# Lagrange and Hermite algorithms

## 1 Introduction

In the previous lecture, we motivated the task of basis reduction: solving lattice problems via the Nearest-Plane algorithm or the Fincke-Pohst Enumeration algorithm was more successful or faster when the input basis was balanced, or more precisely when the Gram-Schmidt vectors were as closer in length to one another. More formally, we consider the *profile* of a basis as defined below, and keep in mind the rate of change of the associated function $\{1, \dots, n\} \to \mathbb{R}$ given by $i \mapsto \ell_i$.

Definition 1 *The profile of a basis* $\mathbf{B}$ *of a lattice of rank* $n$ *is the sequence* $(\ell_1, \dots, \ell_n)$ *of logarithms of its Gram-Schmidt norms:*

$$\ell_i = \log \|\mathbf{b}_i^\star\|_2 \quad \text{for } i \in \{1 \dots n\}.$$

The task of lattice reduction is to change the basis $\mathbf{B}$ so as to make the resulting profile as *flat* as possible, in the sense that each of the $\ell_i$ are a similar magnitude.

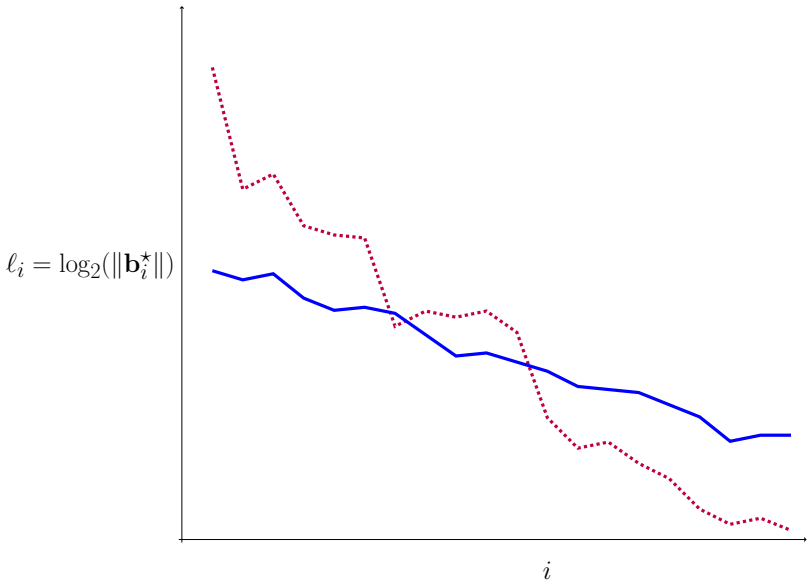


Figure 1: The profile of a basis in dimension 20 before (dashed) and after (plain) a reduction

The volume invariant $\prod \|\mathbf{b}_i^\star\|_2 = \det(L)$ gives us $\sum \ell_i = \log \det(L)$, and thus the area under the plot of $i \mapsto \ell_i$ must remain constant. Also, because the first Gram-Schmidt vector is also a basis vector $\mathbf{b}_1 = \mathbf{b}_1^\star$, a successful reduction of a lattice basis leads to finding a short basis vector. Typically, we start from bases with large vectors $\mathbf{b}_i$. In particular $\mathbf{b}_1$ is large and the later Gram-Schmidt norms are significantly smaller, that is, the profile decreases rapidly.

Note however that there are some lattices that do not admit flat profiles. For example, the lattice generated by $\left(\begin{smallmatrix} \varepsilon & 0 \\ 0 & 1/\varepsilon \end{smallmatrix}\right)$ for $0 < \varepsilon < 1$. For this lattice, there is no basis for which the shortest Gram-Schmidt norm is less than $\varepsilon$. What we can guarantee in general is that the profile does not *decrease* too sharply, but in general, we cannot control sharp *increase* of the profile. Or more

precisely, we can only control them relatively to the so-called successive minima, that we will introduce first.

In this lecture we will see Lagrange reduction, which reduces a 2-dimensional basis. Remarkably, this will always provide a shortest vector of that lattice in polynomial time. Such a Lagrange reduction algorithm will then allow us to generalize reduction notions and algorithms to higher dimensions, using the two dimensional algorithm on projected lattices. These are called *Hermite* and *LLL* reduction. We will prove that the LLL algorithm runs in polynomial time, and provides a solution to $\alpha$-SVP for and approximation factor $\alpha = 2^{O(n)}$ exponential in the dimension $n$.

## 2 Successive Minima

Definition 2 (Successive Minima) *Let $L \subseteq \mathbb{R}^n$ be a rank $k \geq 1$ lattice. For $1 \leq i \leq k$, we define the $i^{th}$ minima of $L$ as*

$$\lambda_i(L) = \inf\{s \geq 0 : \dim(L \cap s\mathfrak{B}^n) \geq i\}.$$

Remark 3 We first note that for $i = 1$, the above definition of $\lambda_1 = \inf\{s \geq 0 : \dim(s\mathfrak{B}^n \cap L) \geq 1\}$ seems somewhat different from the original definition $\lambda_1(L) = \inf_{\mathbf{y} \in L \setminus \{\mathbf{0}\}} \|\mathbf{y}\|$. To see that the definitions are equivalent, note that $\dim(s\mathfrak{B}^n \cap L) \geq 1 \Leftrightarrow \exists \mathbf{y} \in L \setminus \{\mathbf{0}\}$ s.t. $\|\mathbf{y}\| \leq s$. From this, it is direct to see that both definitions yield exactly the same value.

By definition, it is clear that $\lambda_1(L) \leq \lambda_2(L) \leq \cdots \leq \lambda_k(L)$. We now show that successive minima are in fact well-defined, and that there are lattice vectors that attain them.

Lemma 4 *Let $L \subseteq \mathbb{R}^n$ be a $k \geq 1$ dimensional lattice. Then there exists linearly independent vectors $\mathbf{y}_1, \ldots, \mathbf{y}_k \in L$ such that $\|\mathbf{y}_i\| = \lambda_i(L)$. In particular, $\lambda_i(L) < \infty$ for all $i \in [k]$.*

Proof: Let $\mathbf{b}_1, \ldots, \mathbf{b}_k$ denote a basis for $L$. Let $R = \max_{1 \leq i \leq k} \|\mathbf{b}_i\|$. Clearly $\dim(R\mathfrak{B}^n \cap L) = \dim(L) = k$. Therefore, $\lambda_i(L) \leq R$ for all $i \in [k]$. Hence, if there exists $\mathbf{y} \in L$ such that $\|\mathbf{y}\| = \lambda_i(L)$, for any $i \in [k]$, we must have that $\mathbf{y} \in R\mathfrak{B}^n$.

We recursively choose $\mathbf{y}_1, \ldots, \mathbf{y}_k \in L \setminus \{\mathbf{0}\}$ as follows. Let $V_0 = \{\mathbf{0}\}$, and let $\mathbf{y}_1$ be a shortest vector in $(L \cap R\mathfrak{B}^n) \setminus V_0$. For $i$, $2 \leq i \leq k$, let $\mathbf{y}_i$ be the shortest vector in $L \cap R\mathfrak{B}^n \setminus V_{i-1}$ where $V_{i-1} = \mathrm{span}(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$. We note that $\mathbf{y}_1, \ldots, \mathbf{y}_k$ exist since $L \cap R\mathfrak{B}^n$ is finite (by discreteness of $L$) and since $\dim(L \cap R\mathfrak{B}^n) = k$.

We claim that $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are linearly independent and that $\|\mathbf{y}_i\| = \lambda_i(L)$, $i \in [k]$. Since each vector is chosen outside the span of the previous vectors, we have that $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are linearly independent. Therefore $\dim(V_i) = \mathrm{span}(\mathbf{y}_1, \ldots, \mathbf{y}_i) = i$ for $i \in \{0, \ldots, k\}$.

Furthermore, by construction, it is clear that $\|\mathbf{y}_1\| \leq \|\mathbf{y}_2\| \leq \cdots \leq \|\mathbf{y}_k\|$. For $i \in [k]$, let $r_i = \|\mathbf{y}_i\|$. From here see that $\dim(r_i\mathfrak{B}^n \cap L) \geq \dim(V_i) = i$. Hence $r_i = \|\mathbf{y}_i\| \geq \lambda_i(L)$ by definition. We now show that $r_i \leq \lambda_i(L)$. For $i \in [k]$, and $0 < \varepsilon \leq r_i$, take $\mathbf{y} \in L \cap (r_i - \varepsilon)\mathfrak{B}^n$. We claim that $\mathbf{y} \in V_{i-1}$. If not, then by our choice of $\mathbf{y}_i$, we must have that $\|\mathbf{y}_i\| = r_i \leq \|\mathbf{y}\| \leq r_i - \varepsilon < r_i$, a clear contradiction. Therefore $\dim(L \cap (r_i - \varepsilon)\mathfrak{B}^n) \leq \dim(V_{i-1})) = i - 1$, and hence $r_i \leq \lambda_i(L)$ as needed. □

Note that the set of vectors constructed in such a way for $k = n$ is *not* necessarily a basis of $L$, but might instead generate only a a full-rank sublattice of $L$. An example of when this happens is the lattice $D_n = 2\mathbb{Z}^n \cup (2\mathbb{Z}^n + (1, 1, \ldots, 1))$ for $n \geq 5$: in that case we have $\lambda_1(D_n) = \lambda_2(D_n) = \cdots = \lambda_n(D_n) = 2$, but the lattice generated by the $\mathbf{y}_i$'s is exactly $2\mathbb{Z}^n$.

LEMMA 5 *For any lattice $L$ of rank $n$ we have $\prod_{i=1}^{n} \lambda_i^{(2)}(L) \geq \det(L)$.*

PROOF: Consider the lineraly independant vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ as above, and note that they generate a full-rank sublattice $L'$ of $L$, hence $\det(L) \leq \det(L')$. We also have $\det(L') \leq \prod_{i=1}^{n} \|\mathbf{y}_i\|_2$ and we conclude. □

# 3 Lagrange Reduction Algorithm

In this section, we will describe Lagrange reduction algorithm, which finds a basis that is as short as possible for a two-dimensional lattice.

DEFINITION 6 (LAGRANGE REDUCTION) *Let $L$ be a 2-dimensional lattice. A basis $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2)$ of $L$ is called Lagrange Reduced if:*

- $\mathbf{b}_1$ *is a shortest vector of $L$.*
- $|\langle \mathbf{b}_1, \mathbf{b}_2 \rangle| \leq \frac{1}{2}\|\mathbf{b}_1\|_2^2$.

THEOREM 7 (WRISTWATCH LEMMA) *Every 2-dimensional lattice admits a Lagrange-Reduced basis.*

The mnemonic name "wristwatch" [1] refers to the Figure 2. Note that this is indeed a notion of reduction as we discussed in the introduction. Indeed, the basis provided by the Wristwatch lemma implies that $\|\mathbf{b}_2\|_2 \geq \|\mathbf{b}_1\|_2 = \|\mathbf{b}_1^\star\|_2$, where

$$\|\mathbf{b}_2\|_2^2 \leq \|\mathbf{b}_2^\star\|_2^2 + 1/4\|\mathbf{b}_1^\star\|_2^2.$$

In particular $\|\mathbf{b}_2^\star\|_2^2 \geq 3/4\|\mathbf{b}_1^\star\|_2^2$, or equivalently $\ell_2 \geq \ell_1 - \log\sqrt{4/3}$: the profile does not decrease too sharply.

The proof of this theorem is 'by algorithm', which means that we give an algorithm that, given any basis, computes a shortest basis as in the Wristwatch lemma. The proof then consists of showing that the algorithm terminates, and after termination indeed gives the correct basis.

**Algorithm 1:** Lagrange reduction algorithm

**Input :** A basis $(\mathbf{b}_1, \mathbf{b}_2)$ of a lattice $L$.
**Output:** A basis $(\mathbf{b}_1, \mathbf{b}_2)$ as in the Wristwatch lemma.

**repeat**
  swap $\mathbf{b}_1 \leftrightarrow \mathbf{b}_2$
  $k \leftarrow \lceil \frac{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle}{\|\mathbf{b}_1\|_2^2} \rfloor$
  $\mathbf{b}_2 \leftarrow \mathbf{b}_2 - k\mathbf{b}_1$
**until** $\|\mathbf{b}_1\|_2 \leq \|\mathbf{b}_2\|_2$

PROOF: We will prove the lemma by showing Algorithm 1 terminates and is correct.

[1] Some readers may recognise the top 'strap' of the wristwatch as a fundamental domain of the action of $\mathrm{SL}_2(\mathbb{Z})$ on the complex upper half-plane.

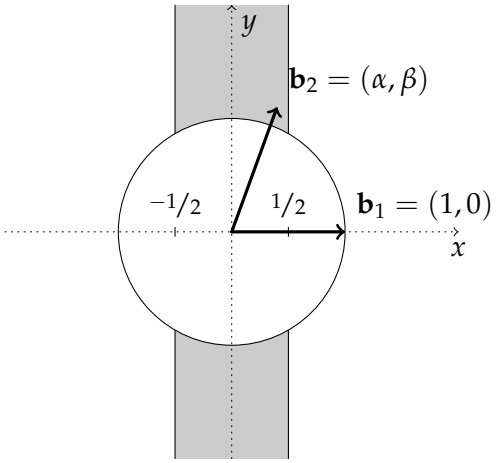


Figure 2: A picture of the wristwatch lemma. The lattice is scaled and rotated such that $\mathbf{b}_1 = (1, 0)$ is of unit length and lies on the $x$-axis. The second basis vector $\mathbf{b}_2$ must then be in one of the two gray areas.

*Algorithm 1 terminates.* This can be seen by the fact that $\|\mathbf{b}_1\|_2$ diminishes after every iteration in the repeat-loop. As $L$ has a minimum non-zero length, and $\mathbf{b}_1$ and $\mathbf{b}_2$ are linearly independent, this necessarily means that the algorithm should terminate.

*The resulting $\mathbf{b}_1, \mathbf{b}_2$ indeed form a basis of the lattice $L$.* This can be seen by the fact that every operation in the loop (swap, row-addition) is a unimodular transformation on the basis, i.e., multiplication by a matrix in $\mathrm{GL}_2(\mathbb{Z})$. The final pair of vectors is therefore a basis, too.

*We indeed have $\langle \mathbf{b}_1, \mathbf{b}_2 \rangle \leq \frac{1}{2}\|\mathbf{b}_1\|_2^2$.* To prove this, we assume, without loss of generality (after scaling and rotating) that $\mathbf{b}_1 = (0, 1)$. Write $\mathbf{b}_2 = (\alpha, \beta)$. Then in the last iteration of the algorithm (omitting the swap), we essentially force that $|\beta| \leq 1/2$. As $\langle \mathbf{b}_1, \mathbf{b}_2 \rangle = \beta$, this proves our claim.

*The vector $\mathbf{b}_1$ is a shortest (non-zero) lattice vector of $L$.* As any vector in $L$ can be written as $\mathbf{v} = m\mathbf{b}_1 + n\mathbf{b}_2$ with $m, n \in \mathbb{Z}$, our aim is to prove that $\|\mathbf{v}\|_2 = \|m\mathbf{b}_1 + n\mathbf{b}_2\|_2 \geq \|\mathbf{b}_1\|_2$. Again, we write $\mathbf{b}_1 = (0, 1)$ and $\mathbf{b}_2 = (\alpha, \beta)$.

For $n = 0$, this is clearly true, as then then $\mathbf{v} = m\mathbf{b}_1$ is a multiple of $\mathbf{b}_1$, which is always at least as long as $\mathbf{b}_1$ itself. The same holds for $m = 0$. For $n, m \neq 0$, we have

$$\|\mathbf{v}\|_2 = \|(n\alpha, m + n\beta)\|_2 = n^2\alpha^2 + (m + n\beta)^2 = n^2(\alpha^2 + \beta^2) + m^2 + 2mn\beta$$

$$\geq n^2 + m^2 - |mn| \geq \min(n^2, m^2) \geq 1 = \|\mathbf{b}_1\|_2.$$

Where the first inequality comes from the fact that $\alpha^2 + \beta^2 = \|\mathbf{b}_2\|_2^2 \geq 1$ and $|\beta| \leq 1/2$. □

Lemma 8 *The Lagrange reduction algorithm terminates after* $25 + \max\left\{0, \log_2 \frac{\|\mathbf{b}_1\|_2}{\sqrt{\det L}}\right\}$ *iterations.*

Proof: Without loss of generality, we may assume that the determinant of the lattice is 1, by scaling. Note that this also means that $\|\mathbf{b}_1^\star\|_2 \|\mathbf{b}_2^\star\|_2 = 1$. We divide the algorithm into two phases; the phase where $\|\mathbf{b}_1\|_2^2 \geq 2$, and the phase where $\|\mathbf{b}_1\|_2^2 < 2$.

- (Phase 1) As $\mathbf{b}_1 = \mathbf{b}_1^\star$, and $\|\mathbf{b}_1\|_2^2 \geq 2$, we must have that $\|\mathbf{b}_2^\star\|_2^2 \leq 1/2 \leq 1/4\|\mathbf{b}_1\|_2^2$. Denote $\mathbf{c}_1$ as being the 'next iteration' $\mathbf{b}_1$. Note that the $\mathbf{c}_1 = \mathbf{b}_2 - k\mathbf{b}_1$ satisfies $|\langle \mathbf{c}_1, \mathbf{b}_1 \rangle| \leq 1/2\|\mathbf{b}_1\|_2^2$

and $\langle \mathbf{c}_1, \mathbf{b}_2^\star \rangle = \langle \mathbf{b}_2, \mathbf{b}_2^\star \rangle = \|\mathbf{b}_2^\star\|_2^2$. It follows that

$$\|\mathbf{c}_1\|_2^2 \leq \frac{1}{4}\|\mathbf{b}_1\|_2^2 + \|\mathbf{b}_2^\star\|_2^2 \leq \frac{1}{2}\|\mathbf{b}_1\|_2^2$$

This means that the square length of $\mathbf{b}_1$ reduces by a factor $1/2$ every iteration. Therefore, the number of iterations in phase 1 is at most $\log_2(\|\mathbf{b}_1\|_2)$.

- (Phase 2) In this phase, $\|\mathbf{b}_1\|_2^2 < 2$. We distinguish the cases $\lambda_2(L)^2 \geq 2$ and $\lambda_2(L)^2 < 2$.

  i) In the first case the algorithm is done, because $\|\mathbf{b}_1\|_2^2 < 2 \leq \lambda_2(L)^2$. Namely, this means that $\mathbf{b}_1$ is a multiple of a shortest vector in $L$. But, as $(\mathbf{b}_1, \mathbf{b}_2)$ is a basis of $L$, $\mathbf{b}_1$ must be a shortest vector itself.

  ii) In the second case, we invoke Lemma 5 to obtain $\lambda_1(L) > 1/\sqrt{2}$. We conclude by a packing argument: the set $P$ of points visited in this phase are in a ball of radius $R = \sqrt{2}$, and are separated by distance $\lambda_1(L)$. That is, the open balls of radius $r = \lambda_1(L)/2$ centered at each $p \in P$ are disjoints, furthermore they all are included in the centered ball of radius $R + r$. Therefore $|P| \leq (R+r)^2/r^2 = (1 + R/r)^2 \leq 5^2 = 25$. Because $\|\mathbf{b}_1\|$ is strictly decreasing at each iteration, each point in $P$ can only be visited once.

□

Definition 9 *Let $L$ be a full rank lattice $L$ of dimension $n$. Then we define $\gamma(L)$ of this lattice as*

$$\gamma(L) := \frac{\lambda_1(L)^2}{\det(L)^{2/n}}$$

Definition 10 (Hermite constant) *The hermite constant $\gamma_n$ is the supremum of $\gamma$ over $n$-dimensional full rank lattices:*

$$\gamma_n := \sup_L \gamma(L)$$

Lemma 11 *The densest sphere packing in dimension 2 is attained by the hexagonal lattice $H$, which achieves $\gamma(H) = \sqrt{4/3}$.*

Proof: Let $H = \mathcal{L}((0,1), (\sqrt{3/4}, 1/2))$. Verify that this indeed has the required Hermite constant. We now have to show that any 2-dimensional lattice has a Hermite constant at least $\gamma(H)$. Let $L$ be any lattice and let $(\mathbf{b}_1, \mathbf{b}_2)$ be a basis as in the Wristwatch lemma. Without loss of generality (see example sheet 4) we may assume $\mathbf{b}_1 = (0,1)$ and $\mathbf{b}_2 = (\alpha, \beta)$. Then $\lambda_1(L) = 1$, and $\det(L) = \det(\mathbf{B}) = \alpha$. As $|\beta| \leq 1/2$ we must have $1 \leq \alpha^2 + \beta^2 \leq \alpha^2 + \frac{1}{4}$. This directly implies $\alpha \geq \sqrt{3/4}$, which proves the claim. □

# 4 Hermite's Bound

Theorem 12 (Hermite) $\gamma_n \leq \gamma_2^{n-1}$.

One could reasonably ask why this theorem is stated here, as we already proved an asymptotically much stronger bound on $\gamma_n$ using Minkowski's convex body theorem. The first reason

why this theorem is mentioned is because of the historic context[2]; before Minkowski's theorem, Hermite's bound was the best known. The second and most important reason why this theorem is treated here, is because of its similarities with a famous basis reduction algorithm: the LLL algorithm. Some consider LLL as an 'algorithmization' of Hermite's bound.

We will soon prove Hermite's theorem 'by algorithm', see algorithm 2.

DEFINITION 13 *Let $L = L(\mathbf{B})$ a lattice generated by the basis $\mathbf{B}$. We will denote by $\mathbf{B}_{i:j}$ the basis $(\pi_i(\mathbf{b}_i), \ldots, \pi_i(\mathbf{b}_j))$, where $\pi_i(\cdot)$ is given as in Gram-Schmidt Orthogonalization, and by $\mathbf{B}^\star_{i:j}$ its Gram-Schmidt orthogonalization.*

Note that because $\pi_{i+a} \circ \pi_i = \pi_{i+a}$ for any $a \geq 0$, it holds that the $a+1$-th column vector of $\mathbf{B}^\star_{i:j}$ is exactly the $(i+a)$-th column vector of $\mathbf{B}^\star$. We can also write the following.

FACT 14 $\mathbf{B}^\star_{i:j} = (\mathbf{B}^\star_{i:k} | \mathbf{B}^\star_{k+1:j})$ *for any $i \leq k \leq j$.*

DEFINITION 15 *Let $L = \mathcal{L}(\mathbf{B})$ a lattice generated by the basis $\mathbf{B}$. We will denote $L_{i:j}$ for the lattice generated by $\mathbf{B}_{i:j} = (\pi_i(\mathbf{b}_i), \ldots, \pi_i(\mathbf{b}_j))$, i.e.,*

$$L_{i:j} = \mathcal{L}(\pi_i(\mathbf{b}_i), \ldots, \pi_i(\mathbf{b}_j)).$$

One easily check the following.

LEMMA 16 $\det(L_{i:j}) = \prod_{k=i}^{j} \|\mathbf{b}^\star_k\|_2$.

---

**Algorithm 2:** Hermite reduction algorithm

---

**Input :** A basis $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of a lattice $L$.
**Output:** A basis $\mathbf{B}$ such that $\|\mathbf{b}_1\|_2^2 \leq \gamma_2^{n-1} \cdot \det(L)^{2/n}$.

**while** $\exists i$ *such that* $\mathbf{B}_{i:i+1} = (\pi_i(\mathbf{b}_i), \pi_i(\mathbf{b}_{i+1}))$ *is not Lagrange reduced* **do**
  Find matrix $\mathbf{U} \in \mathbb{Z}^{2\times 2}$ such that $\mathbf{B}_{i:i+1}\mathbf{U}$ is Lagrange reduced
  $(\mathbf{b}'_i, \mathbf{b}'_{i+1}) \leftarrow (\mathbf{b}_i, \mathbf{b}_{i+1})\mathbf{U}$
  $\mathbf{B} \leftarrow (\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, \mathbf{b}'_i, \mathbf{b}'_{i+1}, \mathbf{b}_{i+2}, \ldots, \mathbf{b}_n)$
**end**
return $\mathbf{B}$

---

PROOF: (of Hermite's theorem)
First, we prove the correctness of Algorithm 2, assuming it terminates. After that, we will show that the algorithm does in fact terminate. Lastly, we see why the correctness and termination of the Hermite reduction algorithm implies Hermite's bound.

**Correctness** If the algorithm terminates, any pair of basis vectors $\mathbf{b}^\star_i, \mathbf{b}^\star_{i+1}$ satisfies $\|\mathbf{b}^\star_i\|_2 \leq \gamma_2 \|\mathbf{b}^\star_{i+1}\|_2$. Using inductive reasoning, one can conclude that $\|\mathbf{b}_1\|_2 = \|\mathbf{b}^\star_1\|_2 \leq \gamma_2^{i-1} \|\mathbf{b}^\star_i\|_2$. Multiplying together for all $i$ and using the determinant formula and the triangular number formula, yields

$$\|\mathbf{b}_1\|_2^n \leq \prod_{i=1}^{n} \gamma_2^{i-1} \|\mathbf{b}^\star_i\|_2 = \gamma_2^{\frac{n(n-1)}{2}} \det(L).$$

[2]Minkowski (1864-1909) established his convex body theorem in 1891; Hermite (1822-1904) stated his bound around 1850.

Taking $n/2$-th roots shows that the output of the Hermite reduction algorithm – if it indeed terminates – satisfies the requirements.

**Termination** Before proving termination of the algorithm, we would like to show that Lagrange-reduction on a pair of vectors $(\mathbf{b}_i, \mathbf{b}_{i+1})$ indeed leads to inequality $\|\mathbf{b}_i^\star\|_2 \leq \gamma_2 \|\mathbf{b}_{i+1}^\star\|_2$. After Lagrange-reduction, we have $\|\mathbf{b}_i^\star\|_2^2 \leq \gamma_2 \cdot \det(\mathbf{B}_{i:i+1}) = \gamma_2 \|\mathbf{b}_i^\star\|_2 \|\mathbf{b}_{i+1}^\star\|_2$, where the last equality comes from Lemma 16. Dividing out appropriately yields $\|\mathbf{b}_i^\star\|_2 \leq \gamma_2 \|\mathbf{b}_{i+1}^\star\|_2$. So Lagrange-reduction indeed 'resolves' the wrong-way inequality $\|\mathbf{b}_i^\star\|_2 > \gamma_2 \|\mathbf{b}_{i+1}^\star\|_2$.

To prove that the algorithm terminates one can use an induction argument. Let us assume, by hypothesis, that the Hermite reduction algorithm always terminates on lattices with dimension smaller than $n$. We will prove that this algorithm also terminates on lattices with dimension precisely $n$.

To show that, we need a few claims.

- The norm of $\mathbf{b}_1$ doesn't change if a Lagrange reduction doesn't involve $\mathbf{b}_1$. This is obviously true, because then $\mathbf{b}_1$ is not affected and the norm stays the same.
- When a Lagrange reduction *does* involve $\mathbf{b}_1$, it will replace $\mathbf{b}_1$ with a new one with strictly smaller norm. This is true by the following reasoning. Because the Hermite reduction algorithm only applies Lagrange reduction whenever $\|\mathbf{b}_1\|_2 > \gamma_2 \|\mathbf{b}_2^\star\|_2$, we know that $\|\mathbf{b}_1\|_2^2 > \gamma_2 \|\mathbf{b}_2^\star\|_2 \|\mathbf{b}_1\|_2 = \gamma_2 \cdot \det(L_{1:2})$ *before* the Lagrange reduction happens. However, after Lagrange reduction, we have $\|\mathbf{b}_1\|_2 \leq \gamma_2 \cdot \det(L_{1:2})$. So the new $\mathbf{b}_1$ has indeed a strictly smaller norm.
- $\|\mathbf{b}_1\|_2$ has a lower bound. Namely, $\lambda_1(L) \leq \|\mathbf{b}_1\|_2$.

Since $L$ is a discrete subset of $\mathbb{R}^n$, the norm of $\mathbf{b}_1$ has a lower bound and the norm is decreasing during the algorithm, this norm must eventually stabilize. This means that no Lagrange-reduction involving $\mathbf{b}_1$ happens after that stabilization.

Therefore, after this stabilization, the algorithm takes place in $L_{2:n}$ alone (because no operations on $\mathbf{b}_1$ are done anymore). By the induction hypothesis, it must terminate. Thus, the entire algorithm on $L$ terminates, too.

**Implication of Hermite's bound** Lastly, the correctness and termination of this algorithm implies Hermite's bound; in any lattice of dimension $n$ we can find a vector $\mathbf{b}_1$ whose square norm is bounded by $\gamma_2^{n-1} \cdot \det(L)^{2/n}$. Thus,

$$\gamma(L) = \frac{\lambda_1(L)^2}{\det(L)^{2/n}} \leq \frac{\|\mathbf{b}_1\|_2^2}{\det(L)^{2/n}} \leq \frac{\gamma_2^{n-1} \cdot \det(L)^{2/n}}{\det(L)^{2/n}} = \gamma_2^{n-1}.$$

Therefore, $\gamma_n = \sup_L \gamma(L) \leq \gamma_2^{n-1}$. □