Counting rational points of a Grassmannian

Seungki Kim

Received: 8 May 2022 / Accepted: 21 September 2022 / Published online: 5 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract
We prove an estimate on the number of rational points on the Grassmannian variety of bounded twisted height, refining the classical results of Schmidt (Duke Math J 35:327–339, 1968) and Thunder (Compos Math 88(2):155–186, 1993) over the rational field: most importantly, our formula counts all points. Among the consequences are a couple of new implications on the classical subject of counting rational points on flag varieties.

Keywords Rational points · Grassmannians · Flag varieties · Manin’s conjecture

Mathematics Subject Classification 11H06 · 11G50 · 14G05

1 Introduction

1.1 Main result

For $L \subseteq \mathbb{R}^n$ a lattice, $1 \leq d < \text{rk} L$ and $H > 0$, let $P(L, d, H)$ be the number of primitive rank $d$ sublattices of $L$ of determinant less than or equal to $H$. The purpose of this paper is to investigate the quantitative behavior of $P(L, d, H)$. The earliest result of this kind goes back to the mid-twentieth century, due to Schmidt [12]:

Theorem 1.1 (Schmidt [12], Theorem 1) Let

$$a(n, d) = \frac{1}{n} \binom{n}{d} \prod_{i=1}^{d} \frac{V(n - i + 1)}{V(i)} \cdot \frac{\xi(i)}{\xi(n - i + 1)},$$

$$b(n, d) = \max \left( \frac{1}{d}, \frac{1}{n - d} \right),$$
where $V(i) := \pi^{i/2}/\Gamma(i/2 + 1)$ is the volume of the unit ball in $\mathbb{R}^i$ and $\zeta(s)$ is the Riemann zeta function, except that we understand $\zeta(1) = 1$ for convenience. Then
\[
P(\mathbb{Z}^n, d, H) = a(n, d) H^n + O(H^{n-b(n,d)}),
\]
where the implicit constant depends on $n$ only.

For $\mathcal{L} \subseteq \mathbb{R}^n$ of full rank, $P(\mathcal{L}, d, H)$ may also be understood in terms of a counting problem on the Grassmannian variety $\text{Gr}(n, d)$ consisting of the $d$-dimensional subspaces of $\mathbb{R}^n$. A rational point $P$ on $\text{Gr}(n, d)$ is a $d$-dimensional subspace such that $P \cap \mathbb{Z}^n$ is a rank $d$ sublattice of $\mathbb{Z}^n$. Its height is given by the determinant of $P \cap \mathbb{Z}^n$, and given a real $n \times n$ matrix $L$, its height twisted by $L$ is given by the determinant of the rank $d$ lattice $(P \cap \mathbb{Z}^n)L = \{vL : v \in P \cap \mathbb{Z}^n\} \subseteq \mathbb{R}^n$

(we write vectors horizontally, so that matrices multiply from the right). Observe that, if $L$ and $L'$ are two $n \times n$ matrices whose row vectors generate the same lattice $\mathcal{L}$, i.e. $L = gL'$ for some $g \in \text{GL}(n, \mathbb{Z})$, then the number of the rational points whose heights are bounded by $H$ is the same, regardless of whether one twists the height by $L$ or $L'$. In addition, one checks that this number is precisely $P(\mathcal{L}, d, H)$ as defined above. We refer the reader to Thunder ([19, Introduction], [20, Part I]) for the general definition of a twisted height, in which it is first introduced.

From this perspective, Thunder [20] proved a vast generalization of Theorem 1.1 above, extending it to any lattice and to any number field $K$ (here $O_K$-modules play the role of lattices). His result, from the 1990’s, remains state-of-the-art to this day. We state his result in case $K = \mathbb{Q}$:

**Theorem 1.2** (Thunder [20], Theorem 3) Let $\mathcal{L} \subseteq \mathbb{R}^n$ be a lattice of full rank. In addition to the notations in Theorem 1.1, define
\[
\mathcal{L}_i = \mathcal{L} \cap \text{span}_\mathbb{R}(v_1, \ldots, v_i),
\]
where $v_1, \ldots, v_n$ are choices of linearly independent vectors in $\mathcal{L}$ such that $\|v_i\| = \lambda_i(\mathcal{L})$, where $\lambda_i(\mathcal{L})$ is the $i$-th successive minimum of $\mathcal{L}$ defined by
\[
\lambda_i(\mathcal{L}) = \inf\{r > 0 : \text{dim span}_\mathbb{R}(\mathcal{L} \cap B(r)) \geq i\},
\]
where $B(r)$ here is the closed ball at origin of radius $r$ in $\mathbb{R}^n$. Let $P_{\mathcal{L}_{n-d}}(\mathcal{L}, d, H)$ be the number of rank $d$ sublattices of $\mathcal{L}$ of determinant $\leq H$ whose intersection with $\mathcal{L}_{n-d}$ is trivial. Then
\[
P_{\mathcal{L}_{n-d}}(\mathcal{L}, d, H) = a(n, d) \frac{H^n}{(\det \mathcal{L})^d} + O\left(\frac{H^{n-b(n,d)}}{(\det \mathcal{L})^{d-b(n,d)}(\det \mathcal{L}_{n-d})^{b(n,d)}}\right),
\]
where the implicit constant depends only on $n$. 

Springer
A notable feature of Theorem 1.2 is that it provides an explicit description of the dependence of the error term on the successive minima $\lambda_1(\mathcal{L}), \ldots, \lambda_n(\mathcal{L})$ of $\mathcal{L}$ (observe that $\det \mathcal{L}_i \sim \lambda_1 \lambda_2 \ldots \lambda_i$ by the Minkowski’s second theorem). Informally speaking, it reflects the “skewness” of the lattice: in case $\mathcal{L}$ is severely skewed, in the sense that $\lambda_i(\mathcal{L})$ is much smaller than $\lambda_{i+1}(\mathcal{L})$ for some $i$, one expects a different behavior of the error term than the case in which most $\lambda_i(\mathcal{L})$ are about equal. Theorem 1.2 may be seen as a realization of this intuition.

However, Thunder [20] does not provide a corresponding estimate for $P(\mathcal{L}, d, H)$, remarking that it would “be a cumbersome task.” In the present paper, we introduce a method that circumvents this difficulty, and prove

**Theorem 1.3’** Continue with the notations in Theorems 1.1 and 1.2 above. Then for all $H > 0$,

$$P(\mathcal{L}, d, H) = a(n, d) \frac{H^n}{(\det \mathcal{L})^d} + O \left( 1 + \left( \frac{H}{\lambda_1(\mathcal{L})^d} \right)^{n-b(n,d)} \right), \quad (1.1)$$

where the implied constant depends only on $n$.

In fact, we prove the more precise

**Theorem 1.3** Let $\mathcal{L} \subseteq \mathbb{R}^n$ be a lattice of full rank. For all $H > 0$,

$$P(\mathcal{L}, d, H) = a(n, d) \frac{H^n}{(\det \mathcal{L})^d} + O \left( \sum_{j \in E_{n,d}} b_j(\mathcal{L}) H^{\gamma_j} \right), \quad (1.2)$$

where the implied constant in the big-O notation depends only on $n$, and $E_{n,d}$ is a finite set of indices, of cardinality less than $n^{3n}$ for $n \geq 2$. Each $j \in E_{n,d}$ is associated with $\gamma_j \in [0, n - b(n, d)]$ and $b_j$ of the form

$$b_j(\mathcal{L}) = \prod_{i=1}^{n} (\det \mathcal{L}_i)^{-\beta_i}$$

for some real $\beta_i \geq 0$ such that $\sum_{i=1}^{n} i \beta_i = d \gamma_j$. This makes the right-hand side of (1.2) scale-invariant i.e. it remains unchanged if $\mathcal{L}$ and $H$ are replaced by $c\mathcal{L}$ and $c^d H$, respectively.

In particular, the leading error term is

$$\frac{H^{n-b(n,d)}}{(\det \mathcal{L})^{d-b(n,d)}(\det \mathcal{L}_{n-d})^{b(n,d)}},$$

as in Theorem 1.2.

**Remark** If $\mathcal{L}$ is of rank $m < n$, we may adapt Theorem 1.3 by observing that, for any isometry $\iota : \mathbb{R}^m \rightarrow \text{span}_{\mathbb{R}}(\mathcal{L})$, it holds that $P(\mathcal{L}, d, H) = P(\iota^{-1}(\mathcal{L}), d, H)$. The same applies to the results of the similar flavor that we state below.
Remark The estimate $|E_{n,d}| < n^{3n}$ is an extremely crude one, provided only to assure that the sum is finite. Describing $E_{n,d}$ exactly from our computations would be quite a laborious task that would not yield a pretty formula and whose fruits seem unclear.

Thus one may feel that the complicated statement of Theorem 1.3 is unnecessary. However, we leave it as it is, since the precise knowledge of at least a part of the error term may be useful for certain applications. For instance, to prove Corollary 1.1 below, we really need the (close-to-)optimal version of the leading error term stated as in Theorem 1.3. If we could compute the coefficients $b_j$’s optimally for more $j$’s, we expect to be able to strengthen Corollary 1.1 accordingly.

We also state the subsequent results in the precise form. If one desires simplification, one may replace $b_j$ by an appropriate power of $\lambda_1$’s, as in (1.1).

Before we go on to discuss a few applications of Theorem 1.3, let us present a few of its variants that may also be of use.

**Theorem 1.4** Let $N(\mathcal{L}, d, H)$ be the number of (not necessarily primitive) rank $d$ sublattices of $\mathcal{L}$ of determinant $\leq H$. Also let

$$c(n, d) = a(n, d) \prod_{i=1}^{d} \zeta(n - i + 1).$$

Then similarly to Theorem 1.3, for a full-rank lattice $\mathcal{L} \subseteq \mathbb{R}^n$ we have

$$N(\mathcal{L}, d, H) = c(n, d) \frac{H^n}{(\det \mathcal{L})^d} + O \left( \sum_{j \in F_{n,d}} b_j(\mathcal{L}) H^{\gamma_j} \right),$$

where the implied constant in the big-$O$ notation depends only on $n$, and $F_{n,d}$ is a set of indices of cardinality at most $n^{3n}$. The description of $b_j$ (resp. $\gamma_j$) is the same as that of $b_j$ (resp. $\gamma_j$) in Theorem 1.3. Also, if $d < n - 1$, the leading error term is the same as in Theorem 1.3. If $d = n - 1$, then the largest $\gamma_j$ for $j \in F_{n,d}$ may be chosen to be $n - 1 + \eta$, for any $\eta > 0$.

**Theorem 1.5** For a lattice $\mathcal{L} \subseteq \mathbb{R}^n$, choose a primitive sublattice $S \subseteq \mathcal{L}$ of rank $\leq n - d$, and let $P_S(\mathcal{L}, d, H)$ be the number of primitive rank $d$ sublattices of $\mathcal{L}$ whose intersection with $S$ is trivial. For $\mathcal{L} \subseteq \mathbb{R}^n$ of full rank, we have

$$P_S(\mathcal{L}, d, H) = a(n, d) \frac{H^n}{(\det \mathcal{L})^d} + O \left( \sum_{j \in G_{n,d}} b_j(\mathcal{L}) H^{\gamma_j} \right),$$

where again the description of the error term is identical to that of Theorem 1.3, except that $G_{n,d}$ is now a set of cardinality less than $3n^{3n}$. In particular, the leading error term is the same as in Theorem 1.3. Moreover, this formula is independent of $S$.

The analogous statement holds for $N_S(\mathcal{L}, d, H)$, which counts both primitive and non-primitive lattices.
1.2 Applications

Below we demonstrate a few immediate applications of Theorem 1.3 and the techniques used in its proof. Its main strength lies in the fact that it counts all the sublattices, and that it provides information regardless of how skewed the given lattice is, in particular relative to \( H \). In comparison, its precedent Theorem 1.2 misses the sublattices that nontrivially intersect \( L_{n-d} \), and thus it does not say anything about the lattices for which \( \det L / \det L_{n-d} > H \).

We expect there to be more uses of Theorem 1.3; for instance, see a recent work of Le Boudec ([7]), which employs the \( d = 1 \) case (due to Schmidt ([12]), see (2.1) below) as the main device.

1.2.1 Rational points of flag varieties

It is natural to expect that a counting formula on Grassmannians should yield a counting formula for general flag varieties. Indeed, Thunder [20] derives such a formula as a relatively simple application of Theorem 1.2. We present its simplest case to initiate the discussion:

**Theorem 1.6** (Thunder [20], Theorem 5) Let \( L \subseteq \mathbb{R}^n \) be a lattice of rank \( n \), and suppose \( H \) is sufficiently large. Then the number of flags \( S^e \subseteq S^d \subseteq L \) of type \( (e, d) \) (hence \( \text{rk } S^i = i \) for \( i \in \{e, d\} \), and \( S^e \cap S_{d-e}^d = 0 \), whose height twisted by \( L \) is at most \( H \) is

\[
\frac{aH}{(\det L)^d} \log \frac{H}{(\det L)^d} + O \left( \frac{H}{(\det L)^{d-b(n,d)(n-d)/n} (\det L_{n-d})^{b(n,d)}} \right),
\]

(1.3)

where \( a \) is some explicit constant depending only on \( n, d, e \), and the implicit constant in the error depends only on \( n \).

In this context, the height of the flag \( S^e \subseteq S^d \) is the quantity \( (\det S^e)^d (\det S^d)^{n-e} \); see e.g. Thunder [20] for details.

In comparison, we can derive from Theorem 1.3 the following

**Corollary 1.1** Let \( L \) be a lattice of rank \( n \geq 3 \), and \( H \) be sufficiently large—more precisely, \( \lambda_n(L) \ll_n H^{1/nd} \). Then the number of flags \( S^e \subseteq S^d \subseteq L \) of type \( (e, d) \) such that \( S^e \cap S_{d-e}^d = 0 \) whose height twisted by \( L \) is at most \( H \) is

\[
\frac{aH}{(\det L)^d} \log \frac{H}{\varepsilon_{e}^d \varepsilon_{d}^{n-e}} + O \left( \sum_{j \in E_{n,e,d}} b_j(L) H^{\gamma_j} \right),
\]

(1.4)

where \( \varepsilon_e = \min \{ \det \chi : \chi \subseteq L, \text{rk } \chi = e \} \) and likewise for \( \varepsilon_d \), \( a \) is the same as in (1.3), the implicit constant depends on \( n \) only, \( E_{n,e,d} \) is an index set of cardinality at most \( 3n^{3n+1} \), and \( b_j(L) \) and \( \gamma_j \) are as in the statement of Theorem 1.3, except that \( \sum \beta_i = dn \gamma_j \).
Furthermore, the largest $\gamma_j$ is 1, and the next largest is either

$$1 - \frac{b(n, d)}{n}, \quad 1 - \frac{b(d, e)(n - e)}{nd}, \quad \text{or} \quad 1 - \frac{1}{n} \left( 1 - \frac{2b(d, e)}{d} + \frac{1 - b(d, e)}{n - e} \right);$$

if $d \leq n/2$, it is always $1 - b(n, d)/n = 1 - 1/dn$.

In order to keep the proof relatively short and simple, we had to keep some of the assumptions made by Theorem 1.6. Still, it has a couple of new implications that may be of interest. First, it shows that there must exist a gap between Theorem 1.6 and an ideal counting formula that would count all the rational flags, and that it must be at least of size

$$O \left( H \log \frac{(\det \mathcal{L})^d}{e_d e^{n-e}_d} \right).$$

For a heavily skewed $\mathcal{L}$, for instance if $\det \mathcal{L} = 1$ but $e_d \ll H^{-1/n}$, then this is of comparable size to the main term.

The second implication has to do with the error term in the well-known theorem of Franke, Manin, and Tschinkel [4] on the number of rational points on flag varieties. In the corollary to Theorem 5 therein, which says that the number of rational points on a flag variety $V$ of (“untwisted”) height bounded by $H$ is

$$Hp(\log H) + o(H),$$

where $p$ is a polynomial of degree $\text{rk} \text{Pic}(V) - 1$, they conjecture that the error term is of size $O(H^{1-\varepsilon})$ with $\varepsilon = (\dim V)^{-1}$. On the other hand, when $V$ is a Grassmannian, the literature (for instance Schmidt [12], and Thunder [19, 20]) suggests that $\varepsilon = b(n, d)/n$, as their analyses seem fairly sharp. Our Corollary 1.1 extends this to flag varieties of type $(e, d)$, suggesting that we have $\varepsilon = b(n, d)/n$ again, at least when $d \leq n/2$. In case $e \geq n/2$, one may be able to estimate $\varepsilon = 1/(n - e)n$ by duality. But in general the nature of $\varepsilon$ appears rather complicated.

It is possible to prove by a similar argument an analogue of Corollary 1.1 for a flag variety of any type that is strong enough to yield these same implications. On the other hand, we expect the ideal formula that would count all points on a flag variety to require another substantial amount of effort along the lines of the present paper. As stated in the remark after Theorem 1.3, if we could find explicit expressions for every $b_j$ in (1.2), preferably containing large powers of $\det \mathcal{L}$, it would allow our error-bounding techniques to apply immediately. Using the methods of the present paper, it may be possible to achieve this for the first few small values of $d$.

### 1.2.2 Mean value theorems over lattices

Mean value theorems over random lattices provide a method of averaging lattice-point counting formulas over the space $X_n := \text{SL}(n, \mathbb{Z}) \backslash \text{SL}(n, \mathbb{R})$ of determinant 1 full-
rank lattices in $\mathbb{R}^n$. The first known such theorem is the famous Siegel integration formula:

**Theorem 1.7** (Siegel [17], Theorem on p.341) Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a Borel measurable and compactly supported function. Then

$$
\int_{X_n} \sum_{v \in \mathcal{L} \setminus \{0\}} f(v) d\mu_n(\mathcal{L}) = \int_{\mathbb{R}^n} f(x) dx,
$$

(1.5)

where $d\mu_n$ is the normalized Haar measure on $SL(n, \mathbb{R})$, and $dx$ is the Lebesgue measure.

For example, if one sets $f$ to be the characteristic function of a set $S \subseteq \mathbb{R}^n$, then the sum on the left-hand side of (1.5) equals $|S \cap \mathcal{L} \setminus \{0\}|$, and thus Theorem 1.7 implies that a random lattice sampled according to $d\mu_n$ has on average $\text{vol}(S)$ nonzero vectors contained in $S$. Another important example is the Rogers integral formula, one of the main tools in geometry of numbers (see e.g. [5, 15, 18] for some of the applications):

**Theorem 1.8** (Rogers [9], (essentially) Theorem 4) For $k < n$, let $f : (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ be a Borel measurable and compactly supported function. Then

$$
\int_{X_n} \sum_{\substack{v_1, \ldots, v_k \in \mathcal{L} \text{ independent}}} f(v_1, \ldots, v_k) d\mu_n(\mathcal{L}) = \int_{\mathbb{R}^n} \ldots \int_{\mathbb{R}^n} f(x_1, \ldots, x_k) dx_1 \ldots dx_k,
$$

where $d\mu_n$ is the normalized Haar measure on $SL(n, \mathbb{R})$, and each $dx_i$ is the Lebesgue measure.

In the author’s work [6], written concurrently with the present paper, the machinery that turns a lattice-point counting formula, such as Theorems 1.3 and 1.5, into a mean-value theorem has been developed, inspired by the argument of Rogers [9]. As a result, the following extension of Theorem 1.8 to Grassmannians is obtained from Theorem 1.5

**Theorem 1.9** (Kim [6], Theorem 3) Suppose $1 \leq k < n$, $1 \leq d_1, \ldots, d_k < n$ and $d_1 + \ldots + d_k < n$. Define $P(\mathcal{L}, d_1, \ldots, d_k, H_1, \ldots, H_k)$ to be the number of independent primitive sublattices $A_1, \ldots, A_k$ of $\mathcal{L}$ of ranks $d_1, \ldots, d_k$ and determinants bounded by $H_1, \ldots, H_k$, respectively. Then

$$
\int_{X_n} P(\mathcal{L}, d_1, \ldots, d_k, H_1, \ldots, H_k) d\mu_n(\mathcal{L}) = \prod_{i=1}^k a(n, d_i) H_i^n.
$$

We note that Thunder [21] proved the $k = 1$ case of this result. Theorem 1.9 has several implications on the statistics of the randomized heights of the points on Grassmannians and flag varieties, such as the following.
Corollary 1.2 (Kim [6], Corollary to Theorem 5) Let \( k \geq 2 \), and \( 1 = d_0 \leq d_1 < \ldots < d_k < d_{k+1} = n \). For a lattice \( \mathcal{L} \subseteq \mathbb{R}^n \) and its flag \( S^{d_1} \subseteq \ldots \subseteq S^{d_k} \subseteq \mathcal{L} \) of type \( \mathcal{D} = (d_1, \ldots, d_k) \), its height is defined as the quantity

\[
\prod_{i=1}^{k} (\det S^{d_i})^{d_{i+1} - d_i - 1}.
\]

Let \( P(\mathcal{L}, \mathcal{D}, H) \) be the number of type \( \mathcal{D} \) flags of \( \mathcal{L} \) whose heights are bounded by \( H \). Then

\[
\int_{X_n} P(\mathcal{L}, \mathcal{D}, H) d\mu_n(\mathcal{L}) = \infty.
\]

It may be interesting to compare this result with Theorem 1.6 and Corollary 1.1. The author speculates that the divergence here is related to the main term of \((1.4)\) in the statement of Corollary 1.1 being dependent on the skewness of the lattice in question.

1.3 Method of proof

All previous works on this topic [12, 19, 20] count “upwards,” i.e. they construct the \( d \)-dimensional sublattice from either a \((d-1)\)-dimensional sublattice or a \( d \)-dimensional sublattice lying in an \((n-1)\)-dimensional ambient space. Our main idea is to take the dual approach, and count “downwards” instead: we project all the \( d \)-dimensional sublattices to a hyperplane, and count the cardinality of each fiber. This lets us bypass some of the technical difficulties that arise when counting upwards.

To elaborate, we prove Theorem 1.3 by the following inductive procedure that resembles the Pascal’s triangle method of computing the binomial coefficients. In case \( d = 1 \) or \( d = n - 1 \), the formulas are well-known. Otherwise, let \( \tilde{\mathcal{L}} \) be the projection of \( \mathcal{L} \) onto the orthogonal complement of a shortest nonzero vector of \( \mathcal{L} \). Then we have

\[
P(\mathcal{L}, d, H) = P(\tilde{\mathcal{L}}, d - 1, \frac{H}{\lambda_1(\mathcal{L})}) + \Phi(P(\tilde{\mathcal{L}}, d, H)), \tag{1.6}
\]

where \( \Phi \) can be regarded as a certain integral transformation. For a choice of a basis \( \{v_1, \ldots, v_n\} \) of \( \mathcal{L} \) and a sublattice \( \mathcal{B} \subseteq \mathcal{L} \) of rank \( d \), let us say \( \mathcal{B} \) is of \( d \)-type \((\alpha_1, \ldots, \alpha_n)\) —“d” stands for “dual”—if the projection of \( \mathcal{B} \) onto \( \text{span}_{\mathbb{R}}(v_1, \ldots, v_{n-i+1})^\perp \) has rank \( \alpha_i \). Then the first term on the right-hand side of (1.6) is counting the sublattices of \( d \)-types \((\ast, \ldots, \ast, d - 1, d)\), and the second term is counting those of \( d \)-types \((\ast, \ldots, \ast, d, d)\).

In comparison, Theorem 1.2 counts precisely the sublattices of \( d \)-type \((1, 2, \ldots, d_1, \ldots, d)\). The upward counting method forces one to count the sublattices of each \( d \)-type separately, which is precisely what Thunder refers to as being “cumbersome.” The downward method resolves this difficulty.

Most of this paper is devoted to explicitly writing out and estimating \( \Phi(P(\tilde{\mathcal{L}}, d, H)) \). Many parts of the computation can be done by slightly refining the methods of Schmidt.
However, the fact that $\mathcal{L}$ can be arbitrarily skewed presents a new difficulty, especially when bounding the error terms. This is resolved by comparing the gaps between the successive minima to $H^{1/d}$: if $\lambda_{i+1} - \lambda_i \ll H^{1/d}$ for all $i = 1, \ldots, n-1$, the lattice may be considered not so severely skewed, as the classical techniques continue to apply. If in contrast $\lambda_{i+1} - \lambda_i \gg H^{1/d}$ for some $i$, we exploit this gap to finesse the desired error bound.

1.4 Organization

In Sect. 2, we introduce the definitions and notations used throughout the paper, and state the known formulas for $P(\mathcal{L}, 1, H)$ and $P(\mathcal{L}, n-1, H)$. In Sect. 3, we set up the induction argument, establishing the precise version of (1.6). Sections 4 and 5 are devoted to the main and error term estimates, respectively. Section 6 collects all the computations and concludes the proof of Theorem 1.3. The variants are all proved in Sect. 7.

2 Some backgrounds

2.1 Definitions, notations, and conventions

Unless mentioned otherwise, the definitions and notations of this section apply.

Generalities

The lowercase letter $p$ denotes a prime. Let us write $\Gamma = \text{GL}(d, \mathbb{Z})$ for short. We use capital letters such as $L, M$ to refer to matrices, and calligraphic fonts such as $\mathcal{L}, \mathcal{M}$ to denote lattices. $n \in \mathbb{Z}_{> 0}$ and $d \in \{1, \ldots, n-1\}$ are fixed integers throughout the paper.

As in the statement of Theorem 1.2, $\lambda_i(\mathcal{L})$ is the $i$-th successive minimum of $\mathcal{L}$, and $\mathcal{L}_i$ denotes (a choice of) a primitive $i$-dimensional sublattice containing $v_1, \ldots, v_i \in \mathcal{L}$, which are linearly independent with $\|v_i\| = \lambda_i(\mathcal{L})$. The $(i, j)$-entry of a matrix is denoted by the lowercase letter of the name of the matrix indexed by $ij$. For example, if $A$ is a $d \times n$ matrix, then $A = (a_{ij})_{1 \leq i \leq d}$. Similarly, if $x \in \mathbb{R}^n$, then the $i$-th entry of $x$ is denoted by $x_i$.

Later, given a $d \times (n-1)$ matrix $A$ and a $d \times 1$ vector $v$, we will need to consider the $d \times n$ matrix $B$ whose $i$-th row equals $(a_{i1}, \ldots, a_{i,n-1}, v_i)$. We write $B = (A; v)$ to describe such a matrix.

For two quantities $f$ and $g$, $f \ll g$ means $f < Cg$, where $C$ is a positive constant possibly depending on $d$ and $n$ but no other variables. $f \sim g$ means $f \ll g$ and $g \ll f$. For example, Minkowski’s second theorem says that $\det \mathcal{L} \sim \prod \lambda_i(\mathcal{L})$.

For two matrices $A$ and $B$ with $d$ rows, $A \sim B$ means they differ by the left multiplication by an element of $\Gamma$. If the rows of each of $A$ and $B$ respectively span $A$ and $B$ in $\text{Gr}(\mathcal{M}, d)$ (whose precise definition is given below), $A \sim B$ means that $A = B$. 

Springer
Later in the paper, we will need a few facts from reduction theory. Let \( \{v_1, \ldots, v_m\} \) be a basis of a lattice \( \mathcal{M} \), and \( \{v_1^*, \ldots, v_m^*\} \) be its Gram-Schmidt orthogonalization, that is, each \( v_i^* \) is the projection of \( v_i \) to the orthogonal complement of \( \text{span}(v_1, \ldots, v_{i-1}) \). Let us say the basis is \textit{reduced} if each \( v_i \sim \lambda_i(\mathcal{M}) \) and \( \langle v_i^*, v_j \rangle / \|v_i^*\|^2 \leq 1/2 \) for all \( i < j \). It is known by reduction theory (see e.g. [2, Chapter 1]) that any lattice has a reduced basis. Moreover, the LLL algorithm [8] outputs a reduced basis of any lattice, given any basis of that lattice.

\textbf{Gr}(\mathcal{M}, d)\ and\ the\ determinant/height

A \( d \times n \) integral matrix \( X \in \text{Mat}_{d \times n}(\mathbb{Z}) \) is said to be \textit{primitive} if \( X \) can be completed to an element of \( \text{GL}(n, \mathbb{Z}) \). When \( d = 1 \), this agrees with the standard definition of a primitive vector. We denote the set of all primitive \( d \times n \) matrices by \( \text{Mat}^{\text{pr}}_{d \times n}(\mathbb{Z}) \).

For a lattice \( \mathcal{M} \subseteq \mathbb{R}^n \) of rank \( m \leq n \), a sublattice \( \mathcal{A} \subseteq \mathcal{M} \) is said to be \textit{primitive} if \( \text{span}(\mathcal{A}) \cap \mathcal{M} = \mathcal{A} \). We denote \( \text{Gr}(\mathcal{M}, d) \) for the set of all rank \( d \) primitive sublattices of \( \mathcal{M} \) inside \( \mathbb{R}^n \). Choose a basis \( \{v_1, \ldots, v_m\} \) of \( \mathcal{M} \), and a basis \( \{w_1, \ldots, w_d\} \) of \( \mathcal{A} \in \text{Gr}(\mathcal{M}, d) \). Let \( M \) and \( A \) respectively denote the \( m \times n \) and \( d \times m \) matrices, such that the \( i \)-th row of \( M \) is \( v_i \), and the \( i \)-th row of \( AM \) is \( w_i \). One checks that \( A \in \text{Mat}^{\text{pr}}_{d \times m}(\mathbb{Z}) \) by the fact that \( A \in \text{Gr}(\mathcal{M}, d) \).

Suppose one chooses a different basis \( \{w'_1, \ldots, w'_d\} \) of \( \mathcal{A} \), and let \( A' \in \text{Mat}^{\text{pr}}_{d \times m}(\mathbb{Z}) \) be the matrix such that the \( i \)-th row of \( A'M \) is \( w'_i \). Then \( A = gA' \) for some \( g \in \Gamma \). Conversely, if \( A = gA' \) for some \( g \in \Gamma \) and \( A' \in \text{Mat}^{\text{pr}}_{d \times m}(\mathbb{Z}) \), then the rows of \( AM \) and \( A'M \) span the same element of \( \text{Gr}(\mathcal{M}, d) \). Therefore, with a choice of \( M \in \text{Mat}_{m \times n}(\mathbb{R}) \) whose rows span \( \mathcal{M} \) over \( \mathbb{Z} \), there exists a bijection between \( \text{Gr}(\mathcal{M}, d) \) and the set of orbits \( \Gamma \backslash (\text{Mat}^{\text{pr}}_{d \times m}(\mathbb{Z}) \cdot M) \) of the action of \( \Gamma \) on \( \text{Mat}^{\text{pr}}_{d \times m}(\mathbb{Z}) \cdot M \) by left multiplication.

To make this even more explicit, recall that each element of \( \Gamma \backslash \text{Mat}^{\text{pr}}_{d \times m}(\mathbb{Z}) \) is uniquely represented by a primitive Hermite normal form over \( \mathbb{Z} \). Thus there is also a bijection between \( \text{Gr}(\mathcal{M}, d) \) and the set of all elements of form \( AM \) where \( A \) is a \( d \times m \) primitive Hermite normal form over \( \mathbb{Z} \). Whenever convenient, we will use these identifications of \( \text{Gr}(\mathcal{M}, d) \) interchangeably throughout the paper.

In order to simplify some notations, we adopt the unusual convention that all determinants are nonnegative (the groups \( \text{GL}(n, \cdot) \) and \( \text{SL}(n, \cdot) \) maintain their usual meanings, though). Specifically, for a square matrix \( X \), we write \( \det X \) for the absolute value of its usual definition of determinant. For a non-square matrix \( X \), we define \( \det X = \sqrt{\det XX^\top} \). For a lattice \( \mathcal{A} \subseteq \mathbb{R}^n \), we define \( \det \mathcal{A} \) to be its covolume within its \( \mathbb{R} \)-span. For \( \mathcal{A} \in \text{Gr}(\mathcal{M}, d) \), note that \( \det \mathcal{A} = \det AM \) holds, where \( M \) is any choice of a matrix whose row vectors form a basis of \( \mathcal{M} \), and \( A \) is any choice of an element of \( \text{Mat}^{\text{pr}}_{d \times m}(\mathbb{Z}) \) such that the row vectors of \( AM \) form a basis of \( \mathcal{A} \).

For a matrix \( M \), we define

\[
 f_H(M) = \begin{cases} 
 1 & \text{if } \det M \leq H \\
 0 & \text{otherwise}, 
\end{cases}
\]

and similarly for a lattice \( \mathcal{M} \).
It is easy to see that, for any compactly supported function $F$ defined on a subset of $\mathbb{R}$, we have

$$
\sum_{A \in \text{Gr}(M, d)} F(\det A) = \sum_{B \in \text{Gr}(\mathbb{Z}^m, d)} F(\det BM) = \sum_{B \in \Gamma \setminus \text{Mat}_{d \times m}(\mathbb{Z})} F(\det BM),
$$

where $BM$ is understood as the image of $B$ by the linear map $\mathbb{R}^n \to \mathbb{R}^n$ induced by the matrix $M$. Again we will switch freely between these notations as we see fit.

**Orthogonality notions**

Following Schmidt [12], we define the polar lattice $M^P$ of $M$ by $M^P = \{ w \in \text{span}_{\mathbb{R}}(M) : \langle v, w \rangle \in \mathbb{Z}, \forall v \in M \}$. If $S \in \text{Gr}(M, d)$, we define its orthogonal lattice $S^\perp \in \text{Gr}(M^P, m - d)$ by $S^\perp = \{ w \in M^P : \langle v, w \rangle = 0, \forall v \in S \}$.

In addition, for a $m \times n$ matrix $M$ whose $i$-th row vector is denoted by $v_i$, we define its polar matrix $M^P$ as the $m \times n$ matrix whose $j$-th row vector $v^P_j$ lies in $\text{span}_{\mathbb{R}}(v_1, \ldots, v_m)$ and satisfies $\langle v_i, v^P_j \rangle = \delta_{ij}$. Then the rows of $M^P$ generate the polar lattice of the lattice generated by the rows of $M$.

**2.2 Base cases**

In case $d = 1$, Theorem 1.3 is precisely Theorem 4 in [20] (also Lemma 2 of [12]), which states that

$$
P(\mathcal{L}, 1, H) = a(n, 1) \frac{H^n}{\det \mathcal{L}} + O \left( \sum_{i=1}^{n} \frac{H^{n-i}}{\det \mathcal{L}^{n-i}} \right). \tag{2.1}
$$

Below in Lemma 3.6, we present an extension of (2.1) to an affine lattice, which we will need later.

In case $d = n - 1$, we apply the duality theorem (see Section 2 of [20]) to (2.1), which states that, for a sublattice $S \subseteq \mathcal{L}$ and its orthogonal lattice $S^\perp \subseteq \mathcal{L}^P$,

$$
\det S^\perp = \frac{\det S}{\det \mathcal{L}}
$$

holds, and thus

$$
P(\mathcal{L}, d, H) = P(\mathcal{L}^P, n - d, \frac{H}{\det \mathcal{L}}). \tag{2.2}
$$

Therefore (2.1) implies

$$
P(\mathcal{L}, n - 1, H) = a(n, n - 1) \frac{H^n}{(\det \mathcal{L})^n \det \mathcal{L}^P} + O \left( \sum_{i=1}^{n} \frac{H^{n-i}}{(\det \mathcal{L})^{n-i} \det(\mathcal{L}^P)^{n-i}} \right).
$$

 Springer
By the well-known facts that \( \det L \cdot \det L^P = 1 \) and \( \lambda_i(L)\lambda_{n-i+1}(L^P) \geq 1 \) (in fact, \( \lambda_i(L)\lambda_{n-i+1}(L^P) \sim 1 \), by [1, Theorem 2.1]),

\[
\det(L^P)_{n-i} \gg \det L_i / \det L,
\]

so we can rewrite the above as

\[
P(L, n - 1, H) = a(n, n - 1) \cdot \frac{H^n}{(\det L)^{n-1}} + O \left( \sum_{i=1}^{n} \frac{H^{n-i}}{(\det L_i \cdot (\det L)^{n-1-i})} \right).
\]

### 3 Division into two parts

#### 3.1 Preliminaries

Until the end of Sect. 6, we fix \( n \geq 4 \) and \( 2 \leq d \leq n - 2 \). We will divide \( P(L, d, H) \) into two parts, and deal with them one at a time. We induct on \( n \), assuming that \( P \) has been computed for all lattices of rank \( < n \).

Throughout the rest of the paper, we fix a basis \( \{v_1, \ldots, v_n\} \) of \( L \), and denote by \( L \) the \( n \times n \) matrix whose \( i \)-th row is \( v_i \). Define \( \tilde{L} = L/\langle v_n \rangle \), and identify it with the projection of \( L \) onto the subspace of \( \mathbb{R}^n \) orthogonal to \( v_n \), i.e. we think of \( \tilde{L} \) as a subset of \( \mathbb{R}^n \). Let \( \tilde{v}_i \) be the component of \( v_i \) orthogonal to \( v_n \), so that \( v_i = \tilde{v}_i + a_i v_n \) for some \( a_i \in \mathbb{R} \) and \( \tilde{L} = \text{span}_{\mathbb{Z}}(\tilde{v}_1, \ldots, \tilde{v}_{n-1}) \). Also let \( \tilde{L} \) be the \( (n - 1) \times n \) matrix whose \( i \)-th row is \( \tilde{v}_i \).

We write

\[
P(L, d, H) = P^1(L, d, H) + P^2(L, d, H),
\]

where \( P^1(L, d, H) \) equals the number of rank \( d \) primitive sublattices of \( L \) of determinant \( \leq H \) such that its projection to \( \tilde{L} \) is also of rank \( d \), and \( P^2(L, d, H) \) equals the number of those whose projection is of rank \( d - 1 \). Equivalently, \( P^1 \) counts primitive sublattices whose \( \mathbb{R} \)-span does not contain \( v_n \), and \( P^2 \) counts those that does.

As discussed in Section 1.4 above, we may identify \( A \in \text{Gr}(L, d) \) with an orbit \( \Gamma M L \) of the left multiplication of \( \Gamma \) on \( \text{Mat}^p_{d \times n}(\mathbb{Z}) \cdot L \), for some \( M = (c_{ij})_{1 \leq i \leq d, 1 \leq j \leq n} \in \text{Mat}^p_{d \times n}(\mathbb{Z}) \). Also, let \( \tilde{L} \) be the \( n \times n \) matrix whose \( i \)-th row vector equals \( \tilde{v}_i \) for \( 1 \leq i \leq n - 1 \), and \( v_n \) for \( i = n \), so that

\[
L = \begin{pmatrix}
1 & a_1 \\
1 & a_2 \\
\vdots & \vdots \\
1 & a_{n-1} \\
1 & a_n
\end{pmatrix}
\]
Then we can also write $\mathcal{A}$ in the form $\Gamma(C; c + c') \tilde{L}$, where $C = (c_{ij})_{1 \leq i \leq d, 1 \leq j \leq n-1}$ is the first $d \times (n-1)$ submatrix of $M$, and $c = (c_1, \ldots, c_{dn})^T$ and $c' = (\sum_j a_j c_{1j}, \ldots, \sum_j a_j c_{dj})^T$ are vectors in $\mathbb{R}^d$.

### 3.2 Computing $P^2(\mathcal{L}, d, H)$

Consider first the case rank $C = d - 1$, so that $\mathcal{A}$ contributes to $P^2$. We may assume that $M$ is a Hermite normal form, so that $C$ is too. Because $M$ is primitive, so is $C$, and the $d$-th entry of the vectors $c$ and $c'$ must be equal to 1 and 0 respectively. This forces each of the other entries of $c + c'$ to have only one choice modulo the left action of $\Gamma$. Thus

\[
P^2(\mathcal{L}, d, H) = P(\bar{\mathcal{L}}, d - 1, \frac{H}{\|v_n\|}),
\]

(3.1)

to which we can simply apply Theorem 1.3 (see the remark after its statement).

### 3.3 Some lemmas

Working with $P^1$ is much more involved. Most of the remainder of this paper is devoted to this task. The goal of this section is to derive the expression (3.5) for $P^1$ that is amenable to computation.

We start by recalling the standard choice of the representatives of the right cosets of $\Gamma$ in the double coset $\Gamma a \Gamma$, where $a \in \text{Mat}_{d \times d}(\mathbb{Z})$ has determinant $k > 0$. Such a representative, say $h = (h_{ij})_{1 \leq i, j \leq d}$, is a lower triangular matrix with determinant $k$, with the condition that $0 \leq h_{ji} < h_{ii}$ for all $j > i$. Of course, $\Gamma h \subseteq \Gamma a \Gamma$ if and only if $a$ and $h$ have the same invariant factors.

**Lemma 3.1** Given an integral $d \times n$ matrix $(C; c)$ with rank $C = d$, there exists a unique triple $(h, B, b)$, where $h$ is one of the right coset representatives described above, $B$ is a $d \times (n-1)$ primitive Hermite normal form of rank $d$, and $b \in \mathbb{Z}^n$, such that $(C; c) \sim (hB; b)$.

**Proof** By the theory of the Smith normal form, we have $(C; c) \sim (aB_0; b_0)$ where $a$ is an invariant factor matrix—that is, $a = \text{diag}(a_1, \ldots, a_d)$ with $a_i|a_{i+1}$—$B_0$ is a primitive $d \times (n-1)$ matrix of full rank, and $b_0 \in \mathbb{Z}^d$. Write $B_0 = \gamma B$, where $B$ is the Hermite normal form of $B_0$ and $\gamma \in \Gamma$. Then there exists $\gamma' \in \Gamma$ and $h$ a coset representative of $\Gamma a \Gamma$ such that $\gamma'h = a\gamma$. Therefore, writing $b = \gamma'^{-1}b_0$, we have $(C; c) \sim (hB, b)$.

Suppose we have another triple $(h', B', b')$ such that $(hB, b) \sim (h'B', b')$. This is possible only if the row vectors of $B$ and $B'$ generate the same lattice. Since both $B$ and $B'$ are in the Hermite normal form, $B = B'$. This in turn implies $h = h'$ and $b = b'$.

**Lemma 3.2** Again given an integral $d \times n$ matrix $(C; c)$, write $C = \gamma aB$, where $\gamma \in \Gamma$, $a = \text{diag}(a_1, \ldots, a_d)$ is an invariant factor matrix, and $B$ is primitive. Thus $(C; c) \sim (aB; \gamma^{-1}c) = (aB; b)$, where $b := \gamma^{-1}c$. 

Springer
Then \((aB; b)\) is primitive if and only if \(a_1 = \ldots = a_{d-1} = 1\) and \(b_d\) is coprime to \(a_d\).

**Proof** Without loss of generality, we may assume \(B\) to be the matrix which has 1’s in the diagonal and 0’s elsewhere. \((aB, b)\) is imprimitive if and only if there exist integers \(0 \leq r_i < a_i\) for \(i = 1, \ldots, d, r_i\) not all zero, such that \((r_1, \ldots, r_d, 0, \ldots, 0, \sum_i b_i r_i / a_i) \in \mathbb{Z}^n\), or equivalently \(\sum_i b_i r_i / a_i \in \mathbb{Z}\).

Suppose \(a_{d-1} \neq 1\). We claim that, for any \(b_d\), \(b_d - 1\) and \(p\), \(b_d - 1 + b_d p \equiv 0\) (mod \(p\)).

Next suppose \(a_{d-1} = 1\). We are led to consider the condition \(b_d r_d / a_d \in \mathbb{Z}\). This is impossible if and only if \((b_d, a_d) = 1\), which completes the proof.

**Lemma 3.3** Write \(e(p^\alpha) = \text{diag}(1, \ldots, 1, p^\alpha)\). Then the necessary and sufficient condition for \(h \in \text{Mat}_{d \times d}(\mathbb{Z})\) to be one of the standard form right coset representatives of \(\Gamma \in \text{Mat}_{d \times d}(\mathbb{Z})\) is as follows: \(h\) is a lower triangular matrix with \(h_{ii} = p^{a_i}\), where \(a_i \geq 0\) and \(\sum a_i = \alpha\), \(0 \leq h_{ji} < h_{ii}\) for \(j > i\), and in addition if \(i < j\) are two indices such that \(a_i, a_j \geq 1\) and \(a_{i+1} = \ldots = a_{j-1} = 0\)—i.e. all diagonal entries between \(h_{ii}\) and \(h_{jj}\) are trivial—then \((h_{ji}, p) = 1\).

**Proof** Let \(h\) be a coset representative of some double coset of a matrix of determinant \(p^\alpha\), in the form that we chose in the beginning of this section. Then all but the last condition are automatically satisfied. For the last condition, choose the three smallest indices \(i < j < k\) for which \(a_i, a_j, a_k > 0\). We consider the \(3 \times 3\) matrix

\[
\begin{pmatrix}
p^{a_i} & h_{ji} & p^{a_j} \\
h_{ki} & h_{kj} & p^{a_k}
\end{pmatrix}.
\]

We will show that this matrix has invariant factors \((1, 1, p^{a_i+a_j+a_k})\) if and only if \(h_{ji}\) and \(h_{kj}\) are coprime to \(p\). Then the proof is complete because we can repeatedly apply this argument to \(h\) to compute the invariant factors of \(h\).

If \(h_{ji}\) and \(p\) are coprime, there exist integers \(x, y\) such that \(y h_{ji} - x p^{a_i} = 1\), so that the matrix

\[
\begin{pmatrix}
h_{ji} & p^{a_i} & 0 \\
x & y & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

has determinant 1. Multiplying this on the left of (3.2), we have

\[
\begin{pmatrix}
0 & p^{a_i+a_j} & 0 \\
1 & y p^{a_j} & 0 \\
h_{ki} & h_{kj} & p^{a_k}
\end{pmatrix}.
\]
which, upon multiplying by suitable elements of $\Gamma$ from both sides, becomes
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & p^{a_i + a_j} & 0 \\
0 & h_{kj} - yp^{a_j}h_{ki} & p^{a_k}
\end{pmatrix}.
\]

If furthermore $h_{kj}$ is coprime to $p$, then so is $h_{kj} - yp^{a_j}h_{ki}$, so we can use the same trick to see that (3.2) has invariant factors $1$, $1$, $p^{a_i + a_j + a_k}$ indeed.

Now go back to (3.2) and consider the case $h_{ji} = cp^b$; we can assume $1 \leq b < a_j$ and $(c, p) = 1$. We restrict our attention to the $2 \times 2$ upper-left corner submatrix of (3.2), and temporarily use $\approx$ to denote the equivalence under the left and right multiplication by $\Gamma$. Then, by a similar argument as earlier, for an appropriate integer $y$,
\[
\begin{pmatrix}
p^{a_i} \\
cp^b \\
p^{a_j}
\end{pmatrix} = \begin{pmatrix}
p^{a_i - b} \\
c \\
p^{a_j}
\end{pmatrix} \begin{pmatrix}
p^b \\
1 \\
p^b
\end{pmatrix} \approx \begin{pmatrix}0 & p^{a_i + a_j - b} \\
1 & yp^{a_j} \\
1 & yp^{a_j}
\end{pmatrix} \begin{pmatrix}
p^b \\
0 \\
p^b
\end{pmatrix} \approx \begin{pmatrix}0 & p^{a_i + a_j - b} \\
p^b \\
0
\end{pmatrix},
\]
so $p^b$ appears as one of the invariant factors. \hfill \square

**Lemma 3.4** Write $e(k) = \text{diag}(1, \ldots, 1, k)$, as in the previous lemma. Then the number of the right cosets of $\Gamma$ in $\Gamma e(k)\Gamma$ equals
\[
\prod_{p^d \parallel k} p^{(a - 1)(d - 1)}(1 + p + \ldots + p^{d - 1}).
\]

**Proof** From the general theory of Hecke operators (see Chapter 3 of Shimura [16]), it suffices to prove the lemma for the case $k = p^\alpha$. We proceed by induction on $\alpha$.

In case $\alpha = 1$, there exist $p^{d - 1}$ coset representatives which has $a_{ii} = p$ and $a_{ij} = 1$ for all $j \neq i$. This exhausts all the representatives of $\Gamma e(p)\Gamma$, so the lemma holds true in this case.

For the general case, it suffices to match, to each representative $h$ of $\Gamma e(p^{\alpha - 1})\Gamma$, $p^{d - 1}$ representatives of $\Gamma e(p^\alpha)\Gamma$, different for each $h$. Suppose $j$ is the smallest number for which $h_{jj}$ is a power of $p$. Then modifying $h_{jj}$ to $ph_{jj}$ and $h_{kj}(k > j)$ to $h_{kj} + c_kh_{jj}$, for any choice of $0 \leq c_k < p$, yields a representative of $\Gamma e(p^\alpha)\Gamma$, accounting for $p^{d - j}$ out of $p^{d - 1}$ total. Also, for each $i < j$, replacing $h_{ii}(= 1)$ by $p$, a choice of each $h_{ki}(k \neq j)$ from $\{0, \ldots, p - 1\}$ and of $h_{ji}$ from $\{1, \ldots, p - 1\}$ ($h_{ji}$ cannot be 0 by the previous lemma) yields a representative of $\Gamma e(p^\alpha)\Gamma$, and there are $p^{d - i - 1}(p - 1)$ of this kind. Therefore, for each $h$ there is a total of $p^{d - j} + p^{d - j}(p - 1) + p^{d - j + 1}(p - 1) + \ldots + p^{d - 2}(p - 1) = p^{d - 1}$ coset representatives of $\Gamma e(p^\alpha)\Gamma$ constructed in this manner, as desired. It remains to show that these representatives do not overlap with those constructed from a different choice of $h$. But this is immediate since, given a representative of $\Gamma e(p^\alpha)\Gamma$, one can read off which representative of $\Gamma e(p^{\alpha - 1})\Gamma$ it came from, by discarding the first factor of $p$ that appears in its diagonal. \hfill \square
3.4 A computable expression for $P^1(L, d, H)$

For $A \in \text{Gr}(L, d)$, recall we defined $f_H(A) = 1$ if $\det A \leq H$ and 0 otherwise. Also, as in the statement of Lemma 3.4 write $e(k) := \text{diag}(1, \ldots, 1, k)$. Thanks to Lemmas 3.1, 3.2 and 3.4, we can rewrite $P^1(L, d, H)$ as

$$\sum_{B \in \Gamma \setminus \text{Mat}^{\text{pr}}_{d \times (n-1)}(\mathbb{Z})} \sum_{h \in \mathbb{Z}} \sum_{b \in \mathbb{Z}^d_{(hB;b) \text{prim}}} f_H((hB; b) L),$$

where the sum over $h$ is taken over all coset representatives of $\Gamma e(k) \Gamma$ in the standard form.

Fix $h, k, B$ for a moment, and consider the innermost summation in (3.3). For some $B' \sim B$, it is equal to (cf. Lemma 3.2)

$$\sum_{b \in \mathbb{Z}^d} f_H((e(k)B'; b) L)$$

$$= \sum_{l | k} \mu(l) \sum_{b \in \mathbb{Z}^d} f_H((e(k)B'; e(l)b) L)$$

$$= \sum_{l | k} \mu(l) \sum_{b \in \mathbb{Z}^d} f_H((e(k)B'; e(l)b + e(k)t) \tilde{L})$$

$$= \sum_{l | k} \mu(l) \sum_{b \in \mathbb{Z}^d} f_H(e(k)B' \tilde{L} + (e(l)b + e(k)t) v_n),$$

where $\mu$ is the Möbius function, and we wrote

$$t = \left( \begin{array}{c} \sum_j a_j b'_{1j} \\ \vdots \\ \sum_j a_j b'_{dj} \end{array} \right)$$

for short. Note that $v_n$ is a row vector, whereas $b$ and $t$ are column vectors.

Temporarily write $X = (e(l)b + e(k)t) v_n$ and $Y = e(k)B' \tilde{L}$. We need to compute the determinant of $X + Y$. First observe that

$$(X + Y)(X + Y)^{\text{tr}} = XX^{\text{tr}} + XY^{\text{tr}} + YX^{\text{tr}} + YY^{\text{tr}} = XX^{\text{tr}} + YY^{\text{tr}},$$

because $v_n \tilde{L}^{\text{tr}} = 0$, and also

$$XX^{\text{tr}} = \|v_n\|^2 (e(l)b + e(k)t) (e(l)b + e(k)t)^{\text{tr}}.$$

This motivates the use of the matrix-determinant lemma, which asserts that for a $d \times d$ matrix $A$ and (row) vectors $x, y \in \mathbb{R}^d$, $\det(A + x^{\text{tr}}y) = \det A \cdot (1 + y A^{-1} x^{\text{tr}})$. To this end, we also need the following lemma.
Lemma 3.5 Let $Y$ be a full-rank $d \times n$ matrix whose $i$-th row equals $y_i \in \mathbb{R}^n$. Let $z_1, \ldots, z_d \in \mathbb{R}^n$ such that they form the basis of the polar lattice spanned by $y_1, \ldots, y_d$ and that $\langle z_i, y_j \rangle = \delta_{ij}$. Let $Z = Y^P$ be the $d \times n$ matrix whose $i$-th row equals $z_i$. Then the inverse of $Y Y^\mathrm{tr}$ is given by $Z Z^\mathrm{tr}$.

**Proof** Complete $Y$ to an invertible $n \times n$ matrix $\bar{Y} = \begin{pmatrix} Y \\ \nu \end{pmatrix}$, such that the rows of $\bar{Y}$ are orthogonal to the rows of $Y$. Similarly complete $Z$ to $\bar{Z} = \begin{pmatrix} Z \\ \nu \end{pmatrix}$, so that the rows of $\bar{Z}$ form the dual basis to that formed by the rows of $\bar{Y}$. Then the rows of $Z'$ are orthogonal to the rows of $Z$ as well.

Since $\bar{Z}$ and $\bar{Y}^\mathrm{tr}$ are inverses of each other, we have $\bar{Y} Y^\mathrm{tr} \bar{Z} Z^\mathrm{tr} = I$. By abuse of language, write $Y = \begin{pmatrix} Y \\ 0 \end{pmatrix}$, $Y' = \begin{pmatrix} 0 \\ Y \end{pmatrix}$, and similarly with $Z$. Then

$$Y Y^\mathrm{tr} \bar{Z} Z^\mathrm{tr} = (Y + Y')(Y^\mathrm{tr} Z + Y'^\mathrm{tr} Z')(Z^\mathrm{tr} + Z'^\mathrm{tr}) = Y Y^\mathrm{tr} Z Z^\mathrm{tr} + Y' Y'^\mathrm{tr} Z' Z'^\mathrm{tr},$$

and observe that the first term on the right is zero outside the first $d \times d$ submatrix, and the second term is zero outside the “last” $(n - d) \times (n - d)$ submatrix. This completes the proof. \hfill \Box

Thanks to the above lemma, with $Z = Y^P$ we compute that $\det(X + Y)$ is the square root of

$$\det(X X^\mathrm{tr} + Y Y^\mathrm{tr}) = \det \left( \left\| v_n \right\|^2 (e(l)b + e(k)t) (e(l)b + e(k)t)^\mathrm{tr} + Y Y^\mathrm{tr} \right) = \det(Y Y^\mathrm{tr}) \left( 1 + \left\| v_n \right\|^2 (e(l)b + e(k)t)^\mathrm{tr} (Z Z^\mathrm{tr}) (e(l)b + e(k)t) \right) = k^2 \det(B' L)^2 \left( 1 + \left\| v_n \right\|^2 (e(l)b + e(k)t)^\mathrm{tr} e(k^{-1})(B' L)^P \right)^2.$$

In the last line, we used the fact that $Z = (e(k)B' L)^P = e(k^{-1})(B' L)^P$.

Let

$$K(B) = \frac{1}{\left\| v_n \right\|} \sqrt{\frac{H^2}{k^2 \det(B L)^2} - 1}$$

if $H \geq k \det(B L)$, and set $K(B) = 0$ otherwise. We will use this notation throughout the rest of the paper. Then (3.4) becomes

$$\sum_{l|k} \mu(l) \cdot \left( \text{number of vectors(nonzero, if } k \neq 1) \text{in the lattice spanned by the rows of } e(l/k)(B' L)^P \text{ whose translates by } l^\mathrm{tr}(B' L)^P \text{ have length } \leq K(B') \right).$$

The lemma below ensures that the translation of the vectors by $l^\mathrm{tr}(B' L)^P$ does not present any extra difficulty in our estimate of this sum.

---

Springer
Lemma 3.6 Let $\Lambda \in \mathbb{R}^d$ be a lattice of rank $d$, and $t \in \mathbb{R}^d$. Temporarily denote by $N(r)$ the number of points $v \in \Lambda + t$ with $\|v\| \leq r$. Then

$$N(r) = \frac{V(d)r^d}{\det \Lambda} + O \left( \sum_{i=1}^{d} \frac{r^{d-i}}{\det \Lambda_{d-i}} \right),$$

where the implicit constant depends on $d$ only.

Proof This is Lemma 2 in [12] generalized to an affine lattice, and is also a special case of Theorem 5.4 in [22]. We provide a proof here for completeness.

We proceed by induction on $d$. The base case $d = 1$ is clear. Now assume the lemma for $d - 1$. By adjusting $\det \Lambda$, we may assume $r = 1$.

First consider the case $\lambda_d \leq 1$. Let $x_i \in \Lambda$, $i \in \{1, \ldots, d\}$, be a vector with $\|x_i\| = \lambda_i$, and consider the parallelepiped spanned by $x_1, \ldots, x_d$. Its diameter is $\leq \lambda_1 + \ldots + \lambda_d \leq d\lambda_d$, and it contains a fundamental parallelepiped $F$ of $\Lambda$, which also has diameter $\leq d\lambda_d$.

Write $B(s)$ for the ball in $\mathbb{R}^n$ at the origin of radius $s$. Then since $B(\max(0, 1 - d\lambda_d)) \subseteq (\Lambda + t) \cap B(1) + F \subseteq B(1 + d\lambda_d)$, we have

$$|N(r)\det \Lambda - V(d)| \leq V(d)((1 + d\lambda_d)^d - \max(0, 1 - d\lambda_d)^d) \leq V(d)(2d\lambda_d)^d,$$

and thus

$$N(r) = O \left( \frac{\lambda_d}{\det \Lambda} \right) = O \left( \frac{1}{\det \Lambda_{d-1}} \right),$$

where the second equality follows from the Minkowski’s second theorem.

It remains to consider the case $\lambda_d > 1$. Then $(\Lambda + t) \cap B(1)$ lies in at most two translates of $\Lambda_{d-1}$ in the direction of $x_d$. Thus the induction hypothesis implies $N(r) = O \left( \sum_{i=1}^{d-1} 1/ \det \Lambda_{d-i} \right)$. Also we have

$$\frac{1}{\det \Lambda} < \frac{\lambda_d}{\det \Lambda} = O \left( \frac{1}{\det \Lambda_{d-1}} \right)$$

as above. This completes the proof.

It follows that (3.4) equals

$$\sum_{l|k} \mu(l) \left( \frac{V(d)K(B')^d}{\det \mathcal{L}(l/k, (B'\bar{L})^P)} + O \left( \sum_{i=1}^{d} \frac{K(B')^{d-i}}{\det \mathcal{L}(l/k, (B'\bar{L})^P)^{d-i}} \right) \right),$$

where $\mathcal{L}(x, M)$ here denotes the lattice spanned by the row vectors of $e(x)M$. We have $\mathcal{L}(l/k, (B'\bar{L})^P) = \mathcal{L}(k/l, B'\bar{L})^P$, and $\det(\mathcal{L}(k/l, B'\bar{L})^P)_{d-i}$
det $\mathcal{L}(k/l, B' \bar{L})_i$ / det $\mathcal{L}(k/l, B' \bar{L})$ by (2.3). Also, det $\mathcal{L}(k/l, B' \bar{L})_i \gg$ det $\mathcal{L}(1, B' \bar{L})_i$, so the above sum can be rewritten as

$$\sum_{l | k} \mu(l) \frac{k}{l} \left( \frac{V(d) K(B')^d}{\det \mathcal{L}(1, B' \bar{L})^p} + O \left( \sum_{i=1}^{d} \frac{K(B')^{d-i} \det \mathcal{L}(1, B' \bar{L})}{\det \mathcal{L}(1, B' \bar{L})_i} \right) \right),$$

which we in turn rewrite as, for the lattice $\mathcal{B} \in \text{Gr}(\mathbb{Z}^{n-1}, d)$ spanned by $B'$,

$$\sum_{l | k} \mu(l) \frac{k}{l} \left( \frac{V(d) K(B)^d}{\det (\bar{B}L)^p} + O \left( \sum_{i=1}^{d} \frac{K(B)^{d-i} \det \bar{B}L}{\det (\bar{B}L)_i} \right) \right).$$

Summing up all our work in this section, we deduce that (3.3) equals

$$\sum_{\mathcal{B} \in \text{Gr}(\mathbb{Z}^{n-1}, d)} \prod_{k \geq 1} \frac{p^{-(\alpha-1)(d-1)}(1 + p + \ldots + p^{d-1})}{p^\alpha \|k\} \sum_{l | k} \mu(l) \frac{k}{l} \left( \frac{V(d) K(B)^d}{\det (\bar{B}L)^p} \right) + O \left( \sum_{i=1}^{d} \frac{K(B)^{d-i} \det \bar{B}L}{\det (\bar{B}L)_i} \right)$$

$$= \sum_{k \geq 1} \prod_{p \|k} \frac{p^{-(\alpha-1)(d-1)}(1 + p + \ldots + p^{d-1}) \varphi(k) V(d)}{p^\alpha \|k} \sum_{\mathcal{B} \in \text{Gr}(\mathbb{Z}^{n-1}, d)} \left( K(B)^d \det \bar{B}L \right) + O \left( \sum_{i=1}^{d} \frac{K(B)^{d-i} \det \bar{B}L}{\det (\bar{B}L)_i} \right).$$

(3.5)

Here $\varphi(k) = \sum_{l | k} \mu(l) \frac{l}{k}$ is the Euler totient.

The remainder of this paper is devoted to computing (3.5). Because $K(B)$ depends on $k$, we cannot deal with the constant factor just yet. However, we will later use

**Lemma 3.7** For $m > d + 1$, 

$$\sum_{k \geq 1} \prod_{p \|k} \frac{p^{-(\alpha-1)(d-1)}(1 + p + \ldots + p^{d-1}) \varphi(k) k^{-m}}{p^\alpha \|k} = \frac{\xi(m-d)}{\xi(m)}.$$

**Proof** We can write the expression under question multiplicatively as

$$\sum_{k \geq 1} \prod_{p \|k} p^{-(m-d)\alpha} \left( 1 - \frac{1}{p^d} \right) = \prod_{p} \left( 1 + \sum_{i \geq 1} (1 - p^{-d}) p^{-i(m-d)} \right),$$
which then becomes
\[
\prod_p \left( \sum_{i \geq 0} p^{-i(m-d)} - p^{-m} \sum_{i \geq 0} p^{-i(m-d)} \right) = \prod_p (1 - p^{-m})(1 - p^{-m-d})^{-1} = \zeta(m - d) \zeta(m).
\]

\[\Box\]

### 4 Main term of (3.5)

In this section, we estimate the intended main term of (3.5), namely
\[
\sum_{\mathcal{B} \in \text{Gr}(\mathbb{Z}^{n-1}, d)} K(\mathcal{B})^d \det(\mathcal{B} \tilde{L}),
\]
(4.1)

for each \(k \geq 1\) and \(2 \leq d \leq n - 2\). We may also assume \(H \geq k \min_{\mathcal{B}} \det(\mathcal{B} \tilde{L})\), since otherwise (4.1) is equal to 0. Our approach is essentially that of Schmidt [12], who uses summation by parts. We improve it somewhat by adopting the language of the Riemann-Stieltjes integral, in order to simplify the computation and to derive pretty error terms.

Let us rewrite (4.1) as
\[
\frac{1}{\|v_n\|^d k^d} Q(k, H),
\]
(4.2)

where
\[
Q(k, H) := \sum_{\mathcal{B} \in \text{Gr}(\mathbb{Z}^{n-1}, d)} \psi(\det(\mathcal{B} \tilde{L}))
\]

and
\[
\psi(t) = \begin{cases} t((H/t)^2 - k^2)^{d/2} & \text{for } 0 < t \leq H/k, \\ 0 & \text{otherwise} \end{cases}
\]

It is easy to check that \(\psi(t)\) is a twice differentiable function on \(0 < t \leq H/k\), with \(\psi'(t) = -((d-1)(H/t)^2 + k^2)((H/t)^2 - k^2)^{(d/2-1)} \leq 0\).

Choose a \(\delta > 0\) with \(\delta \leq \min_{\mathcal{B}} \det(\mathcal{B} \tilde{L})\). Write \(H/k = (\alpha + s)\delta\) with \(\alpha \in [0, 1)\) and \(s \in \mathbb{Z}\). Also, let \(P_1(t)\) be the number of elements \(\mathcal{B} \in \text{Gr}(\mathbb{Z}^{n-1}, d)\) such that \(t < \det(\mathcal{B} \tilde{L}) \leq t + \delta\), and \(P_2(t) = P_1(t - \delta)\). Then for \(i = 1, 2\),
\((-1)^i \left( Q(k, H) - \sum_{j=0}^{s-1} \psi((\alpha + j)\delta)P_i((\alpha + j)\delta) \right) \geq 0. \)

Write \( R_1(t) = P(\bar{\mathcal{L}}, d, t + \delta) \) and \( R_2(t) = P(\bar{\mathcal{L}}, d, t) \). Since \( \psi((a + s)\delta) = 0 \), by the summation by parts,
\[
(-1)^i \left( Q(k, H) - \sum_{j=0}^{s-1} R_i((\alpha + j)\delta)(\psi((\alpha + j)\delta) - \psi((\alpha + j + 1)\delta)) \right) \geq 0.
\]

Thus we have bounded \( Q(k, H) \) from both sides by certain Riemann-Stieltjes sums. We need to show that those sums converge as \( \delta \to 0 \) (and thus \( s \sim H/k\delta \to \infty \)). First, observe that, since \( R_i \)'s are supported strictly away from zero by \( \epsilon = \min B \det(B \bar{\mathcal{L}}) \), we may assume the same of \( \psi \), so that \( \psi \) is of bounded variation. Second, \( R_i \) are clearly not continuous, but by the induction hypothesis on \( n \), we know it is bounded from both sides by a polynomial in \( t \). More precisely,
\[
R_2(t) = a(n - 1, d) \frac{t^{n-1}}{\det(\bar{\mathcal{L}})^d} + O \left( \sum_{j \in E_{n-1,d}} c_j t^\gamma_j \right)
\]
where \( c_j = b_j(\bar{\mathcal{L}}) \) is as in Theorem 1.3, and
\[
R_1(t) = R_2(t) + O_{t, \bar{\mathcal{L}}} (\delta).
\]

By Theorem 6.8 of Rudin [10], we have shown that
\[
Q(k, H) = a(n - 1, d) \frac{t^{n-1}}{\det(\bar{\mathcal{L}})^d} \int_\epsilon^{H/k} t^{n-1} \psi'(t) dt + O \left( \sum_{j \in E_{n-1,d}} c_j \int_\epsilon^{H/k} t^\gamma_j \psi'(t) dt \right).
\]

Since the same argument will be used repeatedly later in this paper, we summarize our discussion so far in the form of a lemma:

**Lemma 4.1** Assume Theorem 1.3 for \( n = m \), and let \( \mathcal{M} \) be a lattice of rank \( m \). Suppose \( \psi \) is a decreasing twice differentiable function supported on \( [a, b] \). Then
\[
\sum_{B \in \text{Gr}(\mathcal{M}, d)} \psi(\det B) = a(m, d) \frac{t^{n-1}}{\det(\mathcal{M})^d} \int_a^b -t^m \psi'(t) dt + O \left( \sum_{j \in E_{m,d}} b_j(\mathcal{M}) \int_a^b -t^\gamma_j \psi'(t) dt \right).
\]

We return to estimating (4.3). Recall \( \epsilon = \min B \det(\mathcal{B} \mathcal{L}) \sim \prod_{i=1}^d \lambda_i(\bar{\mathcal{L}}) \). In (4.3), for the integrals inside the \( O \)-notation, there is no harm in replacing \( \epsilon \) with 0 if \( \gamma > d - 1 \). For the main term, we can do the same at the cost of
\[
\frac{1}{(\det \mathcal{L})^d} \int_0^e -t^{n-1} \psi'(t) dt \ll \frac{1}{(\det \mathcal{L})^d} \int_0^e H^d t^{n-d-1} dt \sim \frac{H^d e^{n-d}}{(\det \mathcal{L})^d} \ll \frac{H^d}{e^{d-1}}.
\]

Now the main term of \( Q(k, H) \) contributes

\[
a(n - 1, d) \left( -t^{n-1} \psi(t) \right)_{H/k}^0 + (n - 1) \int_0^{H/k} t^{n-2} \psi(t) dt = \frac{a(n - 1, d)}{(\det \mathcal{L})^d} n \int_0^{H/k} t^{n-1} \left( \frac{H^2}{t^2} - k^2 \right)^{d/2} dt = \frac{a(n - 1, d)}{(\det \mathcal{L})^d} n \frac{H^d k^{-n+d}}{(n-d)V(n)} = \frac{a(n, d)}{(\det \mathcal{L})^d} \frac{\zeta(n)}{(n-d)V(d)} H^n k^{-n+d}.
\]

For the second last equality, we used the identity on the beta function (see e.g. [3, Section 6.2.1])

\[
B(a, b) = 2 \int_0^1 x^{2a-1} (1 - x^2)^{b-1} dx,
\]

and the last equality follows from the definition of \( a(n, d) \).

Now accounting for the factor of \( 1/\|v_n\|^d k^d \) in (4.2), we obtain for the intended main term of (4.1)

\[
\frac{a(n, d)}{(\det \mathcal{L})^d \|v_n\|^d k^d} \frac{\zeta(n)}{(n-d)V(d)} H^n k^{-n+d} = \frac{a(n, d)}{(\det \mathcal{L})^d \zeta(n-d)V(d)} H^n k^{-n}.
\]

It is clear that this term is scale-invariant i.e. invariant under replacing \( \mathcal{L} \) and \( H \) by \( c\mathcal{L} \) and \( c^d H \) for any \( c > 0 \).

The error terms of \( Q(k, H) \) are dealt with in a similar way, only simpler. For \( j \in E_{m,d} \) with \( \gamma_j > d - 1 \), the corresponding term contributes

\[
c_j H/k \int_0^{H/k} -t^{\gamma_j} \psi'(t) dt \leq c_j \int_0^{H/k} -t^{\gamma_j} \psi'(t) dt = -c_j t^{\gamma_j} \psi(t)_{H/k}^0 + c_j \gamma_j \int_0^{H/k} t^{\gamma_j-1} \psi(t) dt
\]

\( \mathbb{C} \) Springer
\[ \ll c_j \int_0^{H/k} t^{\gamma_j} (H/t)^d \, dt \ll c_j H^{\gamma_j+1}k^{d-\gamma_j-1}. \]

It is apparent that \(c_j H^{\gamma_j+1}k^{d-\gamma_j-1}/(\|v_n\|^d k^d)\) is scale-invariant, since both \(c_j H^{\gamma_j}\) and \(H/\|v_n\|^d\) are.

For those with \(\gamma_j \leq d - 1\), we proceed as follows:

\[
\begin{align*}
& c_j \int_{\varepsilon}^{H/k} -t^{\gamma_j} \psi'(t) \, dt \\
& = -c_j t^{\gamma_j} \psi(t) \bigg|_{\varepsilon}^{H/k} + c_j \gamma_j \int_{\varepsilon}^{H/k} t^{\gamma_j-1} \psi(t) \, dt \\
& \leq c_j \varepsilon^{\gamma_j+1-d} H^d + c_j \gamma_j \int_{\varepsilon}^{H/k} t^{\gamma_j} (H/t)^d \, dt \\
& \ll \begin{cases} 
  c_j H^d \varepsilon^{-d+\gamma_j+1} & \text{if } \gamma_j < d - 1 \\
  c_j H^d (1 + \log \frac{H}{k\varepsilon}) & \text{if } \gamma_j = d - 1
\end{cases}.
\end{align*}
\]

In case \(\gamma_j = d - 1\), we used our assumption \(H \geq k\varepsilon\). Also, to retain the polynomial shape of the error term, we note \(c_j H^d (1 + \log \frac{H}{k\varepsilon}) = O\left(c_j H^{d+\eta}(k\varepsilon)^{-\eta}\right)\) for any \(\eta > 0\), and use this bound instead. The scale-invariance can be checked in a straightforward manner.

In conclusion, we proved that (4.1) equals

\[
\frac{a(n, d)}{\det \mathcal{L}^d} \frac{\chi(n)}{\xi(n-d)\nu(d)} (H/k)^n + O\left(\sum_{j \in E(1)} c_j' (H/k)^{\gamma_j}\right),
\]

where \(E(1)\) is an index set of cardinality \(|E_{n-1,d}| + 1\), each \(c_j'\) is a reciprocal of products of \(\lambda_i(\tilde{L})\)'s and \(\|v_n\|\), so that \(c_j' (H/k)^{\gamma_j}\) is invariant under the appropriate scaling. The leading error term is of degree \(n - b(n-1, d)\).

5 Error term of (3.5)

In this section, we work on the intended error term of (3.5), namely

\[
\sum_{B \in \mathrm{Gr}(\mathbb{Z}^{n-1}, d)} \frac{K(B)^{d-i} \det(B\tilde{L})}{\det(B\tilde{L})_i} (5.1)
\]
for $1 \leq i \leq d$. Rewrite (5.1) as $1/(\|v_n\|)^{d-i}$ times

$$
\frac{1}{kd-i} \sum_{B \in \text{Gr}(\mathbb{Z}^{n-1},d)} \frac{\det(B\bar{L})}{\det(BL)_i} \left( \frac{H^2}{\det(BL)^2} - k^2 \right)^{d-i/2} f_{H/k}(B\bar{L}),
$$

which we simplify and bound from above by

$$(H/k)^{d-i} \sum_{B \in \text{Gr}(\tilde{L},d)} \frac{f_{H/k}(B)}{(\det B)^{d-i-1} \det B_i}. \tag{5.2}$$

Our analysis of (5.2) depends on the “skewness” of $B$ and $\tilde{L}$. We will first explain how to deal with (5.2) in case all $\lambda_i(\tilde{L})$ is of size $(H/k)^{1/d}$—i.e. $\tilde{L}$ is not too skewed—and then work out the general case.

In addition, for the rest of this section, we assume $k = 1$ for simplicity. To restore the general case, one could simply replace $H$ by $H/k$.

5.1 When $\tilde{L}$ is “not skewed”

Assume $\lambda_{n-1}(\tilde{L}) \leq 2^{n-1} H^{1/d}$. For each $0 \leq d' \leq d$, consider the restriction of the sum (5.2) to those $B \in \text{Gr}(\tilde{L}, d)$ for which $d'$ is the lowest number such that

$$
\lambda_{d'}(B) \leq 2^{d'} H^{1/d} \text{ and } \lambda_{d'+1}(B) - \lambda_{d'}(B) > 2^{d'} H^{1/d} \tag{5.3}
$$

(in fact, the former inequality follows from the latter and the minimality of $d'$), where we interpret $\lambda_0 = 0$ and $\lambda_{d+1} = \infty$. Such a sum is then bounded by a constant times

$$
H^{(d'+1)(1-i/d)} \sum_{B \in \text{Gr}(\tilde{L},d)} \frac{f_H(B)}{(\det B_{d'})^{d-i}}, \tag{5.4}
$$

where the sum is over all $B$ satisfying (5.3), since it follows from the Minkowski’s second and (5.3) that

$$
\frac{H^{d-i}}{(\det B)^{d-i-1} \det B_i} \ll \frac{H^{d-i+(d'-i)/d}}{(\det B)^{d-i-1} \det B_{d'}} \ll \frac{H^{d-i+(d'-i)/d-(d-d')(d-i-1)/d}}{(\det B_{d'})^{d-i}} \ll \frac{H^{(d'+1)(1-i/d)}}{(\det B_{d'})^{d-i}}.
$$

The idea for bounding (5.4) is that, because we are assuming $\lambda_{n-1}(\tilde{L}) \leq 2^{n-1} H^{1/d}$, we can proceed as in Section 9 of Schmidt [12]. The lemma below is a refinement of Lemma 6 of [12], so as to make explicit the dependence on the successive minima of $\tilde{L}$ and $B$. 
Lemma 5.1 Let \( \mathcal{L} \) be an \( m(= n - 1 \text{ in our context}) \) dimensional lattice. Fix a \( \mathcal{B}' \in \text{Gr}(\mathcal{L}, d') \), and let \( j = d - d' \). Then the number of \( \mathcal{B} \in \text{Gr}(\mathcal{L}, d) \) such that \( \mathcal{B}_{d-j} = \mathcal{B}' \), \( \lambda_{d'+1}(\mathcal{B}) \gg \lambda_m(\mathcal{L}) \) and \( \det \mathcal{B} \leq H \) is

\[
\ll \left( \frac{\det \mathcal{B}'}{\det \mathcal{L}} \right)^j \left( \frac{H}{\det \mathcal{B}'} \right)^{m-d'},
\]

where the implicit constant here depends only on \( n \) and the implied constants on the bound relating \( \lambda_m(\mathcal{L}) \) and \( \lambda_{d'+1}(\mathcal{B}) \).

Proof We may assume \( \det \mathcal{B}' \ll H^{d'/d} \), because by Minkowski’s second

\[
\det \mathcal{B}' \sim \lambda_1(\mathcal{B}') \cdots \lambda_{d'}(\mathcal{B}') \leq (\lambda_1(\mathcal{B}) \cdots \lambda_d(\mathcal{B}))^{d'/d} \ll H^{d'/d}.
\]

We proceed by induction on \( j \). Suppose first that \( j = 1 \). Let \( \pi : \text{span}_\mathbb{R}(\mathcal{L}) \to \text{span}_\mathbb{R}(\mathcal{L}) \) be the orthogonal projection onto the orthogonal complement of \( \text{span}_\mathbb{R}(\mathcal{B}') \). Then \( \pi(\mathcal{L}) \) (\( \cong \mathcal{L}/\mathcal{B}' \)) is a rank \( m - d' \) lattice of determinant \( \det \mathcal{L}/\det \mathcal{B}' \), and \( \pi(\mathcal{B}) \) is a 1-dimensional primitive sublattice spanned by a vector whose length is \( \det \mathcal{B}/\det \mathcal{B}' \). Therefore, the number of \( \mathcal{B} \) is bounded by the number of primitive vectors of \( \pi(\mathcal{L}) \) of length \( \leq H/\det \mathcal{B}' \).

If \( \mathcal{F} \) is a fundamental domain of \( \mathcal{L} \), then \( \pi(\mathcal{F}) \) is a fundamental domain of \( \pi(\mathcal{L}) \). Since we can choose an \( \mathcal{F} \) of diameter \( \lambda_1(\mathcal{L}) + \cdots + \lambda_m(\mathcal{L}) \leq m \lambda_m(\mathcal{L}) \) and \( \pi \) is a contraction, \( \pi(\mathcal{F}) \) has diameter \( \leq m \lambda_m(\mathcal{L}) \). So the number of vectors of \( \pi(\mathcal{L}) \) of length \( \leq H/\det \mathcal{B}' \) is bounded by a constant times

\[
\frac{\det \mathcal{B}'}{\det \mathcal{L}} \cdot \left( \frac{H}{\det \mathcal{B}'} + m \lambda_m(\mathcal{L}) \right)^{m-d'} \sim \frac{\det \mathcal{B}'}{\det \mathcal{L}} \cdot \left( \frac{H}{\det \mathcal{B}'} \right)^{m-d'}.
\]

Here we used the fact that \( H/\det \mathcal{B}' \geq \det \mathcal{B}/\det \mathcal{B}' \geq \lambda_{d'+1}(\mathcal{B}) \gg \lambda_m(\mathcal{L}). \)

For a general \( j \), by inductive hypothesis what we need to estimate is

\[
\sum \left( \frac{\det \mathcal{C}}{\det \mathcal{L}} \right)^{j - 1} \left( \frac{H}{\det \mathcal{C}} \right)^{m - 1 - d'} \quad (5.6)
\]

where the sum is over all \( \mathcal{C} \in \text{Gr}(\mathcal{L}, d'+1) \) such that \( \mathcal{C}_{d'} = \mathcal{B}' \) and \( \lambda_{d'+1}(\mathcal{C}) \gg \lambda_m(\mathcal{L}) \). In addition, \( \mathcal{C} \) must satisfy \( \det \mathcal{C} \ll \det \mathcal{B}'(H/\det \mathcal{B}')^{1/j} =: h \) say, since \( \lambda_{d'+1}(\mathcal{B}) \ll (H/\det \mathcal{B}')^{1/j} \).

From the (proof of) case \( j = 1 \), the number of \( \mathcal{C} \) with \( \mathcal{C}_{d'} = \mathcal{B}' \), \( \lambda_{d'+1}(\mathcal{C}) \gg \lambda_m(\mathcal{L}) \), and \( \det \mathcal{C} \leq t \) is

\[
\ll \frac{t^{m-d'}}{\det \mathcal{L} \det \mathcal{B}'^{m-1-d'}} \quad \text{if} \quad \frac{t}{\det \mathcal{B}'} \geq (\text{const}) \cdot \lambda_m(\mathcal{L}),
\]

\[
\ll \frac{\det \mathcal{B}'}{\det \mathcal{L}} \cdot \lambda_m(\mathcal{L})^{m-d'} \quad \text{otherwise}.
\]
But \( t \geq \det C \sim \det B' \cdot \lambda_{d'+1}(C) \gg \det B' \cdot \lambda_m(\tilde{L}) \), so we may disregard the latter possibility.

Therefore, we can apply Lemma 4.1, the Riemann-Stieltjes argument in the previous section, and deduce that (5.6) is bounded by a constant times

\[
\int_0^h \frac{t^{m-d'} \cdot t^{-m+1+d'+j}}{\det \tilde{L} \det B' \cdot \lambda_{d'+1}(C)} H^{m-d' - 1} \frac{\det B'}{(\det \tilde{L})^{j-1}} dt,
\]

which turns out to be equal to a constant times

\[
\left( \frac{\det B'}{\det \tilde{L}} \right)^{j} \left( \frac{H}{\det B'} \right)^{m-d'},
\]

as desired. \(\square\)

We proceed to estimating (5.4). Thanks to Lemma 5.1, for some constant \( C > 0 \) depending only on \( n \) such that \( \det B' < CH^{d'/d} \) (which exists by Minkowski’s second), we can bound it by

\[
H^{(d'+1)(1-i/d)} \sum_{B' \in \text{Gr}(\tilde{L}, d')} f_{CH^{d'/d}}(B') \left( \frac{\det B'}{\det \tilde{L}} \right)^{d-d'} \left( \frac{H}{\det B'} \right)^{n-1-d'},
\]

This can be handled again as in the previous section using Lemma 4.1, yielding \( |E_{n-1,d'}| + 1 \) terms of \( H \)-degree at most \( n - i/d \) satisfying all the miscellaneous conditions such as the scaling invariance.

### 5.2 The skewed case

Now assume that \( 0 \leq l < n - 1 \) is the lowest number such that

\[
\lambda_l(\tilde{L}) \leq 2^l H^{1/d} \quad \text{and} \quad \lambda_{l+1}(\tilde{L}) - \lambda_l(\tilde{L}) > 2^l H^{1/d}.
\]

As earlier, we again restrict the sum (5.2) to those \( B \in \text{Gr}(\tilde{L}, d) \) for which \( 0 \leq d' \leq d \) is the lowest number such that

\[
\lambda_{d'}(B) \leq 2^{d'} H^{1/d} \quad \text{and} \quad \lambda_{d'+1}(B) - \lambda_{d'}(B) > 2^{d'} H^{1/d}.
\]

Then we must have \( d' \leq l \) and \( B_d' \subseteq \tilde{L}_l \). There is a decomposition

\[
\tilde{L} = \tilde{L}_l \oplus \mathcal{M},
\]

\(\copyright\) Springer
where $\mathcal{M}$ is an $n-1-l$ dimensional lattice chosen as follows: take a reduced basis $\{x_1, \ldots, x_{n-1}\}$ of $\mathcal{L}$ such that $\|x_i\| \sim \lambda_i(\mathcal{L})$ and spanning $\{x_1, \ldots, x_l\} = \mathcal{L}_l$. Then we let $\mathcal{M} = \text{span}\{x_{l+1}, \ldots, x_{n-1}\}$. Also, let $\mathcal{M}$ to be the orthogonal projection of $\mathcal{M}$ onto $\text{span}_{\mathbb{R}}(\mathcal{L}_l) \perp \text{span}_{\mathbb{R}}(\mathcal{L})$. An important fact we will use later is that $\lambda_1(\mathcal{M}) \gg H^{1/d}$ by construction.

We further restrict (5.2) to those $B$ for which $\text{rk} B \cap \mathcal{L}_l = r$ for $r \in \{d', \ldots, \min(l, d)\}$, and call $B_r = B \cap \mathcal{L}_l$. Note that $(B_r)_{d'} = B_{d'}$. We also let $\mathcal{A} \subseteq \mathcal{M}$ be the projection of $B$ onto $\mathcal{M}$. Clearly $\det B = \det B_r \det \mathcal{A}$, and since $\det \mathcal{A} \gg H^{(d-r)/d}$ we have $\det B_r \ll H^{r/d}$.

Our considerations so far lead us to bound the restriction of (5.2) by, for some constant $C > 0$,

$$H^{(d'+1)(1-i/d)} \sum_{B \in \text{Gr}(\mathcal{L}, d)} \frac{f_H(B)}{(\det B_{d'})^{d-i}} \ll H^{(d'+1)(1-i/d)} \sum_{B \in \text{Gr}(\mathcal{L}, r)} \frac{f_{CH}^{r/d}(B_{r})}{(\det(B_r)_{d'})^{d-i}} \sum_{\mathcal{A} \in \text{Gr}(\mathcal{M}, d-r)} \frac{f_H}{\det B_r}(\mathcal{A}).$$

Using the induction hypothesis on our main theorem, and the fact that $\lambda_1(\mathcal{M}) \gg H^{1/d}$, we can rewrite the inner sum so that this becomes

$$\ll H^{(d'+1)(1-i/d)} \sum_{B \in \text{Gr}(\mathcal{L}, r)} \frac{f_{CH}^{r/d}(B_{r})}{(\det(B_r)_{d'})^{d-i}} \sum_{\gamma \in \{\gamma_j: j \in E_{n-1-l}, d-r\}} \left( \frac{H}{\det B_r} \right)^\gamma H^{-\gamma(d-r)/d}.$$ 

Let us look at one $\gamma$ at a time, and consider

$$H^{\frac{d'}{d} \gamma + (d'+1)(1-\frac{i}{d})} \sum_{B \in \text{Gr}(\mathcal{L}, r)} \frac{f_{CH}^{r/d}(B_{r})}{(\det(B_r)_{d'})^{d-i}} \sum_{\gamma \in \{\gamma_j: j \in E_{n-1-l}, d-r\}} \left( \frac{H}{\det B_r} \right)^\gamma H^{-\gamma(d-r)/d}.$$ 

By Lemma 5.1 as in the previous “not skewed” section, we obtain that this is

$$= H^{\frac{d'}{d} \gamma + (d'+1)(1-\frac{i}{d})} \sum_{B \in \text{Gr}(\mathcal{L}, d') \cap \text{Gr}(\mathcal{L}, r)} \frac{f_{CH}^{r/d}(B')}{(\det B')^{d-i+\gamma}} \sum_{B \in \text{Gr}(\mathcal{L}, r)} \frac{f_{CH}^{r/d}(B_{r})}{(\det(B_r)_{d'})^{d-i}}$$

$$\ll H^{\frac{d'}{d} \gamma + (d'+1)(1-\frac{i}{d})} \sum_{B \in \text{Gr}(\mathcal{L}, d') \cap \text{Gr}(\mathcal{L}, r)} \frac{f_{CH}^{r/d}(B')}{(\det B')^{d-i+\gamma}} \left( \frac{\det B'}{\det \mathcal{L}_l} \right)^{r-d'} \left( \frac{H^{r/d}}{\det B'} \right)^{l-d'}$$

$$= \frac{H^{\frac{d'}{d} \gamma + (d'-l)+(d'+1)(1-\frac{i}{d})}}{(\det \mathcal{L}_l)^{r-d'}} \sum_{B \in \text{Gr}(\mathcal{L}, d')} \frac{f_{CH}^{r/d}(B')}{(\det B')^{d-i+\gamma+r-l}}. \quad (5.8)$$

 Springer
Applying Lemma 4.1, it is seen that (5.8) may be bounded by at most $|E_{I,r}| + 1$ error terms. We need to make sure that the $H$-degree of those terms are strictly below $n$. Here we only discuss the terms of the highest degrees, as the rest can be dealt with in a similar fashion.

If $-d + i - \gamma + r \neq 0$, estimating the sum in (5.8) using Lemma 4.1 yields a term of $H$-degree $\frac{d'}{d}(-d + i - \gamma + r)$. Therefore, the $H$-degree of (5.8) equals

$$\frac{rl}{d} + 1 - \frac{i}{d},$$

which attains its maximum $n - i/d$ only if $r = d$ and $l = n - 1$. But recall that we are assuming $l < n - 1$.

If $-d + i - \gamma + r = 0$, the sum is of size $O(\log H)$, in which case we can say that, for a small $\eta > 0$, the $H$-degree is $\leq n - i/d - id'/d + \eta$ if $d' \neq 0$, and is $\leq n - 1 - i/d + \eta$ if $d' = 0$.

### 5.3 The number of the error terms

We summarize and estimate the maximum number of error terms arising from our estimate of (5.2) so far. If $\tilde{\mathcal{L}}$ is “not skewed,” then our estimate yielded $(d + 1)E_{n-1}$ error terms, where we write

$$E_{n-1} = \max_{0 \leq d' \leq n-1} |E_{n-1,d'}| + 1,$$

and we understand $E_{n-1,0}$ to be the empty set.

As for the skewed case, it is really $n - 1$ separate cases corresponding to the parameter $0 \leq l < n - 1$, and for each $l$ we obtained at most $(l + 1)E_{n-1-l}E_l$ error terms. Hence, regardless of $H$ and $d$, we are able to estimate (5.2) using at most

$$\sum_{l=0}^{n-1} (l + 1)E_{n-1-l}E_l$$

terms.

### 6 Summary, and a proof of Theorem 1.3

#### 6.1 A polynomial expression for $P(\mathcal{L}, d, H)$

Summing up all our work so far, we have that

$$P^1(\mathcal{L}, d, H) = \sum_{k=1}^{H/e} \prod_{\substack{p|k \mid p^d \parallel k}} p^{\alpha d} \left( 1 - \frac{1}{p^d} \right) \left( \frac{\zeta(n-d)}{(\det \mathcal{L})^d \zeta(n)} \right) H^n k^{-n}$$

$\copyright$ Springer
\[ + O \left( \sum_{j \in E^{(2)}} c_j H^{|j|} k^{-|j|} \right) \]  \hspace{1cm} (6.1)\]

where \( \varepsilon = \min_{B \in \text{Gr}(\mathbb{Z}^{n-1}, d)} \det(B\bar{L}) \), \( E^{(2)} \) is an index set of cardinality at most

\[ E_{n-1} + d \sum_{l=0}^{n-1} (l + 1) E_{n-1-l} E_l \]

(collecting all error terms from the previous two sections), and each \( c_j \) is a reciprocal of products of \( \lambda_i(\bar{L}) \)'s and \( \|v_n\| \) so that \( c_j H^{|j|} \) is scale-invariant. In this section, we will estimate the sum (6.1), and then make a choice of \( v_n \in \mathcal{L} \) so that the dependence on \( \lambda_i(\bar{L}) \)'s turns into dependence on \( \lambda_i(\mathcal{L}) \)'s. This will prove our main theorem.

We treat (6.1) one monomial at a time. The highest degree term contributes

\[ \frac{H/\varepsilon}{\sum_{k=1}^{H/\varepsilon} \prod_{p \mid k} p^{a_d} \left( 1 - \frac{1}{p^d} \right) \left( \frac{a(n, d)}{(\det\mathcal{L})^d} \frac{\xi(n)}{\xi(n-d)} H^n k^{-n} \right) } \]

The corresponding infinite series, by Lemma 3.7, equals

\[ \frac{a(n, d)}{(\det\mathcal{L})^d} H^n, \]

the desired main term. It remains to bound the tail, which we can, up to a constant factor, approximate as

\[ \frac{1}{(\det\mathcal{L})^d} \sum_{k \gg H/\varepsilon} H^n k^{d-n}, \]

which is of size

\[ \frac{H^{d+1} \varepsilon^{n-d-1}}{(\det\mathcal{L})^d}. \]

We need to show that \( \varepsilon^{n-d-1}/(\det\mathcal{L})^d \) is bounded by a reciprocal of a product of \( \lambda_i(\mathcal{L}) \)'s. Since \( \varepsilon \sim \prod_{i=1}^d \lambda_i(\bar{L}) \) and \( \lambda_i(\mathcal{L}) \leq \lambda_{i+1}(\mathcal{L}) \) (which can be seen by projecting a \((i+1)\)-dimensional subspace of \( \mathbb{R}^n \) onto the orthogonal complement of \( v_n \)), we have \( \varepsilon \ll \prod_{j=1}^d \lambda_{i+1}(\mathcal{L}) \), and thus \( \varepsilon^{n-d-1} \ll \prod_{i=1}^d \lambda_{i+1}(\mathcal{L})^{n-d-1} \). On the other hand, \( (\det\mathcal{L})^d \sim \prod_{j=1}^n \lambda_j(\mathcal{L})^d \), which contains the factor \( \prod_{j=d+2}^n \lambda_j(\mathcal{L}) \) \( d \) times. For any \( i \leq d + 1, \lambda_i(\mathcal{L})^{n-d-1} / \prod_{j=d+2}^n \lambda_j(\mathcal{L}) \leq 1 \), so \( \varepsilon^{n-d-1}/(\det\mathcal{L})^d \ll \prod_{j=d+1}^n \lambda_j(\mathcal{L})^{-d} \), as desired.
We return to other monomials in (6.1). For the indices \( j \) with \( \gamma_j > d + 1 \), the sum under consideration is

\[
c_j H^{\gamma_j} \sum_{k=1}^{H/\varepsilon} \prod_{p\mid k} p^{\alpha \parallel k \parallel} \left( 1 - \frac{1}{p^{\gamma_j}} \right) k^{-\gamma_j},
\]

which we can bound by the corresponding infinite series and apply Lemma 3.7, obtaining \( O(c_j H^{\gamma_j}) \). If \( \gamma_j < d + 1 \), the sum is of size

\[
c_j H^{\gamma_j} \sum_{k=1}^{H/\varepsilon} k^{d-\gamma_j} \sim \frac{c_j H^{d+1}}{\varepsilon^{d-\gamma_j+1}},
\]

and if \( \gamma_j = d + 1 \), it is

\[
c_j H^{\gamma_j} \sum_{k=1}^{H/\varepsilon} k^{-1} \sim c_j H^{\gamma_j} \log \frac{H}{\varepsilon} \ll \frac{c_j H^{\gamma_j + \eta}}{\varepsilon^\eta}
\]

for any \( \eta > 0 \). Hence, together with the expression (3.1) of \( P^2 \), we conclude that

\[
P(\mathcal{L}, d, H) = \frac{a(n, d)}{(\det \mathcal{L})^d} H^n + O \left( \sum_{j \in E_{n,d}} b_j H^{\gamma_j} \right)
\]

for some index set \( E_{n,d} \) of cardinality at most \( d n \sum_{l=0}^{n-1} E_{n-1-l} E_l \), and where each \( b_j \) is a product of reciprocals of \( \lambda_i(\mathcal{L})'s, \lambda_i(\bar{\mathcal{L}})'s, \) and \( \|v_n\| \), so that \( b_j H^{\gamma_j} \) is scale-invariant.

At this point, choose \( v_n \) to be one of the shortest nonzero vectors of \( \mathcal{L} \). Then the following lemma shows that we can replace \( \lambda_{i-1}(\bar{\mathcal{L}}) \) by \( \lambda_i(\mathcal{L}) \) for each \( i \), so that \( b_j \) would depend only on \( \mathcal{L} \).

**Lemma 6.1** Recall that \( \bar{\mathcal{L}} \) is the orthogonal projection of \( \mathcal{L} \) onto the complement of a vector \( v_n \in \mathcal{L} \). If we choose \( v_n \) to be a shortest nonzero vector of \( \mathcal{L} \), then \( \lambda_{i-1}(\bar{\mathcal{L}}) \sim \lambda_i(\mathcal{L}) \) for all \( i = 2, \ldots, n \).

**Proof** Let \( \{w_1, \ldots, w_n\} \) be a reduced basis of \( \mathcal{L} \) containing \( v_n = w_1 \). Then, writing \( \tilde{w}_i \) for the projection of \( w_i \) to the complement of \( v_n \), \( \{\tilde{w}_2, \ldots, \tilde{w}_n\} \) is a reduced basis of \( \bar{\mathcal{L}} \). Therefore, \( \|w_i\| \sim \lambda_i(\mathcal{L}) \) and \( \|\tilde{w}_i\| \sim \lambda_i(\bar{\mathcal{L}}) \).

On the other hand, by the definition of a reduced basis, \( \|\tilde{w}_i\|^2 = \|w_i\|^2 - \mu^2 \|w_1\|^2 \) for some \( \|\mu\| \leq 1/2 \). This immediately implies \( \|\tilde{w}_i\| \leq \|w_i\| \), and also, since \( \|w_1\| \leq \|w_i\| \), we have \( \|w_i\| \gg \|\tilde{w}_i\| \), completing the proof. \( \square \)
6.2 The number of the error terms

Let us give a quick, crude estimate of \( E_n = \max_{0 \leq d \leq n} |E_{n,d}| + 1 \). From the above discussion, we have

\[
E_n \leq n^2 \sum_{l=0}^{n-1} E_{n-1,l}.
\]

Recall that \( E_0 = 1 \) by definition. Also, it is clear from Section 2.2 that \( E_1 = 2 \), \( E_2 = 3 \), \( E_3 = 4 \). We claim in general that \( E_n \leq n^{3n} \) for \( n \geq 2 \). Indeed, the base case is obvious, and assuming the truth for the \( n-1 \) case, it follows from the above inequality that

\[
E_n \leq n^2 \cdot n \cdot n^{3(n-1)} = n^{3n}.
\]

6.3 The primary error term, \( d \leq n/2 \)

Finally, we provide an estimate on the primary error term of \( P(\mathcal{L}, d, H) \), again assuming \( \|v_n\| = \lambda_1(\mathcal{L}) \). We temporarily assume \( d \leq n/2 \), and argue the cases \( d > n/2 \) by duality. Tracing back our estimates so far, there are two candidates for the primary error term: one is from the estimate of the “main part” (4.1), which contributes

\[
O\left( \frac{b(\tilde{\mathcal{L}})}{\|v_n\|^d} H^{n-b(n-1,d)} \right) \tag{6.2}
\]

where \( b(\tilde{\mathcal{L}}) = b_j(\tilde{\mathcal{L}}) \) for \( j \in E_{n-1,d} \) corresponding to the leading error term, and the other is from the estimate of the “error part” (5.1) in case \( i = 1 \), which contributes

\[
O\left( \frac{H^{n-b(n,d)}}{(\det \mathcal{L})^{d-1} \det \tilde{\mathcal{L}}} \right),
\]

but by rewriting everything in terms of \( \lambda_i(\mathcal{L}) \)’s with help of Lemma 6.1, we find that this is bounded by

\[
O\left( \frac{H^{n-b(n,d)}}{(\det \mathcal{L})^{d-b(n,d)} (\det \mathcal{L}_{n-d})^{b(n,d)}} \right). \tag{6.3}
\]

The reason we use this slightly inferior bound is that this possesses a convenient symmetry under duality, as we will see below.

We claim by induction that the main error term has degree \( n - b(n,d) \), and that (6.2) is no greater than (6.3). In the base case \( n = 4, d = 2 \), this is clear. In general, if \( d = n/2 \), (6.2) is of degree strictly less than \( n - b(n,d) \), and we are done. If \( d < n/2 \),
then by the fact that $\|v_n\| = \lambda_1(\mathcal{L})$ and Lemma 6.1,

$$\|v_n\|^d(\det \bar{\mathcal{L}})^{d-1/d}(\det \bar{\mathcal{L}}_{n-d})^{1/d} \sim (\det \mathcal{L})^{d-1/d}(\det \mathcal{L}_{n-d})^{1/d},$$

which shows that (6.2) has the same size as (6.3), by the inductive hypothesis on $b(\bar{\mathcal{L}})$. This proves the claim.

### 6.4 The primary error term, $d > n/2$

Write $d' = n - d$ for short. By the duality theorem (2.2), $P(\mathcal{L}, d, H) = P(\mathcal{L}^P, d', H')$, where $H' = H / \det \mathcal{L}$. Observe that both has the same main term, that is,

$$a(n, d) \frac{H^n}{(\det \mathcal{L})^d} = a(n, d') \frac{H'^n}{(\det \mathcal{L}^P)^{d'}}.$$

Moreover, from the previous section, $P(\mathcal{L}^P, d', H')$ has the leading error term of size

$$H'^{n-b(n,d')} \frac{H^n}{(\det \mathcal{L}^P)^{d'-b(n,d')} (\det (\mathcal{L}^P)_{n-d'})^{b(n,d')}},$$

which is equal to

$$\frac{H^{n-b(n,d)}}{(\det \mathcal{L})^{n-b(n,d)} (\det \mathcal{L}^P)^{n-d-b(n,d)} (\det (\mathcal{L}^P)_{d})^{b(n,d)}} \frac{H'^{n-b(n,d')}}{(\det \mathcal{L}^P)^{d'-b(n,d')} (\det (\mathcal{L}^P)_{d'})^{b(n,d')}}$$

$$\ll \frac{H^{n-b(n,d)}}{(\det \mathcal{L})^{d-b(n,d)} (\det \mathcal{L}_{n-d})^{b(n,d)}},$$

as desired.

### 7 Proofs of the variants

#### 7.1 Formula for $N(\mathcal{L}, d, H)$

An asymptotic formula on $N(\mathcal{L}, d, H)$ can be derived easily from that of $P(\mathcal{L}, d, H)$ by a standard Möbius inversion, as in Schmidt ([12, Sections 3,4,10]). As in [12], define $\sigma_d(m)$ inductively by

$$\sigma_1(m) = 1,$$

$$\sigma_d(m) = \sum_{r|n} r^{d-1} \sigma_{d-1}(m/r).$$
It is shown in [12, Section 3] that $\sigma_d(m)$ equals the number of index $m$ sublattices of a rank $d$ lattice, and that

$$\sigma_d(m) \ll (m \log \log m)^{d-1},$$

$$\sum_{m=1}^{\infty} \sigma_d(m)/m^n = \prod_{i=1}^{d} \zeta(n+1-i) \quad (7.1)$$

for $d \leq n-1$. From the latter it follows that

$$N(\mathcal{L}, d, H) = \sum_{m=1}^{H/\varepsilon} P(\mathcal{L}, d, H/m)\sigma_d(m)$$

$$= \frac{a(n,d)}{(\det \mathcal{L})^d} \sum_{m=1}^{H/\varepsilon} (H/m)^n\sigma_d(m) + O \left( \sum_{j \in E_{n,d}} \sum_{m=1}^{H/\varepsilon} b_j(\mathcal{L})(H/m)^{\gamma_j}\sigma_d(m) \right),$$

where $\varepsilon := \min_{B \in \text{Gr}(\mathcal{L},d)} \det B$. We handle each sum over $m$ one at a time. For the main term, we have

$$\sum_{m=1}^{H/\varepsilon} (H/m)^n\sigma_d(m) = \sum_{m=1}^{\infty} (H/m)^n\sigma_d(m) - \sum_{m>H/\varepsilon} (H/m)^n\sigma_d(m).$$

On the right-hand side, the first sum is $H^n \prod_{i=1}^{d} \zeta(n+1-i)$ by (7.1), which is exactly what we need. The second sum is bounded by a constant times

$$\sum_{m>H/\varepsilon} m^{d-n-1+\eta} H^n \sim \frac{H^{d+\eta}}{\varepsilon^{d-n+\eta}}$$

for any $\eta > 0$.

In the error term, for those $j \in E_{n,d}$ with $\gamma_j > d$ we can replace the sum $\sum_{m=1}^{H/\varepsilon}$ by the infinite sum $\sum_{m=1}^{\infty}$, and apply (7.1). For those with $\gamma_j \leq d$, we see that

$$\sum_{m=1}^{H/\varepsilon} \sigma_d(m)m^{-\gamma_j} \ll \sum_{m=1}^{H/\varepsilon} m^{d-1-\gamma_j+\eta} \sim \left( \frac{H}{\varepsilon} \right)^{d-\gamma_j+\eta}$$

for any $\eta > 0$. If $d < n-1$, $\eta$ can be set small enough, so that the secondary term has $H$-degree $n - b(n,d)$. If $d = n-1$, the secondary term has degree $n - 1 + \eta$.

The required properties of the coefficients $b'_j(\mathcal{L})$ can be checked straightforwardly, so we omit the proof.
Remark One may wonder what the formula for $N(\mathcal{L}, n, H)$ would be. In this case, the skewness of $\mathcal{L}$ induces no subtlety at all, and simply

\[
N(\mathcal{L}, n, H) = c \cdot \left( \frac{H}{\det \mathcal{L}} \right)^n + O \left( \left( \frac{H}{\det \mathcal{L}} \right)^{n-1+\eta} \right)
\]

for any $\eta > 0$. The proof is identical to the argument in Section 3 of [12]; indeed, observe that $N(\mathcal{L}, n, H) = N(\mathbb{Z}^n, n, H)$ for any full-rank $\mathcal{L}$ of covolume 1.

### 7.2 Formula for $P_S(\mathcal{L}, d, H)$

Let $\mathcal{S} \subseteq \mathcal{L}$ be a sublattice of rank $e \leq n - d$. By choosing the basis $\{v_1, \ldots, v_n\}$ of $\mathcal{L}$ so that $\{v_{n-e+1}, \ldots, v_n\}$ is a basis of $\mathcal{S}$, and proceeding analogously as in Section 3 with $\mathcal{L}/\mathcal{S}$ instead of $\mathcal{L}$, we obtain an estimate of $P_S(\mathcal{L}, d, H)$ analogous to that of $P(\mathcal{L}, d, H)$ in (1.2), with the coefficients $b_j$ being a product of reciprocals of $\lambda_i(\mathcal{L})$ and $\lambda_i(\mathcal{L}/\mathcal{S})$ (here we identify $\mathcal{L}/\mathcal{S}$ with the projection of $\mathcal{L}$ onto the orthogonal complement of span$_\mathbb{R}(\mathcal{S})$ in $\mathbb{R}^n$). However, the reciprocal of $\lambda_i(\mathcal{L}/\mathcal{S})$ could be arbitrarily large, which may cause difficulties in some applications of Theorem 1.3. For instance, suppose one wants to compute

\[
\sum_{A,B \in \text{Gr}(\mathcal{L}, d), A \cap \mathcal{B} = [0]} f_{H_1}(A) f_{H_2}(B) = \sum_A f_{H_1}(A) \sum_{B, A \cap \mathcal{B} = [0]} f_{H_2}(B).
\]

Here one is eventually led to sum the multiples of the reciprocals of $\lambda_i(\mathcal{L}/A)$ over sublattices $A$ of height bounded by $H_1$. It seems to be a nontrivial task to show that such a sum is asymptotically small.

Fortunately, with minor modifications to our proof of Theorem 1.3, it is possible to provide a formula for $P_S(\mathcal{L}, d, H)$ independent of $\mathcal{S}$, avoiding the above complication altogether. In this section, we explain what modifications are to be made.

Consider first the base cases $d = 1$ or $n - 1$. If $d = 1$, $P_S(\mathcal{L}, 1, H) = P(\mathcal{L}, 1, H) - P(\mathcal{S}, 1, H)$, and bounding the contribution from $P(\mathcal{S}, 1, H)$ in terms of $\mathcal{L}$ using $\lambda_i(\mathcal{S}) \geq \lambda_i(\mathcal{L})$ (because $\mathcal{S} \subseteq \mathcal{L}$), we obtain the same type of estimate as in (2.1). In case $d = n - 1$, we must have $\text{rk} \mathcal{S} = 1$, and thus for $B \in \text{Gr}(\mathcal{L}, n - 1)$, $B \cap \mathcal{S} = \{0\}$ if and only if $B^\perp \cap \mathcal{S}^\perp = \{0\}$; hence the proof follows from the $d = 1$ case and the duality theorem.

For other values of $d$, we proceed by induction on $n$, and split $P_S = P_S^1 + P_S^2$ as in Section 3. For $P_S^2$, we simply bound it by $P^2$. As for $P_S^1$, observe that, analogously to (3.3), we can write

\[
P_S^1(\mathcal{L}, d, H) = \sum_{B \in \Gamma \setminus \text{Mat}_{d \times (n-1)}(\mathbb{Z})} \sum_{k \geq 1} \sum_{h \in \mathbb{Z}} \sum_{b \in \mathbb{Z}^d \text{ prim.}} f_H((hB; b) \mathcal{L}),
\]

where $\mathcal{L}(M)$ denotes the lattice spanned by the row vectors of $M$. The idea is that the main contribution of the above sum comes from those $B$ with $\mathcal{L}(B \mathcal{L}) \cap \mathcal{S} = \{0\}$, where...
\( \tilde{S} \) is the projection of \( S \) onto \( \tilde{L} \). Since \( \mathcal{L}(B\tilde{L}) \cap \tilde{S} = \{0\} \) implies \( \mathcal{L}((hB; b)L) \cap S = \{0\} \), we may write

\[
P_S^1 = P_S^{1,1} + O(P_S^{1,2}),
\]

where

\[
P_S^{1,1} = \sum_{B \in \Gamma \setminus \text{Mat}_{d \times (n-1)}(\mathbb{Z})} \sum_{k \geq 1} \sum_{h \in \mathbb{Z}^d} \sum_{b \in \mathbb{Z}^d, (hB; b) \text{ prim.}} f_H \left( (hB; b) L \right),
\]

\[
P_S^{1,2} = \sum_{B \in \Gamma \setminus \text{Mat}_{d \times (n-1)}(\mathbb{Z})} \sum_{k \geq 1} \sum_{h \in \mathbb{Z}^d} \sum_{b \in \mathbb{Z}^d, (hB; b) \text{ prim.}} f_H \left( (hB; b) L \right).
\]

But since \( P_S^{1,1} + P_S^{1,2} = P_1 \), it also holds that

\[
P_S^1 = P_1 + O(P_S^{1,2}).
\]

Therefore all we need to show is that \( P_S^{1,2} \) is small, or equivalently, that \( P_S^{1,1} \) is close to \( P_1 \).

Estimating \( P_S^{1,1} \) amounts to considering an analogous expression to (3.5) where the sum over \( B \) is further restricted to those for which \( B\tilde{L} \cap \tilde{S} = \{0\} \). With the same restriction added to all the subsequent computations, all of the arguments in Section 4 goes through, including Lemma 4.1, since by the induction hypothesis \( P \) and \( P_S \) satisfy the same asymptotics on lattices of rank \( \leq n - 1 \). As for the error terms of \( P_S^{1,1} \), we may simply bound them by those of \( P_1 \), namely (5.2) for \( i = 1, \ldots, d \), and so there are no changes to make. This shows that \( P_1 \) and \( P_S^{1,1} \) satisfy the same asymptotics, and hence that \( P_S^{1,2} \) is bounded by terms of leading degree strictly less than \( n \).

To count the number of error terms, recall that \( P_S = P_1 + O(P_2) + O(P_S^{1,2}) \). \( P_1 + O(P_2) \) have \( < n^{3n} \) error terms, and since \( P_S^{1,2} = P_1 - P_S^{1,1} \), it has at most \( 2n^{3n} \) error terms. Thus the number of error terms in our formula for \( P_S \) is no more than \( 3n^{3n} \).

It remains to determine the main error term. If \( d \leq n/2 \), the argument of Sect. 6.3 carries over, showing that it is the same as in Theorem 1.3. If \( d > n/2 \), extend \( S \) to a sublattice \( S' \subseteq \mathcal{L} \) of rank precisely \( n - d \). Then

\[
P_{S'}(\mathcal{L}, d, H) \leq P_S(\mathcal{L}, d, H) \leq P(\mathcal{L}, d, H)
\]

on the one hand, and on the other hand,

\[
P_{S'}(\mathcal{L}, d, H) = P_{S',\perp}(\mathcal{L}^\perp, n - d, H/\det \mathcal{L}).
\]
Now we can argue as in Sect. 6.4, using duality, to show that $P_{S'}$ has the leading error term of the same size, and therefore so does $P_S$. This completes the proof of Theorem 1.5; for $N_S$, one may proceed as in the last section.

7.3 Flag varieties of type $(e, d)$

Let $L \subseteq \mathbb{R}^n$ be a lattice, and let $1 \leq e < d < n$. Our goal is to estimate the sum

$$\sum_{\mathcal{W} \in \text{Gr}(L, d)} P(\mathcal{W}, e, (H/(\det \mathcal{W})^{n-e})^{1/d}) = \sum_{\mathcal{W} \in \text{Gr}(L, d)} \frac{a(d, e)H}{(\det \mathcal{W})^n} + O\left(\frac{H^{1-b/d}}{(\det \mathcal{W})^{n-b(d+e-e)/d}(\det \mathcal{W}_{d-e})^b}\right). \tag{7.2}$$

Here $b = b(d, e)$. Note that there is also the constraint

$$(\det \mathcal{W})^{n-e}(\det \mathcal{W}_e)^d \leq H \tag{7.3}$$

coming from the definition of height.

Estimating the sum over the main term is very similar to the computation in Section 4, so we will be brief. The largest term of (7.2), obtained by applying Lemma 4.1 to $a(d, e)H/(\det \mathcal{W})^n$, comes from the integral

$$\int_{\varepsilon_d}^{(H/\varepsilon_e)^{(n-e)/d}} \frac{a(n, d)t^n}{(\det L)^d} \left(\frac{a(d, e)H}{t^{n-e}}\right) \frac{dt}{t} = \frac{a(n, d)a(d, e)nH}{(\det L)^d} \int_{\varepsilon_d}^{(H/\varepsilon_e)^{(n-e)/d}} \frac{dt}{t}$$

where $\varepsilon_e = \min\{\det B : B \subseteq L, \dim B = e\}$ and likewise for $\varepsilon_d$. The smaller terms can be computed similarly, and it turns out the largest error term is of $H$-degree 1, and the second largest is of $H$-degree $1 - b(n, d)/(n - e)$, both coming from the leading error term of $P(L, d, (H/\varepsilon_e)^{(1/(n-e)})$. One obtains a total of $2|E_{n,d}| < 2E_n$ error terms from here.

The harder part of (7.2) is the sum over the error term, namely

$$\sum_{\mathcal{W} \in \text{Gr}(L, d)} \frac{H^{1-b/d}}{(\det \mathcal{W})^{n-b(d+e-e)/d}(\det \mathcal{W}_{d-e})^b}. \tag{7.4}$$

To bound this, we employ our method in Section 5 above. In order to avoid repetitive and unenlightening computations, we only show how to compute the first two largest $H$-degree terms, and suppress the $\lambda_i(L)$ factors from the expressions for the error terms.
As in Section 5, for each $0 \leq d' \leq d$, we restrict the sum in (7.4) to those $\mathcal{W}$ for which $d'$ is the smallest number such that

$$\lambda_{d'}(\mathcal{W}) \leq 2^{d'} H^{1/n_d} \text{ and } \lambda_{d'+1}(\mathcal{W}) - \lambda_{d'}(\mathcal{W}) > 2^{d'} H^{1/n_d}.$$ 

Then we can bound (7.4) by

$$H^{1 - \frac{b}{n} + b/n_d (d' - d + e)} \sum_{\mathcal{W}' \in \text{Gr}(\mathcal{L},d')} \frac{1}{(\det \mathcal{W'})^b} \sum_{\mathcal{W} \in \text{Gr}(\mathcal{L},d') \mathcal{W}' = \mathcal{W}} \frac{1}{(\det \mathcal{W})^{n-b(n+d-e)/d}}.$$  

(7.5)

In order to work on the inner sum, we first determine the range of $\det \mathcal{W}$. By Minkowski’s second, we have $(\det \mathcal{W}_{d'})^{d'/d} \ll \det \mathcal{W}$. On the other hand, again by Minkowski’s second we have $\det \mathcal{W}_{d'} H^{(e-d')/n_d} \leq \det \mathcal{W}_e$, so (7.3) implies

$$(\det \mathcal{W})^{n-e} (\det \mathcal{W}_{d'})^d H^{(e-d')/n} \ll H \Rightarrow \det \mathcal{W} \ll H^{n-e+d'/n_d} (\det \mathcal{W}_{d'})^{-d/n-e} \Rightarrow \det \mathcal{W}_{d'} \ll H^{d'/d_n}.$$  

(7.6)

In the last line, we also used $\det \mathcal{W}_{d'} \ll (\det \mathcal{W})^{d'/d}$.  

If $d' = 0$, the outer sum of (7.5) is vacuous, and the inner sum can be computed as in the main term estimate above, yielding $O(H + H^{1-b(n+d-e)/n_d} + H^{1-b(n,d)/n})$ up to lower $H$-degree terms. Similarly, we obtain the same bound in case $d' = d$. Each of these cases add at most $2E_n$ error terms to our estimate.

So assume $1 \leq d' \leq d - 1$. We will apply Lemma 5.1. To do so, it is required that $\lambda_{d'+1}(\mathcal{W}) \gg \lambda_n(\mathcal{L})$, which is true provided $\lambda_n(\mathcal{L}) \ll H^{1/n_d}$. Thus, assuming $H$ is sufficiently large, the inner sum of (7.5) is bounded by a constant times

$$\frac{1}{(\det \mathcal{L})^{d-d'}(\det \mathcal{W'})^{n-d}} \int_{(\det \mathcal{W'})^{d'/d}} H^{\frac{n-e+d'}{(n-e)n_d}} (\det \mathcal{W})^{-\frac{d}{n-e}} \frac{x^{n-d' - 1}}{x^{n-b(n+d-e)/d}} d^d x.$$ 

We divide into cases according to whether $b(n + d - e)/d - d' < 0$ or not:

(i) If $b(n + d - e)/d - d' < 0$, then (7.5) is

$$\ll H^{1 - \frac{b}{n} + b/n_d (d' - d + e)} \sum_{\mathcal{W}' \in \text{Gr}(\mathcal{L},d')} \frac{1}{(\det \mathcal{W'})^{n-d+b}} \int_{(\det \mathcal{W'})^{d'/d}} H^{\frac{n-e+d'}{(n-e)n_d}} (\det \mathcal{W})^{-\frac{d}{n-e}} \frac{x^{n-d' - 1}}{x^{n-b(n+d-e)/d}} d^d x$$

$$\ll H^{1 - \frac{b}{n} + b/n_d (d' - d + e)} \sum_{\mathcal{W}' \in \text{Gr}(\mathcal{L},d')} \frac{1}{(\det \mathcal{W'})^{n-b(n+d-d'-e)/d'}}.$$
(7.6) imposes the additional constraint $\det W' \ll H^{d'/dn}$ to this sum. Therefore, by Lemma 4.1, the contribution from the largest $H$-degree term in our estimate (1.2) of $P(\mathcal{L}, d')$, (const) $H^{d'/dn}$ to (7.5) is

$$
\ll \frac{H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)}}{(\det \mathcal{L})^{d'}} \int_{e_{d'}}^{H^{d'/dn}} \frac{x^{n-1}}{\chi^{-b(n+d-d' - e)/d'} dx} \\
\ll O(H + H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)})
$$

up to the two largest $H$-degree terms, as desired. The contributions from the smaller terms of $P(\mathcal{L}, d')$, (const) $H^{d'/dn}$ will be discussed later.

(ii) If $b(n + d - e)/d - d' > 0$, then we instead have

$$
\ll \frac{H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)}}{(\det \mathcal{L})^{d'}} \sum_{W' \in Gr(\mathcal{L}, d')} \frac{1}{(\det W')^{n-d+b}} \int_{0}^{H^{d'/dn}} \frac{\chi^{-b(n+d-d' - e)/d'} dx}{(\log W')^{d'/\pi}} \\
\ll \frac{H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)}}{(\det \mathcal{L})^{d'}} \sum_{W' \in Gr(\mathcal{L}, d')} \frac{H^{d'/dn} \chi^{-b(n+d-d' - e)/d'}}{\chi^{-b(n+d-d' - e)/d'} \log W'}
$$

Estimating in the same manner as in Case (i) above, this is

$$
O(H + H^{1 - \frac{d'}{\pi}(1 - \frac{2b}{a} + \frac{d' - b}{n-e})})
$$

and since $1 - 2b/d + (d' - b)/(n - e) \geq 0$, we are done.

(iii) In the rare, yet possible, case that $b(n + d - e)/d - d' = 0$, we proceed similarly, and bound (7.5) by

$$
\ll \frac{H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)}}{(\det \mathcal{L})^{d'}} \sum_{W' \in Gr(\mathcal{L}, d')} \frac{1}{(\det W')^{n-d+b}} \log \frac{H^{d'/dn} \chi^{-b(n+d-d' - e)/d'}}{H_{n-\epsilon}^{1/n} \chi^{-d'/d'}} \\
\ll \frac{H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)}}{(\det \mathcal{L})^{d'}} \int_{e_{d'}}^{H^{d'/dn}} x^{d-b-1} \log \left( \frac{H^{1/n} \chi^{d'/d'}}{x^{d'}} \right) dx \\
= \frac{H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)}}{(\det \mathcal{L})^{d'}} \left[ \frac{1}{d-b} \log H^{1/n} \cdot x^{d-b} \\
- \frac{d}{d'} \left( \frac{1}{d-b} x^{d-b} \log x - \frac{1}{(d-b)^2} x^{(d-b)} \right) \right]_{e_{d'}}^{H^{d'/dn}} \\
= O(H + H^{1 - \frac{b}{a} + \frac{b}{ma}(d' - d + e)} \log H).
$$
In addition, in all three cases above, the contribution from the leading error term of $P(L, d', H^{d'/dn})$ is of size $O(H^{1-b(n,d')d'/dn})$, and the number of error terms obtained is at most $2E_n$. This completes the error estimate. We note that the related computation in Thunder [20], lines 5–6 on p.185, contains a minor error: if $d-e=1$, the integral there diverges.

In summary, we estimated (7.2) to be

$$a(n, d)a(d, e)\frac{n}{n-e} \frac{H}{(\det L)^d} \log \frac{H}{\varepsilon_d^e} \sum_{j \in E_{n,e,d}} b_j(L) H^{\gamma_j},$$

where $E_{n,e,d}$ is an index set of cardinality at most $(d+2)2E_n \leq 3n^{3n+1}$, $b_j(L)$’s are appropriate inverse products of $\lambda_j(L)$’s, and the implied constant depends on $n$ only. The largest $\gamma_j$ is 1, and the second largest is one of

$$1 - \frac{b(n, d)}{n}, \quad 1 - \frac{b(d, e)(n-e)}{nd}, \quad \text{or} \quad 1 - \frac{1}{n}\left(1 - \frac{2b(d, e)}{d} + \frac{1 - b(d, e)}{n-e}\right).$$

When $d \leq n/2$, it is always $1 - b(n, d)/n = 1 - 1/dn$, but otherwise both of the other two are possible.

Acknowledgements Part of this work is supported by NSF grant CNS-2034176. The author thanks the anonymous referee, Lillian Pierce, Anders Södergren, and Jeffery Thunder for helpful comments and suggestions.

References

1. Banaszczyk, W.: New bounds in some transference theorems in the geometry of numbers. Math. Ann. 296(4), 625–635 (1993)
2. Borel, A.: Introduction to arithmetic groups. University Lecture Series, Vol. 73. The American Mathematical Society, Providence, RI (2019)
3. Davis, P.J.: 6. Gamma function and related functions. In: Abramowitz, M., Stegun, I. (eds.) Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York (1972)
4. Franke, J., Manin, Y., Tschinkel, Y.: Rational points of bounded height on Fano varieties. Invent. Math. 95(2), 421–435 (1989)
5. Kim, S.: Random lattice vectors in a set of size $O(n)$. Int. Math. Res. Not. 2020(5), 1385–1416 (2020)
6. Kim, S.: Mean value formulas on sublattices and flags of the random lattice. J. Number Theory (to appear)
7. Le Boudec, P.: Height of rational points on random Fano hypersurfaces. arXiv:2006.02288v1
8. Lenstra, A.K., Lenstra, H.W., Jr., Lovász, L.: Factoring polynomials with rational coefficients. Math. Ann. 261(4), 515–534 (1982)
9. Rogers, C.A.: Mean values over the space of lattices. Acta Math. 94, 249–287 (1955)
10. Rudin, W.: Principles of Mathematical Analysis, 3rd edn. McGraw-Hill, USA (1976)
11. Schanuel, S.H.: Heights in number fields. Bull. Soc. Math. France 107, 433–449 (1979)
12. Schmidt, W.M.: Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height. Duke Math. J. 35, 327–339 (1968)
13. Schmidt, W.M.: The distribution of sublattices of $\mathbb{Z}^m$. Monatshefte Math. 125, 37–81 (1998)
14. Schmidt, W.M.: Masstheorie in der Geometrie der Zahlen. Acta Math. 102(3–4), 159–224 (1959)
15. Shapira, U., Weiss, B.: A volume estimate for the set of stable lattices. Comptes Rendus Math. 352(11), 875–879 (2014)
16. Shimura, G.: Introduction to the Arithmetic Theory of Automorphic Functions. Princeton University Press, Princeton (1971)
17. Siegel, C.L.: A mean value theorem in geometry of numbers. Ann. Math. 46, 340–347 (1945)
18. Södergren, A., Strömbergsson, A.: On the generalized circle problem for a random lattice in large dimension. Adv. Math. 345, 1042–1074 (2019)
19. Thunder, J.L.: An asymptotic estimate for heights of algebraic subspaces. Trans. Am. Math. Soc. 331(1), 395–424 (1992)
20. Thunder, J.L.: Asymptotic estimates for rational points of bounded height on flag varieties. Compos. Math. 88(2), 155–186 (1993)
21. Thunder, J.L.: Higher-dimensional analogs of Hermite’s constant. Mich. Math. J. 45(2), 301–314 (1998)
22. Widmer, M.: Counting primitive points of bounded height. Trans. Am. Math. Soc. 362(9), 4793–4829 (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.