}_n(A(K)) = |\det(\boldsymbol{A})| \cdot \mathrm{Vol}_n(K) = \mathrm{Vol}_n(K)$. In other words, determinant-one matrices correspond to *volume preserving* linear transformations. On the other hand, the Gaussian density $\gamma_n(\boldsymbol{x})$ is a lot higher if $\|\boldsymbol{x}\|_2$ is small and $\gamma_n(A(K))$ is not invariant under determinant-one maps. Intuitively, a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $|\det(\boldsymbol{A})| = 1$ that maximizes $\gamma_n(A(K))$ is one that makes K "well-rounded".



A powerful result that is closely related to the techniques mentioned in Section 4.4.3 is the following:

Theorem 8.15. *For any symmetric convex body $K \subseteq \mathbb{R}^n$ with $\mathrm{Vol}_n(K) = 1$ there is a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $|\det(\boldsymbol{A})| = 1$ so that $\gamma_{n,1/t}(A(K)) \geq \frac{2}{3}$ where $t = C\log(n)$ with $C > 0$ large enough.*

Proof. For a convex body Q we define the *width* as $w(Q) = \max_{\boldsymbol{\theta}\sim S^{n-1}}\{|\langle \boldsymbol{\theta}, \boldsymbol{x} - \boldsymbol{y}\rangle| : \boldsymbol{x}, \boldsymbol{y} \in Q\}$, which is a standard quantity studied in convex geometry. It is known that among convex bodies with a fixed volume, Euclidean balls minimize the width, see e.g. [AAGM15]:

Fact I (Urysohn's Inequality). *For any convex body Q one has $w(Q) \geq 2(\frac{Vol_n(Q)}{Vol_n(B_2)})^{1/n}$.*

As $\|\cdot\|_{K^\circ}$ is the dual norm of $\|\cdot\|_K$ (i.e. $\|\boldsymbol{x}\|_{K^\circ} = \max\{\langle \boldsymbol{x}, \boldsymbol{y}\rangle : \boldsymbol{y} \in K\}$), we have $w(K) = 2\,\mathbb{E}_{\boldsymbol{x}\sim S^{n-1}}[\|\boldsymbol{x}\|_{K^\circ}]$. By going from S^{n-1} to γ_n and using concentration it is not hard to show that $w(K) = \Theta(\frac{1}{\sqrt{n}})\ell_{K^\circ}$. Then Theorem 4.39 can be rephrased as:

Fact II ($\ell\ell^\circ$-estimate). *For any symmetric convex body $Q \subseteq \mathbb{R}^n$ there is an invertible linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ so that $w(A(K)) \cdot w(A(K)^\circ) \leq O(\log n)$.*

We use Fact II to find a linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ with $w(A(K)) \cdot w(A(K)^\circ) \leq O(\log n)$. By scaling we may assume that $|\det(\boldsymbol{A})| = 1$ and so $\mathrm{Vol}_n(A(K)) = 1$. Then Fact I gives that

$$w(A(K)) \geq 2 \cdot \left(\frac{\mathrm{Vol}_n(A(K))}{\mathrm{Vol}_n(B_2^n)}\right)^{1/n} \geq \frac{2}{\mathrm{Vol}_n(B_2^n)^{1/n}} \geq \Theta(\sqrt{n}).$$

This implies that $w(A(K)^\circ) \leq O(\frac{\log n}{\sqrt{n}})$. Hence

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_{A(K)}] = \Theta(\sqrt{n}) \cdot w(A(K)^\circ) \leq C\log(n)$$

for some constant $C > 0$. Setting $t := 3C\log(n)$ one has that $\Pr_{\boldsymbol{x}\sim\gamma_n}[\|\boldsymbol{x}\|_{A(K)} \leq t] \geq \frac{2}{3}$ by Markov's Inequality. Then then implies that $\gamma_{n,1/t}(A(K)) = \gamma_n(\sqrt{2\pi}\, t A(K)) \geq \gamma_n(tA(K)) \geq \frac{2}{3}$ using (8.1). □

We would like to mention that it is a central open question in convex geometry whether the $O(\log n)$ term from Fact II can be improved to $O(\sqrt{\log n})$; the $\Theta(\sqrt{\log n})$ factor would be tight even for the cube $K = [-\frac{1}{2}, \frac{1}{2}]^n$.

We say that a measurable set $U \subseteq \mathbb{R}^n$ is in *isotropic t-Gaussian position* if

$$\int_U \rho_t(\boldsymbol{x}) \cdot \boldsymbol{x}\boldsymbol{x}^T \, d\boldsymbol{x} = \alpha \boldsymbol{I}_n$$

for some $\alpha > 0$. We will only consider this notion for symmetric convex bodies K. Intuitively, being isotropic t-Gaussian means that the body K has all its mass (if weighted by $\rho_t(\boldsymbol{x})$) equally spread in all directions. We will require a result of Bobkov which says that a body in isotropic t-Gaussian position maximizes the Gaussian measure:

Theorem 8.16 (Bobkov [Bob11])**.** *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body and let $s > 0$. If K is in isotropic s-Gaussian position then $\gamma_s(K) \geq \gamma_s(A(K))$ for all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $|\det(\boldsymbol{A})| = 1$.*

This is a highly non-trivial fact and it can be derived from the *(B) conjecture*, posed by Banaszczyk and proven by Cordero-Erausquin, Fradelizi and Maurey [CEFM04]. The original conjecture says that for any symmetric convex body $K \subseteq \mathbb{R}^n$, the map $F : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $F(t) = \gamma_n(e^t \cdot K)$ is log-concave. In fact, [CEFM04] prove the more general statement that also the function $G : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $G(t) := \gamma_n(\{(e^{t_1}x_1, \ldots, e^{t_n}x_n) : x \in K\})$ is log-concave. This more general statement can be used to derive Theorem 8.16. Bobkob's Theorem then quickly implies that any body in isotropic Gaussian position must have large Gaussian measure:

Theorem 8.17. *Let $t \geq C\log(n)$ with $C > 0$ large enough and let K be a symmetric convex body with $\mathrm{Vol}_n(K) \geq 1$ that is in isotropic $\frac{1}{t}$-Gaussian position. Then $\gamma_{n,1/t}(K) \geq \frac{2}{3}$.*

Proof. We use Theorem 8.15 and fix a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $|\det(\boldsymbol{A})| = 1$ so that $\gamma_{n,1/t}(A(K)) \geq \frac{2}{3}$. Then as K is in isotropic $\frac{1}{t}$-Gaussian position we know by Bobkov's Theorem (Theorem 8.16) that $\gamma_{n,1/t}(K) \geq \gamma_{n,1/t}(A(K)) \geq \frac{2}{3}$. □

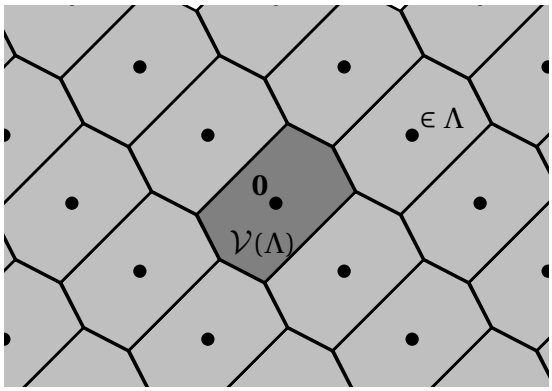
We can make use of a strong concentration of measure effect for Gaussians:

Lemma 8.18 (Boosting)**.** *Let K be a symmetric convex body with $rB_2^n \subseteq K$ and $\gamma_{n,1/t}(K) \geq \frac{1}{2}$. Then $\gamma_{n,\frac{1}{(1+\lambda)t}}(K) \geq 1 - \exp(-\Theta(r^2t^2\lambda^2))$ for any $\lambda > 0$.*

Proof. As we may absorb the $\sqrt{2\pi}$ into the unspecified constant, the claim is equivalent to: *For any symmetric convex K with $rB_2^n \subseteq K$ and $\gamma_n(tK) \geq \frac{1}{2}$ one has $\gamma_n((1+\lambda)tK) \geq 1 - \exp(-\Theta(r^2t^2\lambda^2))$.* A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *L-Lipschitz*, if $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_2$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We use the following standard fact (see e.g. [AAGM15]):
Fact. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be an L-Lipschitz function and let $\lambda \geq 0$. Then $\Pr_{\boldsymbol{x}\sim\gamma_n}[f(\boldsymbol{x}) \geq$ median$(f) + \lambda L] \leq 2\exp(-\lambda^2/4)$.*
Here the median is the value with $\Pr_{\boldsymbol{x}\sim\gamma_n}[f(\boldsymbol{x}) \geq \text{median}(f)] = \frac{1}{2}$. The function that we consider is $f(\boldsymbol{x}) := \|\boldsymbol{x}\|_{tK}$. As $\gamma_n(tK) \geq \frac{1}{2}$ we know that median$(f) \leq 1$. Moreover, as $trB_2^n \subseteq tK$, the function f is $\frac{1}{tr}$-Lipschitz. Then $\Pr_{\boldsymbol{x}\sim\gamma_n}[f(\boldsymbol{x}) \geq 1 + \lambda] \leq 2\exp(-(rt)^2\lambda^2/4)$ as claimed. □

8.5 Gaussian measure of the Voronoi cell

In this section, we revisit the Voronoi cell $\mathcal{V}(\Lambda) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{y}, \boldsymbol{x}\rangle \leq \frac{\|\boldsymbol{y}\|_2^2}{2}\ \forall \boldsymbol{y} \in \Lambda\}$ of a lattice Λ.



Ultimately it will be our goal to prove an *upper bound* on $\rho_s(\Lambda)$ — we will now justify that this is the same as proving a *lower bound* on $\gamma_{n,s}(\mathcal{V}(\Lambda))$.

Lemma 8.19. *For any lattice $\Lambda \subseteq \mathbb{R}^n$ and any $s > 0$ one has $\rho_s(\Lambda) \cdot \gamma_{n,s}(\mathcal{V}(\Lambda)) \leq 1$.*

Proof. After scaling we may assume $s = 1$. We recall that the translates $\boldsymbol{y} + \mathcal{V}(\Lambda)$

with $\boldsymbol{y} \in \Lambda$ form a *tiling* of $\mathbb{R}^n$. We use this to estimate:

$$\begin{aligned} 1 &\overset{\text{tiling}}{=} \sum_{\boldsymbol{y}\in\Lambda} \int_{\mathcal{V}(\Lambda)} e^{-\pi\|\boldsymbol{y}+\boldsymbol{t}\|_2^2} d\boldsymbol{t} \\ &\overset{\text{symmetry}}{=} \sum_{\boldsymbol{y}\in\Lambda} e^{-\pi\|\boldsymbol{y}\|_2^2} \int_{\mathcal{V}(\Lambda)} e^{-\pi\|\boldsymbol{t}\|_2^2} \cdot \underbrace{\left(\frac{1}{2}e^{2\pi\langle \boldsymbol{y},\boldsymbol{t}\rangle} + \frac{1}{2}e^{2\pi\langle \boldsymbol{y},-\boldsymbol{t}\rangle}\right)}_{\geq 1} d\boldsymbol{t} \\ &\geq \rho_1(\Lambda)\cdot\gamma_{n,1}(\mathcal{V}(\Lambda)) \end{aligned}$$

□

We will need to understand how local optima of $\gamma_{n,1/t}(\mathcal{V}(\Lambda(\boldsymbol{B})))$ look like, subject to fixing the determinant of $\boldsymbol{B}$. This is the part where we will skip details in order to keep the exposition short. For those details, we recommend the 2022 Arxiv update of [RS17] which contains a streamlined proof. First, we verify this for an arbitrary set U:

Lemma 8.20. *Let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ continuously differentiable and let $U \subset \mathbb{R}^n$ be a bounded measurable set. Consider the functions $h : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined by*

$$h(\boldsymbol{A}) := \frac{1}{|\det(\boldsymbol{A})|}\int_{A(U)} f(\|\boldsymbol{x}\|_2^2) d\boldsymbol{x}$$

Then h is differentiable at $\boldsymbol{A} = \boldsymbol{I}_n$ with

$$\nabla_{\boldsymbol{A}} h(\boldsymbol{A})_{|\boldsymbol{A}=\boldsymbol{I}_n} = 2\int_U f'(\|\boldsymbol{x}\|_2^2)\boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x}$$

Proof. By a change of variables we obtain $h(\boldsymbol{A}) = \int_U f(\|\boldsymbol{A}\boldsymbol{x}\|_2^2)\, d\boldsymbol{x}$. Then differentiating gives

$$\nabla h(\boldsymbol{A}) = \int_U \left(\nabla_{\boldsymbol{A}} f(\|\boldsymbol{A}\boldsymbol{x}\|_2^2)\right) d\boldsymbol{x} = 2\int_U f'(\|\boldsymbol{A}\boldsymbol{x}\|_2^2)\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x}$$

Then setting $\boldsymbol{A} := \boldsymbol{I}_n$ gives the claim, □

Hence we may conclude what the derivative is for the Voronoi cell:

Lemma 8.21. *Let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ continuously differentiable and let $\Lambda \subseteq \mathbb{R}^n$ be a lattice. Consider the function $g : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined by*

$$g(\boldsymbol{A}) := \frac{1}{|\det(\boldsymbol{A})|}\int_{\mathcal{V}(A(\Lambda))} f(\|\boldsymbol{x}\|_2^2) d\boldsymbol{x}$$

Then g is differentiable at $\boldsymbol{A} = \boldsymbol{I}_n$ with

$$\nabla_{\boldsymbol{A}} g(\boldsymbol{A})_{|\boldsymbol{A}=\boldsymbol{I}_n} = 2\int_{\mathcal{V}(\Lambda)} f'(\|\boldsymbol{x}\|_2^2)\boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x}$$

Proof sketch. We have

$$\begin{aligned} g(\boldsymbol{A}) &= \frac{1}{|\det(\boldsymbol{A})|}\int_{\mathcal{V}(A(\Lambda))} f(\|\boldsymbol{x}\|_2^2)\,d\boldsymbol{x} \\ &\overset{(*)}{=} \frac{1}{|\det(\boldsymbol{A})|}\int_{\mathcal{V}(A(\Lambda))} f(d(\boldsymbol{x},A(\Lambda))^2)\,d\boldsymbol{x} \\ &\overset{(**)}{=} \frac{1}{|\det(\boldsymbol{A})|}\int_{A(\mathcal{V}(\Lambda))} f(d(\boldsymbol{x},A(\Lambda))^2)\,d\boldsymbol{x} \\ &\overset{(***)}{=} \frac{1}{|\det(\boldsymbol{A})|}\int_{A(\mathcal{V}(\Lambda))} f(\|\boldsymbol{x}\|_2^2)d\boldsymbol{x} \quad \pm C_0(\Lambda)\cdot\|\boldsymbol{A}-\boldsymbol{I}_n\|_{\mathrm{op}}^2 \end{aligned}$$

Here we use in $(*)$ that for any $\boldsymbol{x}\in\mathcal{V}(A(\Lambda))$, $\boldsymbol{0}$ is the closest lattice point and so $\|\boldsymbol{x}\|_2 = d(\boldsymbol{x},\mathcal{V}(A(\Lambda)))$. For $(**)$ note that the function $\boldsymbol{x}\mapsto f(d(\boldsymbol{x},A(\Lambda))^2)$ is periodic over the lattice $A(\Lambda)$ and hence the integral is the same for any fundamental body (which is a body K so that $K+A(\Lambda)$ is a tiling of $\mathbb{R}^n$); we use that with the fact that $A(\mathcal{V}(\Lambda))$ is a fundamental body for the lattice $A(\Lambda)$ (even if it might not be the Voronoi cell). We skip the estimates that justify $(***)$ and rather refer to [RS17]. Applying Lemma 8.20 then gives the claim. □

8.6 Proof of the Reverse Minkowski Theorem

The crucial part of the proof is arguing that any stable lattice has a large Voronoi cell. Here we could have picked any polynomially small probability by adjusting the constant C.

Lemma 8.22. *Let $\Lambda\subseteq\mathbb{R}^n$ be a stable lattice. Then $\gamma_{n,1/t}(\mathcal{V}(\Lambda))\geq\exp(-\frac{\mathrm{rank}(\Lambda)}{4n^2})\geq\frac{2}{3}$ for $t:=C\log(n)$ where $C>0$ is a large enough constant.*

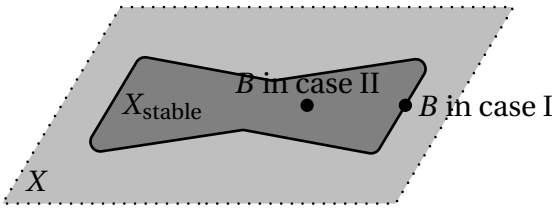
Proof. It suffices to prove the following statement where $C>0$ is a large enough constant:

Claim. *For all $1\leq n\leq N$ and any full rank stable lattice $\Lambda\subseteq\mathbb{R}^n$ one has $\gamma_{n,1/t}(2\mathcal{V}(\Lambda))\geq\exp(-\frac{n}{4N^2})$ where $t=C\ln(N)$.*

The proof works by induction over n. The induction hypothesis is chosen so that we gain some slack as the rank of the considered lattices increases. The claim is clear for $n=1$ where $2t\mathcal{V}(\Lambda)$ is a symmetric interval of length $\Theta(\log n)$, hence let $n\geq 2$. We consider the function $F(\boldsymbol{B}):=\gamma_{n,1/t}(\mathcal{V}(\Lambda(\boldsymbol{B})))$. We define two domains

$$\begin{aligned} X &:= \{\boldsymbol{B}\in\mathbb{R}^{n\times n}\mid\det(\boldsymbol{B})=1\} \\ X_{\mathrm{stable}} &:= \{\boldsymbol{B}\in\mathbb{R}^{n\times n}\mid\Lambda(\boldsymbol{B})\text{ is stable and }\|\boldsymbol{B}\|_F\leq 3n^{2.5}\} \end{aligned}$$

Here, X is a (n^2-1)-dimensional unbounded manifold while the set X_{stable} is closed and bounded, hence it is compact. The function F is continuous and hence there is a minimizer $\boldsymbol{B} \in X_{\text{stable}}$. We set $\Lambda := \Lambda(\boldsymbol{B})$. By Lemma 8.10, we may assume that $\|\boldsymbol{B}\|_F \le 2n^{2.5}$ (otherwise replace $\boldsymbol{B}$ with such a basis). Then $\boldsymbol{B}$ is either in the interior of X_{stable} or if it is on the boundary then there is a sublattice $\Lambda' \subseteq \Lambda(\boldsymbol{B})$ with $\det(\Lambda') = 1$. We try a schematic picture:



Hence we may distinguish two cases:

- **Case I. There is a sublattice $\{\mathbf{0}\} \subset \Lambda' \subset \Lambda$ with $\det(\Lambda') = 1$.** Fix that sublattice Λ' with $\det(\Lambda') = 1$. Then by Lemma 8.11, both Λ' as well as the quotient lattice Λ/Λ' are stable. By induction we have

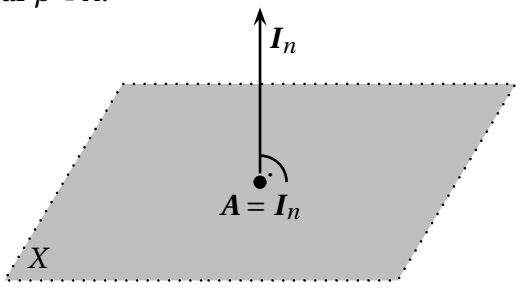
$$\begin{aligned}\gamma_{n,1/t}(2\mathcal{V}(\Lambda)) &\overset{\text{Lem 8.5.(e)}}{\ge} \gamma_{n,1/t}(2\mathcal{V}(\Lambda')) \cdot \gamma_{n,1/t}(2\mathcal{V}(\Lambda/\Lambda')) \\ &\overset{\text{induction}}{\ge} \exp\Big(-\frac{\operatorname{rank}(\Lambda')}{4N^2}\Big) \cdot \exp\Big(-\frac{\operatorname{rank}(\Lambda/\Lambda')}{4N^2}\Big) = \exp\Big(-\frac{n}{4N^2}\Big)\end{aligned}$$

  using that $\operatorname{rank}(\Lambda') + \operatorname{rank}(\Lambda'/\Lambda) = n$.

- **Case II. For all sublattices $\{\mathbf{0}\} \subset \Lambda' \subset \Lambda$ one has $\det(\Lambda') > 1$.** Then the matrix $\boldsymbol{B}$ does not lies on the boundary of X_{stable}. In particular, $\boldsymbol{B}$ is also a *local minimum* of the extended function $F : X \to \mathbb{R}$. Now consider the function $G : \mathbb{R}^{n\times n} \to \mathbb{R}$ with

$$G(\boldsymbol{A}) := \frac{1}{|\det(\boldsymbol{A})|} F(\boldsymbol{AB}) = \frac{1}{|\det(\boldsymbol{A})|} \gamma_{n,1/t}(\mathcal{V}(A(\Lambda))).$$

  Then the restriction $G_{|X}$ has a local minimum at $\boldsymbol{A} = \boldsymbol{I}_n$. Since X is a manifold defined by the single equation $\det(\boldsymbol{A}) = 1$ and the gradient of $\boldsymbol{A} \mapsto \det(\boldsymbol{A})$ at the identity is $(\nabla_{\boldsymbol{A}} \det(\boldsymbol{A}))_{|\boldsymbol{A}=\boldsymbol{I}_n} = \boldsymbol{I}_n$, we may conclude that $(\nabla_{\boldsymbol{A}} G(\boldsymbol{A}))_{|\boldsymbol{A}=\boldsymbol{I}_n} = \beta \boldsymbol{I}_n$ for some scalar $\beta \in \mathbb{R}$.



Next, we want to rewrite the gradient of G using Lemma 8.21. Recall that $\rho_{1/t}(\boldsymbol{x}) = e^{-\pi t^2 \|\boldsymbol{x}\|_2^2}$ and $\rho_{1/t}(\mathbb{R}^n) = t^{-n}$. So the right choice will be to set $f(y) := t^n \cdot e^{-\pi t^2 y}$ so that $f(\|\boldsymbol{x}\|_2^2) = \frac{\rho_{1/t}(\boldsymbol{x})}{\rho_{1/t}(\mathbb{R}^n)}$ and consequently we have

$$G(\boldsymbol{A}) = \frac{1}{|\det(\boldsymbol{A})|}\gamma_{n,1/t}(\mathcal{V}(A(\Lambda))) = \frac{1}{|\det(\boldsymbol{A})|}\int_{\mathcal{V}(A(\Lambda))} f(\|\boldsymbol{x}\|_2^2)d\boldsymbol{x}$$

We verify that $f'(y) = -\pi t^{n+2} e^{-\pi t^2 y}$ and so $f'(\|\boldsymbol{x}\|_2^2) = -\pi t^2 \cdot \frac{\rho_{1/t}(\boldsymbol{x})}{\rho_{1/t}(\mathbb{R}^n)}$. Then

$$\beta \boldsymbol{I}_n = (\nabla_{\boldsymbol{A}} G(\boldsymbol{A}))_{|\boldsymbol{A}=\boldsymbol{I}_n} \overset{\text{Lem 8.21}}{=} 2\int_{\mathcal{V}(\Lambda)} f'(\|\boldsymbol{x}\|_2^2)\boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x} = -\frac{2\pi t^2}{\rho_{1/t}(\mathbb{R}^n)}\int_{\mathcal{V}(\Lambda)} \rho_{1/t}(\boldsymbol{x})\cdot \boldsymbol{x}\boldsymbol{x}^T d\boldsymbol{x}$$

Hence $\mathcal{V}(\Lambda)$ is in isotropic $\frac{1}{t}$-Gaussian position. Moreover $\mathrm{Vol}_n(\mathcal{V}(\Lambda)) = \det(\Lambda) = 1$. Then by Theorem 8.17 one has $\gamma_{1/t}(\mathcal{V}(\Lambda)) \geq \frac{2}{3}$. Since Λ is stable, the shortest vector must have length $\lambda_1(\Lambda) \geq 1$ and consequently $\frac{1}{2}B_2^n \subseteq \mathcal{V}(\Lambda)$. Then applying the boosting from Lemma 8.18 with $r = \frac{1}{2}$ gives that $\gamma_{n,1/t}(2\mathcal{V}(\Lambda)) \geq \exp(-\frac{1}{4N^2})$.

□

We restate and prove the first main theorem of this chapter:

Theorem (Reverse Minkowski Theorem (Theorem 8.1))**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a lattice that satisfies $\det(\Lambda') \geq 1$ for all sublattices $\Lambda' \subseteq \Lambda$. Then for $t = C\log n$ with $C > 0$ large enough one has $\rho_{1/t}(\Lambda) \leq \frac{3}{2}$.*

Proof. First consider a stable lattice $\Lambda \subseteq \mathbb{R}^n$. Then

$$\rho_{1/t}(\Lambda)\cdot \exp\left(-\frac{\mathrm{rank}(\Lambda)}{4n^2}\right) \overset{\text{Lem 8.22}}{\leq} \rho_{1/t}(\Lambda)\cdot\gamma_{n,1/t}(\mathcal{V}(\Lambda)) \overset{\text{Lem 8.19}}{\leq} 1.$$

and hence $\rho_{1/t}(\Lambda) \leq \exp(\frac{\mathrm{rank}(\Lambda)}{4n^2})$.

Now let Λ be an arbitrary lattice Λ that satisfies the assumption. Consider the canonical filtration $\{\boldsymbol{0}\} = \Lambda_0 \subset \Lambda_1 \subset \ldots \subset \Lambda_k = \Lambda$ and abbreviate $r_i := \det(\Lambda_i/\Lambda_{i-1})^{1/\mathrm{rank}(\Lambda_i/\Lambda_{i-1})}$. Note that $r_1 < r_2 < \ldots < r_k$ by Theorem 8.14 and $r_1 = \det(\Lambda_1/\Lambda_0) = \det(\Lambda_1) \geq 1$. Hence $r_i \geq 1$ for all $i = 1,\ldots,k$. Also we know by Theorem 8.14 that the scaled

quotient lattice $\frac{1}{r_i} \cdot \Lambda_i / \Lambda_{i-1}$ is stable. Then

$$\begin{aligned}
\rho_{1/t}(\Lambda) &\overset{\text{Lem 8.5.(c)}}{\leq} \prod_{i=1}^{k} \rho_{1/t}(\Lambda_i/\Lambda_{i-1}) \\
&\overset{r_i \geq 1}{\leq} \prod_{i=1}^{k} \rho_{1/t}\Big(\underbrace{\frac{1}{r_i}\Lambda_i/\Lambda_{i-1}}_{\text{stable lattice}}\Big) \\
&\leq \prod_{i=1}^{k} \exp\Big(\frac{\operatorname{rank}(\Lambda_i/\Lambda_{i-1})}{4n^2}\Big) = \exp\Big(\frac{1}{4n}\Big) \leq \frac{3}{2}
\end{aligned}$$

□

8.7 The covering radius

In this section, we will prove that the covering radius $\mu(\Lambda)$ has a surprisingly tight characterization in terms of density of sublattices. We restate an earlier definition and give two equivalent formulations:

Lemma 8.23. *For any lattice $\Lambda \subseteq \mathbb{R}^n$ one has*

$$\mu_{\det}(\Lambda) := \max_{W \subseteq \mathbb{R}^n \text{ subspace}} \sqrt{\dim(W)} \cdot \det(\Pi_W(\Lambda))^{1/\dim(W)} \tag{8.2}$$

$$= \max_{\Lambda' \subseteq \Lambda \text{ primitive}} \sqrt{\mathit{rank}(\Lambda/\Lambda')} \cdot \det(\Lambda/\Lambda')^{1/\mathit{rank}(\Lambda/\Lambda')} \tag{8.3}$$

$$= \max_{\Lambda' \subseteq \Lambda^*} \sqrt{\mathit{rank}(\Lambda')} \cdot \det(\Lambda')^{-1/\mathit{rank}(\Lambda')} \tag{8.4}$$

Proof. The equivalence of (8.2) and (8.3) follows via $W = \operatorname{span}(\Lambda')^\perp$. The equivalence of (8.2) and (8.4) follows from considering $\Pi_W(\Lambda)^* = \Lambda^* \cap W =: \Lambda'$ (see Lemma 8.3). □

Note that $\Pi_W(\Lambda)$ is a lattice obtained by projecting into a subspace W and $\Pi_W(\Lambda)$ might be arbitrarily dense. But $\mu_{\det}(\Lambda)$ is searching for the least dense projected sublattice; equivalently it searches for the densest sublattice of the dual, see (8.4). It might be worth noting that for the lattice $\Lambda = \mathbb{Z}^n$ one has $\lambda_i(\Lambda) = 1$ for all $i = 1, \ldots, n$ while $\mu(\mathbb{Z}^n) = \frac{\sqrt{n}}{2}$. Hence the lower bound of $\mu(\Lambda) \geq \frac{\lambda_n(\Lambda)}{2}$ from Lemma 1.44 is rather poor in this case. On the other hand, we have indeed $\mu_{\det}(\mathbb{Z}^n) = \sqrt{n}$ by choosing $W = \mathbb{R}^n$.

First, we prove two auxiliary results: Recall that $\eta_\varepsilon(\Lambda)$ is the *smoothing radius*, which is the number s so that $\rho_{1/s}(\Lambda^*) = 1 + \varepsilon$.

Lemma 8.24. *For any lattice $\Lambda \subseteq \mathbb{R}^n$ one has $\mu(\Lambda) \le \sqrt{n} \cdot \eta_{1/2}(\Lambda)$.*

Proof. We assume $n \ge 2$. The claim is invariant under scaling, hence let us assume that $\eta_{1/2}(\Lambda) = 1$ which means that $\rho_1(\Lambda^* \setminus \{\mathbf{0}\}) \le \frac{1}{2}$. Suppose for the sake of contradiction that $\mu(\Lambda) > \sqrt{n}$ and so for some shift $\boldsymbol{u} \in \mathbb{R}^n$ one has $(\boldsymbol{u}+\Lambda) \cap \sqrt{n}B_2^n = \emptyset$. Then

$$\frac{1}{3}\rho_1(\Lambda) = \frac{1-1/2}{1+1/2}\cdot\rho_1(\Lambda) \overset{\text{Cor } 4.28}{\le} \rho_1(\boldsymbol{u}+\Lambda) \overset{(\boldsymbol{u}+\Lambda)\cap\sqrt{n}B_2^n=\emptyset}{=} \rho_1((\boldsymbol{u}+\Lambda)\setminus\sqrt{n}B_2^n) \overset{\text{Lem } 4.24}{\le} 2^{-n}\rho_1(\Lambda)$$

This is a contradiction for $n \ge 2$. □

Lemma 8.25. *If $\Lambda \subseteq \mathbb{R}^n$ is the scaling of a stable lattice, then $\mu(\Lambda) \le O(\sqrt{n}\log(n)) \cdot \det(\Lambda)^{1/n}$.*

Proof. By scaling it suffices to show that for any stable lattice Λ one has $\mu(\Lambda) \le O(\sqrt{n}\log(n))$. Let $t := C\log(n)$. By the Reverse Minkowski Theorem (Theorem 8.1) we have $\rho_{1/t}(\Lambda^* \setminus \{\mathbf{0}\}) \le \frac{1}{2}$ and so $\eta_{1/2}(\Lambda) \le t$. Then $\mu(\Lambda) \le \sqrt{n}\cdot\eta_{1/2}(\Lambda) \le O(\sqrt{n}\log(n))$ by Lemma 8.24. □

Recall that the *AMGM inequality* says that the arithmetic mean is at least the geometric mean. Mathematically, for any $a_1, \dots, a_k > 0$ and $d_1, \dots, d_k > 0$ one has

$$\frac{\sum_{i=1}^k d_i a_i}{\sum_{i=1}^k d_i} \ge \Big(\prod_{i=1}^k a_i^{d_i}\Big)^{1/\sum_{i=1}^k d_i}$$

One can prove that the reverse inequality is true up to some logarithmic term. To be precise, we will require a slight variant which is as follows:

Lemma 8.26 (Reverse AMGM)**.** *Let $0 < a_1 < \ldots < a_k$ and $d_1, \dots, d_k \in \mathbb{N}$ and define $m_j := \sum_{i=j}^k d_i$. Then*

$$\sum_{i=1}^k d_i a_i \le O(\log(m_1)) \cdot \max_{j=1,\dots,k}\Big\{m_j \prod_{i\ge j}(a_i^{d_i})^{1/m_j}\Big\}$$

Proof. We may group indices i so that the a_i's are within a factor of 2 together and drop any indices i with $a_i \le \frac{a_k}{2m_1}$ without decreasing the LHS by more than a constant factor. After that change we may assume $k \le O(\log n)$. We fix an index $i^* \in \{1, \dots, k\}$ so that $d_{i^*}a_{i^*} \ge \frac{1}{k}\sum_{i=1}^k d_i a_i$. Then

$$\frac{1}{k}\sum_{i=1}^k d_i a_i \overset{\text{choice of } i^*}{\le} d_{i^*}a_{i^*} \le m_{i^*}a_{i^*} \overset{a_{i^*}\le a_i\,\forall\, i\ge i^*}{\le} m_{i^*}\cdot\Big(\prod_{i\ge i^*} a_i^{d_i}\Big)^{1/m_{i^*}}$$

□

Now we are ready for the 2nd main result which again we restate:

Theorem (Covering radius approximation – Theorem 8.2)**.** *For any lattice $\Lambda \subseteq \mathbb{R}^n$ one has*

$$\Theta(1) \cdot \mu_{\det}(\Lambda) \leq \mu(\Lambda) \leq O(\log^{3/2} n) \cdot \mu_{\det}(\Lambda)$$

Proof. First we prove the lower bound. Let W be the subspace with $k := \dim(W)$ attaining the minimum. Then

$$\mu(\Lambda) \overset{(*)}{\geq} \mu(\Pi_W(\Lambda)) \overset{\text{Lem 1.43}}{\geq} \frac{\sqrt{k}}{6} \cdot \det(\Pi_W(\Lambda))^{1/k}$$

where $(*)$ follows from the fact that the projection preserves a covering, i.e. if $\Lambda + rB_2^n = \mathbb{R}^n$ then certainly also $\Pi_W(\Lambda) + r\Pi_W(B_2^n) = W$. We have also used the determinant-based lower bound from Lemma 1.43.

Now we come to the upper bound. Let $\{\mathbf{0}\} = \Lambda_0 \subset \Lambda_1 \subset \ldots \subset \Lambda_k = \Lambda$ be the canonical filtration, following Theorem 8.14. Let $d_i = \operatorname{rank}(\Lambda_i/\Lambda_{i-1})$. In wise foresight, we define $r_i := \det(\Lambda_i/\Lambda_{i-1})^{1/d_i}$ and recall that $r_1 < r_2 < \ldots < r_k$. Note that each quotient lattice Λ_i/Λ_{i-1} is the scaling of a d_i-dimensional stable lattice and so by Lemma 8.25 we have

$$\mu(\Lambda_i/\Lambda_{i-1}) \leq O(\sqrt{d_i}\log(d_i)) \cdot r_i$$

Then

$$\begin{aligned}
\mu(\Lambda)^2 &\overset{\text{Lem 8.5.(d)}}{\leq} \sum_{i=1}^{k} O(d_i \log^2 d_i) \cdot r_i^2 \\
&\leq O(\log^2 n) \cdot \sum_{i=1}^{k} d_i \cdot r_i^2 \\
&\overset{\text{Reverse AMGM}}{\leq} O(\log^3 n) \cdot \max_{i=1,\ldots,k} \left\{ \Big(\sum_{j \geq i} d_i \Big) \Big(\prod_{j \geq i} (r_j^2)^{d_j} \Big)^{1/\sum_{j\geq i} d_j} \right\} \\
&= O(\log^3 n) \cdot \max_{i=1,\ldots,k} \Big\{ \underbrace{\operatorname{rank}(\Lambda/\Lambda_{i-1}) \cdot \Big(\det(\Lambda/\Lambda_{i-1})^{1/\operatorname{rank}(\Lambda/\Lambda_{i-1})} \Big)^2}_{\leq \mu_{\det}(\Lambda)^2} \Big\} \\
&\leq O(\log^3 n) \cdot \mu_{\det}(\Lambda)^2
\end{aligned}$$

Here we use $\operatorname{rank}(\Lambda/\Lambda_{i-1}) = \sum_{j\geq i} d_j$ and $\prod_{j\geq i} \det(\Lambda_j/\Lambda_{j-1}) = \det(\Lambda/\Lambda_{i-1})$. Then taking the square root gives the claim. □

The proof hides a bit the subspace that certifies an upper bound on $\mu(\Lambda)$. Again, consider the canonical filtration $\{\mathbf{0}\} = \Lambda_0 \subset \Lambda_1 \subset \ldots \subset \Lambda_k = \Lambda$ of an arbitrary lattice Λ, say with $n := \operatorname{rank}(\Lambda)$. We have no control over the number of

lattices in this filtration other than $k \in \{1,\dots,n\}$. We may form groups of all quotient lattices Λ_i/Λ_{i-1} where r_i is within a constant factor. This way we can select indices $\ell(1) < \ell(2) < \dots < \ell(K)$ that give us a "coarser" filtration $\{\mathbf{0}\} = \Lambda_{\ell(1)} \subset \dots \subset \Lambda_{\ell(K)} = \Lambda$ so that $D_i := \operatorname{rank}(\Lambda_{\ell(i)}/\Lambda_{\ell(i-1)})$ and $R_i := \operatorname{rank}(\Lambda_{\ell(i)}/\Lambda_{\ell(i-1)})^{1/D_i}$ satisfies not only $R_1 < R_2 < \dots < R_K$ but also $R_i \le \frac{1}{2}R_{i+1}$. Now we can pick an index i^* that maximizes $D_i R_i^2$. Then

$$\begin{aligned}\mu(\Lambda) &\le \Theta(\log^{3/2} n)\cdot\sqrt{D_{i^*}}\cdot R_{i^*}\\ &\le \Theta(\log^{3/2} n)\cdot\sqrt{\operatorname{rank}(\Lambda/\Lambda_{\ell(i^*-1)})}\cdot\det(\Lambda/\Lambda_{\ell(i^*-1)})^{1/\operatorname{rank}(\Lambda/\Lambda_{\ell(i^*-1)})}\\ &= \Theta(\log^{3/2} n)\cdot\sqrt{\dim(W)}\cdot\det(\Pi_W(\Lambda))^{1/\dim(W)}\end{aligned}$$

Here $W := \operatorname{span}(\Lambda_{\ell(i^*-1)})^{\perp}$ with $\dim(W) = \operatorname{rank}(\Lambda/\Lambda_{\ell(i^*-1)}) = n - \operatorname{rank}(\Lambda_{\ell(i^*-1)})$.

8.7.1 The Kannan-Lovasz Conjecture

Dadush [Dad12] conjectured that the covering radius of any lattice with respect to any convex body is approximated well by a volume/determinant ratio and attributes this to Kannan and Lovász [KL88].

Conjecture 1 ([KL88])**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $K \subseteq \mathbb{R}^n$ be a convex body with $\mathrm{int}(K)\cap\Lambda = \emptyset$. Then there is a subspace $W \subseteq \mathbb{R}^n$ so that for $k := \dim(W)$ one has*

$$\frac{Vol_k(\Pi_W(K))}{\det(\Pi_W(\Lambda))} \le (O(\log n))^k$$

We would like to point out that Theorem 8.2 implies the Kannan-Lovász Conjecture (Conjecture 1) for ellipsoids (with a slightly worse bound).

Corollary 8.27 (Kannan-Lovász Conjecture for Ellipsoids)**.** Let $\Lambda \subseteq \mathbb{R}^n$ be a full rank lattice and let $K = \boldsymbol{u} + \mathcal{E} \subseteq \mathbb{R}^n$ be a shifted ellipsoid with $\mathrm{int}(K)\cap\Lambda = \emptyset$. Then there is a subspace $W \subseteq \mathbb{R}^n$ so that for $k := \dim(W)$ one has

$$\frac{\mathrm{Vol}_k(\Pi_W(K))}{\det(\Pi_W(\Lambda))} \le \left(O(\log^{3/2} n)\right)^k$$

Proof. The claim is invariant under applying a linear transformation, hence we may assume that $K = \boldsymbol{u} + B_2^n$. Then

$$1 \overset{(\boldsymbol{u}+B_2^n)\cap\Lambda=\emptyset}{\le} \mu(\Lambda) \overset{\text{Thm } 8.2}{\le} O(\log^{3/2} n)\cdot\mu_{\det}(\Lambda) = \Theta(\log^{3/2} n)\cdot\sqrt{k}\det(\Pi_W(\Lambda))^{1/k}$$

where we let $W \subseteq \mathbb{R}^n$ be the subspace attaining $\mu_{\det}(\Lambda)$ with $k := \dim(W)$. Note that $\mathrm{Vol}_k(\Pi_W(K)) = \mathrm{Vol}_k(B_2^k) \leq (\frac{6}{\sqrt{k}})^k$. Then

$$\frac{\mathrm{Vol}_k(\Pi_W(K))}{\det(\Pi_W(\Lambda))} \leq \left(\Theta(\log^{3/2} n) \cdot \sqrt{k}\right)^k \cdot \left(\frac{6}{\sqrt{k}}\right)^k = \Theta(\log^{3/2} n)^k$$

as claimed. □

Remark 1. After conclusion of this course Conj 1 was proven with a slightly weaker term of $\left(O(\log^3 n)\right)^k$ by Reis and the author of this text, see [RR23].

Bibliography

[AAGM15] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D. Milman. *Asymptotic geometric analysis. Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[Ajt98] Miklós Ajtai. The shortest vector problem in l2 is np-hard for randomized reductions (extended abstract). In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pages 10–19, New York, NY, USA, 1998. Association for Computing Machinery.

[AKS01] Miklós Ajtai, Ravi Kumar, and D. Sivakumar. A sieve algorithm for the shortest lattice vector problem. In *Proceedings on 33rd Annual ACM Symposium on Theory of Computing, July 6-8, 2001, Heraklion, Crete, Greece*, pages 601–610, 2001.

[AR05] Dorit Aharonov and Oded Regev. Lattice problems in np intersected conp. *J. ACM*, 52(5):749–765, sep 2005.

[Bab86] László Babai. On lovász' lattice reduction and the nearest lattice point problem. *Comb.*, 6(1):1–13, 1986.

[Bal97] Keith Ball. An elementary introduction to modern convex geometry. In *in Flavors of Geometry*, pages 1–58. Univ. Press, 1997.

[Ban93a] W. Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(1):625–635, 1993.

[Ban93b] W. Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(1):625–635, 1993.

[Ban96] Wojciech Banaszczyk. Inequalities for convex bodies and polar reciprocal lattices in r^n II: application of k-convexity. *Discret. Comput. Geom.*, 16(3):305–311, 1996.

[BLPS99] Wojciech Banaszczyk, Alexander E. Litvak, Alain Pajor, and Stanislaw J. Szarek. The flatness theorem for nonsymmetric convex bodies via the local theory of banach spaces. *Math. Oper. Res.*, 24(3):728–750, 1999.

[BN09] Johannes Blömer and Stefanie Naewe. Sampling methods for shortest vectors, closest vectors and successive minima. *Theor. Comput. Sci.*, 410(18):1648–1665, 2009.

[Bob11] Sergey G. Bobkov. On Milman's ellipsoids and M-position of convex bodies. In *Concentration, functional inequalities and isoperimetry*, volume 545 of *Contemp. Math.*, pages 23–33. Amer. Math. Soc., Providence, RI, 2011.

[CEFM04] D Cordero-Erausquin, M Fradelizi, and B Maurey. The (b) conjecture for the gaussian measure of dilates of symmetric convex sets and related problems. *Journal of Functional Analysis*, 214(2):410–427, 2004.

[Dad12] Daniel Nicolas Dadush. *Integer Programming, Lattice Algorithms, and Deterministic Volume Estimation*. PhD thesis, USA, 2012. AAI3531709.

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, jan 1991.

[FTJ79] T. Figiel and Nicole Tomczak-Jaegermann. Projections onto hilbertian subspaces of banach spaces. *Israel Journal of Mathematics*, 33(2):155–171, 1979.

[Kan87a] Ravi Kannan. Minkowski's convex body theorem and integer programming. *Math. Oper. Res.*, 12(3):415–440, August 1987.

[Kan87b] Ravindran Kannan. Algorithmic geometry of numbers. *Annual Review of Comp. Sci*, 2:231–267, 1987.

[KL88] Ravi Kannan and Laszlo Lovasz. Covering minima and lattice-point-free convex bodies. *Annals of Mathematics*, 128(3):577–602, 1988.

[Len83] Jr. Lenstra, H. W. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8(4):pp. 538–548, 1983.

[Lew79] D. R. Lewis. Ellipsoids defined by banach ideal norms. *Mathematika*, 26(1):18–29, 1979.

[Mat02] Jiri Matousek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[MG02] Daniele Micciancio and Shafi Goldwasser. *Complexity of Lattice Problems: a cryptographic perspective*, volume 671 of *The Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers, Boston, Massachusetts, March 2002.

[MH73] John Milnor and Dale Husemoller. *Symmetric bilinear forms*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 73. Springer-Verlag, New York-Heidelberg, 1973.

[MR07a] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM J. Comput.*, 37(1):267–302, 2007.

[MR07b] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM J. Comput.*, 37(1):267–302, 2007.

[MSW21] Lukas Mayrhofer, Jamico Schade, and Stefan Weltge. Lattice-free simplices with lattice width $2d - o(d)$, 2021.

[MV10] Daniele Micciancio and Panagiotis Voulgaris. A deterministic single exponential time algorithm for most lattice problems based on voronoi cell computations. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 351–358, 2010.

[MV13] Daniele Micciancio and Panagiotis Voulgaris. A deterministic single exponential time algorithm for most lattice problems based on voronoi cell computations. *SIAM J. Comput.*, 42(3):1364–1391, 2013.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[Odl90] A. M. Odlyzko. The rise and fall of knapsack cryptosystems. *Cryptology and Computational Number Theory*, pages 75–88, 1990.

[Pei13] Chris Peikert. Lattices in cryptography, 2013.

[Pis80] Gilles Pisier. Un théorème sur les opérateurs linéaires entre espaces de Banach qui se factorisent par un espace de Hilbert. *Annales scientifiques de l'École Normale Supérieure*, 4e série, 13(1):23–43, 1980.

[Reg] O. Regev. The learning with errors problem.

[Reg09a] Oded Regev. Lecture notes on lattices, 2009.

[Reg09b] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6):34:1–34:40, 2009.

[RR23] Victor Reis and Thomas Rothvoss. The subspace flatness conjecture and faster integer programming. *FOCS 2023 (to appear)*, 2023.

[RS17] Oded Regev and Noah Stephens-Davidowitz. A reverse minkowski theorem. In *STOC*, pages 941–953. ACM, 2017.

[Rud98] Mark Rudelson. Distances between non-symmetric convex bodies and the mm*-estimate. *Positivity*, 4:161–178, 1998.

[Sch99] Alexander Schrijver. *Theory of linear and integer programming*. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1999.

[Ste17] Noah Stephens-Davidowitz. *On the Gaussian Measure Over Lattices*. PhD thesis, New York University, USA, 2017.

[Tal87] Michel Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.

[vEB81] P. van Emde-Boas. *Another NP-complete partition problem and the complexity of computing short vectors in a lattice*. Report. Department of Mathematics. University of Amsterdam. Department, Univ., 1981.

[Ver19] Roman Vershynin. High-dimensional probability. 2019.