Flags and Lattice Basis Reduction

Hendrik W. Lenstra, Jr.

Abstract. In this lecture we give a self-contained introduction to the theory of lattices in Euclidean vector spaces. We reinterpret a large class of lattice basis reduction algorithms by using the concept of a "flag". In our reformulation, lattice basis reduction algorithms are more appropriately called "flag reduction" algorithms. We address a problem that arises when one attempts to find a particularly good flag for a given lattice.

1. Introduction

A *lattice* is a discrete subgroup of a Euclidean vector space. Every lattice has a *basis*, and a *lattice basis reduction algorithm* is an algorithm that transforms a given basis for a given lattice into a basis consisting of relatively short vectors.

The present lecture is devoted to a conceptual discussion of an important class of lattice basis reduction algorithms. This class includes an algorithm that I introduced in 1981 [4], its close relative that is known as the LLL or Lovász basis reduction algorithm from 1982 [3], and many variants of these algorithms that were proposed in subsequent years. The original applications of basis reduction algorithms to integer programming ([5, 1]) and to algorithmic number theory ([3, 2]) exemplify the scope of their importance in pure and applied mathematics.

The central notion in the present discussion is that of a *flag* for a lattice. A flag carries a little less information than a basis; the definition is given in section 6. The basis reduction algorithms that I consider can be reinterpreted in terms of flags; one may say that they transform a given flag for a lattice into a 'reduced' flag for the same lattice. To this end they perform a series of successive steps, each step replacing a flag by a 'neighbouring' one that is closer to being reduced. We picture this procedure by means of a directed graph, of which the vertices represent all flags for a given lattice, and the arcs the steps that are permitted. For the algorithm to be efficient, it is necessary that not too many steps are performed in succession. This leads to the problem of giving an upper bound for the length of a directed path in the graph that starts from a given vertex. In section 8 I present such an upper bound. It may be considered satisfactory if one considers only lattices of fixed rank. It is an interesting open problem to find an upper bound that has a better behaviour as a function of the rank.

2. Euclidean Vector Spaces

A *Euclidean vector space* is a finite dimensional vector space E over the field $\mathbf{R}$ of real numbers equipped with a map $\langle\ ,\ \rangle\colon E\times E\to\mathbf{R}$ satisfying

$$\begin{aligned}\langle w+x,y\rangle&=\langle w,y\rangle+\langle x,y\rangle, &\langle rx,y\rangle&=r\langle x,y\rangle,\\ \langle x,y\rangle&=\langle y,x\rangle, &\langle z,z\rangle&>0\end{aligned}$$

for all $r\in\mathbf{R}$ and $w,x,y,z\in E$, $z\neq 0$. We refer to the map $\langle\ ,\ \rangle$ as the *inner product* on E. Any Euclidean vector space E is a *metric space* with distance function $d\colon E\times E\to\mathbf{R}$ defined by $d(x,y)=\langle x-y,x-y\rangle^{1/2}$. For each non-negative integer n, the vector space $\mathbf{R}^n$ is a Euclidean vector space with the *standard inner product* defined by $\langle(x_i)_{i=1}^n,(y_i)_{i=1}^n\rangle=\sum_{i=1}^n x_iy_i$.

Let E be a Euclidean vector space, and let $D\subset E$ be a subspace. Then the restriction of $\langle\ ,\ \rangle$ to $D\times D$ makes D into a Euclidean vector space. Let $D^\dagger=\mathrm{Hom}(D,\mathbf{R})$ be the dual of D, and write $D^\perp$ for the kernel of the linear map $E\to D^\dagger$ sending $x\in E$ to the map $y\mapsto\langle x,y\rangle$. One has $D\cap D^\perp=0$, so the natural maps $D\to E/D^\perp\to D^\dagger$ are injective, and by $\dim D=\dim D^\dagger$ they are isomorphisms. It follows that each $w\in E$ has a unique representation as $w=x+y$, with $x\in D$ and $y\in D^\perp$; this is the *orthogonal decomposition* of w with respect to D. The quotient space E/D becomes a Euclidean vector space as well, since it is canonically isomorphic to the subspace $D^\perp$ of E. One also concludes that E is canonically isomorphic to $E^\dagger$.

Applying the above to one-dimensional subspaces, one easily proves by induction on $\dim E$ that for every Euclidean vector space E there is a linear isomorphism from the standard Euclidean vector space $\mathbf{R}^{\dim E}$ to E that preserves inner products. For example, if one makes the field $\mathbf{C}$ of complex numbers into a Euclidean vector space by putting $\langle x,y\rangle=(x\bar{y}+y\bar{x})/2$ (= the real part of $x\bar{y}$), then the isomorphism $\mathbf{R}^2\to\mathbf{C}$ sending (a,b) to $a+bi$ preserves inner products.

Proposition 2.1. *Let D be a finite dimensional vector space over $\mathbf{R}$, and let $\langle\ ,\ \rangle\colon D\times D\to\mathbf{R}$ be a map satisfying $\langle w+x,y\rangle=\langle w,y\rangle+\langle x,y\rangle$, $\langle rx,y\rangle=r\langle x,y\rangle$, $\langle x,y\rangle=\langle y,x\rangle$, and $\langle x,x\rangle\geq 0$ for all $r\in\mathbf{R}$, $w,x,y\in D$. Write $\operatorname{rad}D=\{x\in D:\langle x,x\rangle=0\}$. Then $\operatorname{rad}D$ is a subspace of D, and if we write $E=D/\operatorname{rad}D$ then $\langle\ ,\ \rangle$ can be written as the composition of the natural map $D\times D\to E\times E$ and a uniquely determined map $E\times E\to\mathbf{R}$; moreover, the latter map makes E into a Euclidean vector space.*

Proof. Let $w\in\operatorname{rad}D$, $y\in D$. Then one has $2r\langle w,y\rangle+\langle y,y\rangle=\langle rw+y,rw+y\rangle\geq 0$ for every $r\in\mathbf{R}$, and therefore $\langle w,y\rangle=0$. One readily deduces that $\operatorname{rad}D$ is a subspace and that for any $x,x',y,y'\in D$ with $x\equiv x'\bmod\operatorname{rad}D$ and $y\equiv y'\bmod\operatorname{rad}D$ one has $\langle x,y\rangle=\langle x',y'\rangle$. Hence $\langle\ ,\ \rangle$ factors through $E\times E$. The last statement is now immediate. □

3. Lattices in a Euclidean Vector Space

Let E be a Euclidean vector space. A *lattice in* E is an additive subgroup L of E for which there exists a positive real number l such that every $z \in L$, $z \neq 0$, satisfies $\langle z, z\rangle \geq l$; equivalently, it is an additive subgroup of E that is *discrete* in the topology induced by the metric on E.

Proposition 3.1. *Let L be a lattice in a Euclidean vector space E, and let $r, l \in \mathbf{R}$ be real numbers with $r \geq 0$, $l > 0$ such that every $z \in L$, $z \neq 0$, satisfies $\langle z, z\rangle \geq l$. Write n for the dimension of the subspace of E spanned by L. Then we have*

$$\#\{x \in L : \langle x, x\rangle \leq r\} \leq (1 + 2\sqrt{r/l})^n .$$

Proof. Replacing E by the subspace spanned by L we may assume $n = \dim E$. Since L is an additive subgroup, any two distinct elements of L differ by a non-zero element of L and therefore have distance at least $\sqrt{l}$. Hence the open n-dimensional balls with radius $\sqrt{l}/2$ centered at all $x \in L$ are pairwise disjoint. All of these balls whose center x satisfies $\langle x, x\rangle \leq r$ are contained in the open ball with radius $\sqrt{r} + \sqrt{l}/2$ centered at 0. Computing volumes we find

$$\#\{x \in L : \langle x, x\rangle \leq r\} \cdot (\sqrt{l}/2)^n \leq (\sqrt{r} + \sqrt{l}/2)^n .$$

(It is practical to rescale the volume on E so that the unit ball has volume 1.) □

Let $\mathbf{Z}$ denote the ring of integers.

Proposition 3.2. *Let E be a Euclidean vector space, let $b_1, b_2, \ldots, b_n \in E$ be linearly independent, and for each i let $b_i = (b_i - b_i^*) + b_i^*$ be the orthogonal decomposition of b_i with respect to $\sum_{j<i} \mathbf{R} b_j$; so $\langle b_i^*, b_j\rangle = 0$ for $j < i$, and $b_i - b_i^* \in \sum_{j<i} \mathbf{R} b_j$. Then all b_i^* are non-zero, and for each $z \in \sum_i \mathbf{Z} b_i$, $z \neq 0$, one has $\langle z, z\rangle \geq \min_i \langle b_i^*, b_i^*\rangle$.*

Remark. If the b_i are not supposed to be linearly independent, then the dimension of the subspace they span equals the number of i for which $b_i^* \neq 0$.

Proof. We have $b_i^* \neq 0$ since $b_i \notin \sum_{j<i} \mathbf{R} b_j$. For $z = \sum_i n_i b_i \in \sum_i \mathbf{Z} b_i$, $z \neq 0$, choose i maximal with $n_i \neq 0$. Then $z = (z - n_i b_i^*) + n_i b_i^*$ is the orthogonal decomposition of z with respect to $\sum_{j<i} \mathbf{R} b_j$, so $\langle z, z\rangle \geq n_i^2 \langle b_i^*, b_i^*\rangle \geq \langle b_i^*, b_i^*\rangle$. □

Proposition 3.3. *Let E be a Euclidean vector space and let L be a subset of E. Then L is a lattice in E if and only if there is a linearly independent subset $B \subset E$ with $L = \sum_{b \in B} \mathbf{Z} b$.*

If B and L are as in proposition 3.3, then B is called a *basis* for L. Its cardinality $\#B$ equals the number n defined in proposition 3.1, so it depends only on L; it is called the *rank* $\operatorname{rk} L$ of L. One has $\operatorname{rk} L \leq \dim E$.

Proof. The if-part follows from proposition 3.2. For the only if-part, let $C \subset L$ be a basis for the subspace D of E spanned by L. Each $w \in L$ can be written as $w = x + y$ with $x \in M = \sum_{c \in C} \mathbf{Z}c$ and y in the intersection of L with the bounded set $\sum_{c \in C}[0,1)c$. By proposition 3.1, that intersection is finite, so M has finite index m (say) in L. Lagrange's theorem from group theory now implies that $mL \subset M$, so L is a subgroup of finite index of the free abelian group $m^{-1}M$. Therefore L has a basis B over $\mathbf{Z}$ with $\#B = \#C$, and B is linearly independent since its span contains C. □

4. Lattices

We next define lattices in an absolute sense, without reference to a Euclidean vector space. A *lattice* is a finitely generated abelian group L equipped with a map $q\colon L \to \mathbf{R}$ satisfying the following three conditions:

(i) $q(x+y)+q(x-y) = 2q(x)+2q(y)$ for all $x, y \in L$ (the *parallelogram* law);
(ii) $q(z) \neq 0$ for all $z \in L$, $z \neq 0$;
(iii) for each real number r, the set $\{x \in L : q(x) \leq r\}$ is *finite*.

An *isomorphism* from a lattice L, q to a lattice L', q' is a group isomorphism $f\colon L \to L'$ such that for all $x \in L$ one has $q(x) = q'(f(x))$; if such a map exists then the lattices L and L' are called *isomorphic.*

By proposition 3.1, any lattice in a Euclidean vector space becomes a lattice in the sense just defined if we put $q(x) = \langle x, x\rangle$. We prove that, up to isomorphism, any lattice can be obtained in this way.

Proposition 4.1. *Any lattice is isomorphic to a lattice in a Euclidean vector space.*

Proof. Let L, q be a lattice. For $x, y \in L$, define $\langle x, y\rangle = \big(q(x+y) - q(x-y)\big)/4$. The parallelogram law implies $q(x-y) = q(y-x)$, so we have $\langle x, y\rangle = \langle y, x\rangle$. Let $w, x, y \in L$. We have by the parallelogram law

$$\begin{aligned} q(w+x+y) + q(w-x+y) &= 2q(w+y) + 2q(x)\,, \\ q(w+x-y) + q(w-x-y) &= 2q(w-y) + 2q(x)\,, \\ q(w+x+y) + q(w-x-y) &= 2q(x+y) + 2q(w)\,, \\ q(w+x-y) + q(w-x+y) &= 2q(x-y) + 2q(w)\,. \end{aligned}$$

Taking the alternating sum and dividing by 8 we find that $\langle w+x, y\rangle = \langle w, y\rangle + \langle x, y\rangle$. One readily checks that $q(0) = 0$ and $\langle x, x\rangle = q(2x)/4 = q(x)$, for $x \in L$. If $x \in L$ satisfies $q(x) < 0$, then one has $q(mx) = m^2 q(x) < 0$ for all non-zero $m \in \mathbf{Z}$, so x has infinite order, and one obtains a contradiction with (iii); hence $q(x) \geq 0$ for all $x \in L$. Write $D = \mathbf{R} \otimes_{\mathbf{Z}} L$, and let $\langle\ ,\ \rangle\colon D \times D \to \mathbf{R}$ be the $\mathbf{R}$-bilinear function induced by $\langle\ ,\ \rangle\colon L \times L \to \mathbf{R}$. For each positive integer m and each $x \in L$ the element $z = (1/m) \otimes x$ of D satisfies $\langle z, z\rangle = q(x)/m^2 \geq 0$, and since the set of all $z \in D$ of this form is dense in D one has $\langle z, z\rangle \geq 0$ for all $z \in D$. From proposition 2.1 one now obtains a Euclidean vector space $E = D/\operatorname{rad} D$ such that the group homomorphism $f\colon L \to E$ sending x to the coset $(1 \otimes x) \bmod \operatorname{rad} D$ satisfies $q(x) = \langle f(x), f(x)\rangle$. By property (ii) the map is injective, and using (iii) one deduces that $f(L)$ is a lattice in E that is isomorphic to L. (Comparing ranks and dimensions one also finds $\operatorname{rad} D = 0$, so $D = E$.) □

Remark. In the sequel, we shall write $\langle x, y\rangle = \big(q(x+y) - q(x-y)\big)/4$ for x, y in a lattice, and lattices may tacitly be assumed to be embedded in a Euclidean vector space. This is justified by proposition 4.1 and its proof.

RANK AND DETERMINANT Two important numerical invariants attached to any lattice L are its *rank* $\operatorname{rk} L$ and its *determinant* $d(L)$. The rank is the unique non-negative integer n for which there is an isomorphism $L \cong \mathbf{Z}^n$ of abelian groups. The determinant is defined by

$$d(L) = \det\big(\langle b, b'\rangle\big)^{1/2}_{b,b'\in B},$$

where B is a basis of L; if L is a lattice in $\mathbf{R}^n$ with basis equal to the set of columns of a non-singular $n \times n$ matrix $\mathbf{B}$, then one has $d(L) = |\det \mathbf{B}|$. One way to prove that $d(L)$ is well-defined is by showing the limit relation

$$\lim_{r\to\infty} \frac{\#\{x \in L : q(x) \leq r\}}{\omega_n r^n / d(L)} = 1,$$

which is valid for any lattice L of rank n. Here we write $\omega_n = \pi^{n/2}/\frac{n}{2}!$ for the standard volume of the unit ball in $\mathbf{R}^n$; the factor $\frac{n}{2}! = \Gamma(1+\frac{n}{2})$ may be computed from $0! = 1$, $\frac{1}{2}! = \sqrt{\pi}/2$, and $z! = z \cdot (z-1)!$. We have $d(L) = 1$ if $\operatorname{rk} L = 0$.

Proposition 4.2. *Let L, q be a lattice of positive rank n. Then there exists $x \in L$ with $x \neq 0$ and $q(x) \leq n \cdot d(L)^{2/n}$.*

Proof. Assume that L is a lattice in $\mathbf{R}^n$, and write vol for the standard n-dimensional volume. Let $B \subset \mathbf{R}^n$ be a basis for L, and write $F = \sum_{b\in B}[0,1)b$. Then $\operatorname{vol} F = d(L)$, and $\mathbf{R}^n$ is the disjoint union of the sets $x + F$, $x \in L$. Let $l = \min\{q(x) : x \in L, x \neq 0\}$, and write $t = \sqrt{l/n}$, so that the assertion of proposition 4.2 is equivalent to $t^n \leq d(L)$. Let C be the cube $[0,t)^n$ in $\mathbf{R}^n$. Any two elements of C have distance smaller than $t\sqrt{n} = \sqrt{l}$, so their difference is not a non-zero element of L. Hence the sets $-x + C$, $x \in L$, are pairwise disjoint. Since C is the disjoint union of the sets $(x+F) \cap C$, $x \in L$, we conclude that

$$\begin{aligned} t^n = \operatorname{vol} C &= \sum_{x\in L} \operatorname{vol}\big((x+F)\cap C\big) = \sum_{x\in L} \operatorname{vol}\big(F \cap (-x+C)\big) \\ &= \operatorname{vol}\big(F \cap \textstyle\bigcup_{x\in L}(-x+C)\big) \leq \operatorname{vol}(F) = d(L), \end{aligned}$$

as required. □

Remark. Replacing the cube in the proof by an open ball of radius $\sqrt{l}/2$ one finds the better inequality $q(x) \leq 4\omega_n^{-2/n} \cdot d(L)^{2/n}$, with ω_n as above, and further improvements are possible. One has $4\omega_n^{-2/n} = 2n/(\pi e + o(1))$ for $n \to \infty$.

SUBLATTICES AND QUOTIENT LATTICES Let L, q be a lattice, and let K be a subgroup of L. Then the restriction of q to K makes K into a lattice, a *sublattice* of L. We next restrict to *pure* subgroups. In general, a subgroup K of an additively written abelian group L is called *pure* if for all positive integers m one has $mK = K \cap mL$. If L is a lattice, this property is equivalent to L/K being torsion-free; and if L is a lattice in a Euclidean vector space E, then it is equivalent to the existence of a subspace D of E such that $K = L \cap D$, and also to L having a basis that contains a basis for K. Now suppose that K is a pure sublattice of a lattice L. Then the map $q' \colon L/K \to \mathbf{R}$ defined by

$$q'(x+K) = \inf\{q(mx-y)/m^2 : m \in \mathbf{Z}, m \neq 0, y \in K\}$$

makes L/K into a lattice. To prove this, one embeds L as a lattice in a Euclidean vector space E, one defines D to be the subspace of E spanned by K, and one verifies that q' is induced by the inclusion of L/K in the Euclidean vector space E/D. One has

$$\operatorname{rk} K + \operatorname{rk}(L/K) = \operatorname{rk} L, \quad d(K) \cdot d(L/K) = d(L).$$

Proposition 4.3. *Let L be a lattice and let r be a real number. Then the number of sublattices K of L with $d(K) \leq r$ is finite.*

Proof. For any subgroup $K \subset L$, with $\mathbf{R}$-linear span $\mathbf{R} \cdot K$, the subgroup $K' = L \cap (\mathbf{R} \cdot K)$ is pure, the number $m = \operatorname{index}[K' : K]$ is finite, and one has $d(K) = m \cdot d(K')$. Hence we may restrict to *pure* subgroups. We apply induction on $\operatorname{rk} L$. The set of non-zero b in L with $q(b) \leq \max\{i \cdot r^{2/i} : 1 \leq i \leq n\}$ is finite, and by proposition 4.2 any non-zero subgroup $K \subset L$ with $d(K) \leq r$ contains at least one of them. If K is a pure subgroup containing a given such b, then it also contains the pure subgroup $L_b = L \cap \mathbf{R}b$, and K/L_b is a pure subgroup of L/L_b with $d(K/L_b) = d(K)/d(L_b)$. Now apply the induction hypothesis to each L/L_b. □

Remark. An alternative proof of proposition 4.3 makes use of exterior powers. For subgroups of rank 1, one uses defining property (iii) of lattices. Generally, if $K \subset L$ is a subgroup of rank i, then $\wedge^i K \subset \wedge^i L$ is a subgroup of rank 1, and $\wedge^i L$ has a natural lattice structure for which $d(\wedge^i K) = d(K)$; in addition, K is 'almost' determined by $\wedge^i K$ in the sense that another subgroup $J \subset L$ of rank i satisfies $\wedge^i J = \wedge^i K$ if and only if J is a subgroup of $L \cap (\mathbf{R} \cdot K)$ of the same index as K.

Remark. It follows from proposition 4.3 that there is a positive lower bound for the determinants of the subgroups of a given lattice. Explicitly, any subgroup $K \subset L$ with $\operatorname{rk} K = i > 0$ satisfies $d(K) \geq \left(\min\{q(x) : x \in L, x \neq 0\}/i\right)^{i/2}$, by proposition 4.2.

THE DUAL Let L be a lattice in a Euclidean vector space E with $\dim E = \operatorname{rk} L$. Then $L^\dagger = \{x \in E : \langle x, L \rangle \subset \mathbf{Z}\}$ is also a lattice in E, the *dual* (or *polar*) of L. One has

$$\operatorname{rk} L^\dagger = \operatorname{rk} L, \quad d(L^\dagger) = d(L)^{-1}, \quad L^{\dagger\dagger} = L.$$

If L is a lattice in $\mathbf{R}^n$ with basis equal to the set of columns of a certain non-singular matrix, then the columns of the inverse transpose matrix form a basis for $L^\dagger$. If desired, one can also define the dual without reference to a Euclidean vector space, by taking $L^\dagger = \mathrm{Hom}(L, \mathbf{Z})$ and letting $q(f)$, for $f \in L^\dagger$, be the infimum of all non-negative real numbers r with the property that for all $x \in L$ one has $f(x)^2 \leq r \cdot q(x)$.

Let L be a lattice, with dual $L^\dagger$, and let $K \subset L$ be a pure sublattice. Then $K^\perp = \{x \in L^\dagger : \langle x, K\rangle = 0\}$ is a pure sublattice of $L^\dagger$ that may be identified with $(L/K)^\dagger$, and $K^\dagger$ may be identified with $L^\dagger/K^\perp$; in addition, one has $K^{\perp\perp} = K$.

5. Algorithmic Problems

In the present section we discuss a few fundamental and frequently encountered problems concerning lattices. The first is the *homogeneous approximation* problem: *given a non-zero lattice L, find a non-zero element $x \in L$ with $q(x)$ smallest possible.* The informal formulation allows many interpretations. For example, the lattice may be 'given' in some theoretical sense, and 'finding' x may be meant purely existentially, so that proposition 4.2 goes some way towards solving the problem. We are mainly interested in an algorithmic interpretation, in which the lattice is 'given' in some numerical manner, and likewise its elements have a numerical representation; the problem of 'finding' x is then to be interpreted algorithmically, and one wants not just $q(x)$ but also the run time of the algorithm to be small. One will have to allow for a trade-off between the latter two quantities, and the requirement that $q(x)$ be 'smallest possible' may be taken to mean: smallest possible given the time that one is willing to spend.

One way of specifying a lattice L numerically is by means of a real $m \times n$ matrix $\mathbf{B}$ of rank n; then L is embedded in the Euclidean vector space $\mathbf{R}^m$, the columns of $\mathbf{B}$ forming a basis, and an element $x \in L$ is either represented as a real m-vector or as an integral n-vector consisting of the coefficients of x on that basis. In order to avoid rounding problems one may require the entries of $\mathbf{B}$ to be rational. A second way of specifying L is by means of a real positive definite symmetric $n \times n$ matrix $\mathbf{A}$; in this case L is the group $\mathbf{Z}^n$, its elements are represented as integral n-vectors, and $\langle x, y\rangle = x^T \mathbf{A} y$ for $x, y \in L$. Again one may require the entries of $\mathbf{A}$ to be rational. One easily transforms the first type of representation into the second by taking $\mathbf{A} = \mathbf{B}^T \cdot \mathbf{B}$, and this transformation preserves rationality. One can also transform the second representation into the first, but complications arise if one wishes to do this by means of a polynomial time algorithm that preserves rationality and keeps m low. There are other possibilities of representing lattices numerically, but the two that we just mentioned appear to be the most convenient ones for algorithmic purposes.

Of the many algorithmic situations giving rise to the homogeneous approximation problem we mention a single one; namely, the problem of factoring a given one-variable polynomial f with rational coefficients into irreducible factors, which was considered in [3]. In this case, one can take the lattice to consist of integer polynomials of a certain degree that assume a very small value in a suitably constructed p-adic zero of f, and one proves that any sufficiently short non-zero vector in that lattice must be an irreducible factor of f.

The homogeneous approximation problem has also appeared under the following guise: *given a lattice L in a Euclidean vector space E of dimension* $\operatorname{rk} L$, *find $x \in E$ with $L \subset (\mathbf{R}x)^{\perp} + \mathbf{Z}x$ and $\langle x, x\rangle$ largest possible.* Geometrically, this amounts to asking for a hyperplane H in E such that L is contained in the union of a collection of maximally widely spaced translates of H; namely, take $H = (\mathbf{R}x)^{\perp}$ and consider translates with successive distances equal to $\langle x, x\rangle^{1/2}$. Such a hyperplane is useful when one wishes to enumerate elements of L that lie in a certain bounded region, which occurs in the context of integer programming (see [5]). A given non-zero vector $x \in E$ satisfies $L \subset (\mathbf{R}x)^{\perp} + \mathbf{Z}x$ if and only if $x/\langle x, x\rangle$ belongs to the dual $L^{\dagger}$ of L, so the problem is equivalent to the homogeneous approximation problem for $L^{\dagger}$.

Finally, one frequently encounters the *inhomogeneous approximation* problem: *given a lattice L in a Euclidean vector space E, and $x \in E$, find $y \in E$ with $x - y \in L$ and $\langle y, y\rangle$ smallest possible.* In other words, one wishes to 'round' a given element x of E to an element w of L such that the 'error' $d(x, w)$ is minimal. It is a mistake to think that the special case $x = 0$ of the inhomogeneous approximation problem amounts to the homogeneous approximation problem (since one takes $w = y = 0$); but it is true that solving $2^{\operatorname{rk} L} - 1$ inhomogeneous approximation problems suffices to solve the homogeneous approximation problem; namely, let x range over coset representatives of all non-trivial elements of $\frac{1}{2}L/L$.

All problems that we mentioned can to a certain extent be solved if a *reduced basis* of the lattice is available. The notion of a 'reduced basis' has many different definitions, and one usually chooses the most convenient one for the purpose at hand. Different definitions are rarely logically equivalent, but typically bases that are reduced in different senses share many qualitative properties: they consist of 'fairly short' vectors that stand at 'almost right' angles, the product of their lengths is a 'fair' approximation to the determinant of the lattice, and, of course, they yield solutions to the three problems formulated above.

In the next section we shall consider *flags* of a lattice. The notion of a flag is a little weaker than the notion of a basis, but it still carries enough information to assist us in solving our three problems.

Finding a reduced basis for a given lattice is done by means of a *lattice basis reduction algorithm*, which replaces a given basis for a given lattice by a reduced basis for the same lattice. We shall not present any of these. Instead, we describe in very general terms a *flag reduction algorithm*, that is, a procedure that replaces a given flag of a given lattice by what might be called a 'reduced flag' of the same lattice; but we refrain from giving a rigorous definition of the latter term. Many existing lattice basis reduction algorithms, including those presented in [4] and [3], may be interpreted as flag reduction algorithms, and fit as such under our general description.

6. Flags

Let L be a lattice, and write $n = \operatorname{rk} L$. A *flag* of L is a sequence $\mathfrak{F} = (F_i)_{i=0}^n$ of pure sublattices F_i of L satisfying $\operatorname{rk} F_i = i$ (for $0 \leq i \leq n$) and $F_{i-1} \subset F_i$ (for $0 < i \leq n$); clearly we must have $F_0 = \{0\}$ and $F_n = L$. Every basis $(b_i)_{i=1}^n$ of L gives rise to the flag $\left(\sum_{j \leq i} \mathbf{Z} b_j\right)_{i=0}^n$, and one readily checks that every flag of L is of this form. In order to express when two bases $(b_i)_{i=1}^n$ and $(a_i)_{i=1}^n$ of L give rise to the same flag, let $(b_i^*)_{i=1}^n$ be defined as in proposition 3.2, and $(a_i^*)_{i=1}^n$ analogously. Then the two bases give rise to the same flag of L if and only if for each i one has $b_i^* = \pm a_i^*$; or, equivalently, if and only if there are integers c_{ij}, for $1 \leq j \leq i \leq n$, with $b_i = \sum_{j=1}^i c_{ij} a_j$ and $c_{ii} = \pm 1$ for all i.

In an algorithmic context one may wish to represent a flag numerically. Assuming L and its elements to be represented in one of the manners described in section 5, one can do this by specifying a basis $(b_i)_{i=1}^n$ of L; then the flag is $\left(\sum_{j \leq i} \mathbf{Z} b_j\right)_{i=0}^n$. As we just noted, certain changes in the basis do not change the flag. This freedom is often used in order to achieve that the real numbers μ_{ij} for which $b_i - b_i^* = \sum_{j=1}^{i-1} \mu_{ij} b_j^*$ satisfy $|\mu_{ij}| \leq \frac{1}{2}$.

Let $\mathfrak{F} = (F_i)_{i=0}^n$ be a flag of L. The *size* $s(\mathfrak{F})$ of $\mathfrak{F}$ is defined by

$$s(\mathfrak{F}) = \prod_{i=0}^{n} d(F_i)\,.$$

For $1 \leq i \leq n$, the *i-th successive distance* $l_i(\mathfrak{F})$ is defined by $l_i(\mathfrak{F}) = d(F_i/F_{i-1})$; if $\mathfrak{F}$ is obtained from a basis $(b_j)_{j=1}^n$, and $(b_j^*)_{j=1}^n$ is as above, then one has $l_i(\mathfrak{F}) = \langle b_i^*, b_i^* \rangle^{1/2}$. One has $d(F_i) = \prod_{j=1}^i l_j(\mathfrak{F})$ for $0 \leq i \leq n$, and $s(\mathfrak{F}) = \prod_{j=1}^n l_j(\mathfrak{F})^{n+1-j}$. It is an easy consequence of proposition 4.3 that L has, for any real number r, only finitely many flags $\mathfrak{F}$ with $s(\mathfrak{F}) \leq r$.

Let again $\mathfrak{F} = (F_i)_{i=0}^n$ be a flag of L. Then $\mathfrak{F}^\perp = (F_{n-i}^\perp)_{i=0}^n$ is a flag of $L^\dagger$, the flag *dual* to $\mathfrak{F}$. One has

$$l_i(\mathfrak{F}^\perp) = l_{n+1-i}(\mathfrak{F})^{-1}, \quad s(\mathfrak{F}^\perp) = s(\mathfrak{F})/d(L)^{n+1}, \quad \mathfrak{F}^{\perp\perp} = \mathfrak{F}$$

for $1 \leq i \leq n$.

We shall in particular be interested in flags $\mathfrak{F}$ with the property that $l_{i+1}(\mathfrak{F})$ is not much smaller than $l_i(\mathfrak{F})$, for each $i = 1, \ldots, n-1$. The following result and its proof show the relevance of such flags for the homogeneous approximation problem formulated in the previous section.

Proposition 6.1. *Let c be a real number with $c \geq 1$, let L be a non-zero lattice in a Euclidean vector space E of dimension $n = \operatorname{rk} L$, and let $\mathfrak{F}$ be a flag of L with* *the property* $l_{i+1}(\mathfrak{F})^2 \geq c^{-1} \cdot l_i(\mathfrak{F})^2$ *for* $0 < i < n$. *Then we have*

$$c^{1-n} \cdot l_1(\mathfrak{F})^2 \leq \min\{q(x) : x \in L,\, x \neq 0\} \leq l_1(\mathfrak{F})^2$$
$$l_n(\mathfrak{F})^2 \leq \max\{\langle x, x\rangle : x \in E,\, L \subset (\mathbf{R}x)^{\perp} + \mathbf{Z}x\} \leq c^{n-1} \cdot l_n(\mathfrak{F})^2 .$$

Proof. Let $(b_i)_{i=1}^n$ be a basis of L such that $\mathfrak{F} = \left(\sum_{j\leq i} \mathbf{Z}b_j\right)_{i=0}^n$, and let $(b_i^*)_{i=1}^n$ be as in proposition 3.2. By proposition 3.2, we have $\min\{q(x) : x \in L,\, x \neq 0\} \geq \min_i\langle b_i^*, b_i^*\rangle = \min_i l_i(\mathfrak{F})^2$. The hypotheses imply that $l_i(\mathfrak{F})^2 \geq c^{1-n} \cdot l_1(\mathfrak{F})^2$, and the first inequality follows. The second inequality follows from $l_1(\mathfrak{F})^2 = q(b_1)$. One proves the last two inequalities by applying the first two to the dual flag. Note that $x = b_n^*$ satisfies $L \subset (\mathbf{R} \cdot x)^{\perp} + \mathbf{Z}x$ and $l_n(\mathfrak{F})^2 = \langle x, x\rangle$. □

Remark. The flags considered in proposition 6.1 also yield a fairly good solution to the inhomogeneous approximation problem. Namely, if the notation and hypotheses are as in proposition 6.1 and its proof, then for every $x \in E$ there is a unique element $y \in \sum_{i=1}^n (-\frac{1}{2}, \frac{1}{2}] \cdot b_i^*$ with the property $x - y \in L$, and this element y satisfies

$$\frac{c-1}{c^n - 1} \cdot \langle y, y\rangle \leq \min\{\langle z, z\rangle : z \in E,\, x - z \in L\} \leq \langle y, y\rangle ,$$

where one should read $\frac{1}{n}$ for $\frac{c-1}{c^n-1}$ if $c = 1$.

To obtain the best results in proposition 6.1, one should take c smallest possible. In section 8 we shall see that $c = \frac{4}{3}$ can be achieved; that is, every lattice L has a flag $\mathfrak{F}$ with the property $l_{i+1}(\mathfrak{F})^2 \geq \frac{3}{4} \cdot l_i(\mathfrak{F})^2$ for $0 < i < \operatorname{rk} L$. Also, $\frac{4}{3}$ is best possible in the sense that for any $n > 1$ there is a lattice L of rank n such that for every flag $\mathfrak{F}$ of L there exists i with $0 < i < n$ and $l_{i+1}(\mathfrak{F})^2 \leq \frac{3}{4} \cdot l_i(\mathfrak{F})^2$. Namely, one can take L to be the 'orthogonal sum' of the hexagonal lattice $\mathbf{Z}2 + \mathbf{Z}(1 + \mathrm{i}\sqrt{3})$ in $\mathbf{C}$ with the lattice $N\mathbf{Z}^{n-2}$ in $\mathbf{R}^{n-2}$, for N large enough.

7. The Reduction Graph

Let L be a lattice, and let n be its rank. We write $\Gamma(L)$ for the set of flags of L. We make $\Gamma(L)$ into the set of vertices of a directed graph, the *reduction graph* of L, by drawing an arc from $\mathfrak{F} = (F_i)_{i=0}^n$ to $\mathfrak{F}' = (F_i')_{i=0}^n$ if and only if there exists j, $0 < j < n$, with the following properties:

(i) $F_i = F_i'$ for all $i \neq j$;
(ii) $F_j + F_j' = F_{j+1}$;
(iii) $s(\mathfrak{F}')$ is minimal, given (i) and (ii);
(iv) $s(\mathfrak{F}') < s(\mathfrak{F})$.

Condition (iii) means, more formally, that for all flags $\mathfrak{G} = (G_i)_{i=0}^n$ of L satisfying $F_i = G_i$ for all $i \neq j$ and $F_j + G_j = F_{j+1}$ one has $s(\mathfrak{F}') \leq s(\mathfrak{G})$. To reformulate this condition, suppose that (i) and (ii) are satisfied; then we can write $F_j/F_{j-1} = \mathbf{Z}w$, $F_j'/F_{j-1} = \mathbf{Z}x$ for a certain basis w, x of the rank 2 lattice F_{i+1}/F_{i-1}, and (iii) is now equivalent to the inequality $|\langle x, w\rangle| \leq \langle w, w\rangle/2$; also, one has $s(\mathfrak{F}')^2/s(\mathfrak{F})^2 = q(x)/q(w)$, so (iv) is equivalent to $q(x) < q(w)$.

We write $\mathfrak{F} \rightarrow \mathfrak{F}'$ to denote an arc from $\mathfrak{F}$ to $\mathfrak{F}'$, and refer to it as a *step* in $\Gamma(L)$. The *length* of such a step is defined to be $s(\mathfrak{F})^2/s(\mathfrak{F}')^2$, and the number j appearing above is called the *colour* of the step; by (i) and (ii) it is uniquely determined.

One readily checks that there are at most two steps in $\Gamma(L)$ that start from a given flag and have a given colour; and if there are two, then they have the same length.

Let K be a pure sublattice of L. The set of flags of L that comprise K may in an obvious manner be identified with $\Gamma(K) \times \Gamma(L/K)$. With this identification, one has $s((\mathfrak{E}, \mathfrak{F})) = s(\mathfrak{E}) \cdot s(\mathfrak{F}) \cdot d(K)^{(\operatorname{rk} K) \operatorname{rk}(L/K)}$, and there is a step $(\mathfrak{E}, \mathfrak{F}) \rightarrow (\mathfrak{E}', \mathfrak{F}')$ in $\Gamma(L)$ if and only if either $\mathfrak{F} = \mathfrak{F}'$ and there is a step $\mathfrak{E} \rightarrow \mathfrak{E}'$ in $\Gamma(K)$, or $\mathfrak{E} = \mathfrak{E}'$ and there is a step $\mathfrak{F} \rightarrow \mathfrak{F}'$ in $\Gamma(L/K)$; in the former case, $(\mathfrak{E}, \mathfrak{F}) \rightarrow (\mathfrak{E}', \mathfrak{F})$ has the same length and colour as $\mathfrak{E} \rightarrow \mathfrak{E}'$, and in the latter case $(\mathfrak{E}, \mathfrak{F}) \rightarrow (\mathfrak{E}, \mathfrak{F}')$ has the same length as $\mathfrak{F} \rightarrow \mathfrak{F}'$ but the colour is larger by $\operatorname{rk} K$.

Proposition 7.1. *Let L be a lattice. Then the map $\Gamma(L) \rightarrow \Gamma(L^\dagger)$ sending $\mathfrak{F}$ to $\mathfrak{F}^\perp$ is an isomorphism of directed graphs. Corresponding steps have the same length, and their colours add up to* $\operatorname{rk} L$.

Proof. This is entirely straightforward, and left to the reader. □

The following result decribes the effect of a step on the successive lengths.

Proposition 7.2. *Let L be a lattice, let $\mathfrak{F} \rightarrow \mathfrak{F}'$ be a step in $\Gamma(L)$, and let j be its colour. Then one has $l_i(\mathfrak{F}) = l_i(\mathfrak{F}')$ for all $i \neq j$, $j+1$, and*

$$l_{j+1}(\mathfrak{F}) \leq l_j(\mathfrak{F}') < l_j(\mathfrak{F}), \quad l_{j+1}(\mathfrak{F}) < l_{j+1}(\mathfrak{F}') \leq l_j(\mathfrak{F}).$$

Proof. The relation $l_i(\mathfrak{F}) = d(F_i)/d(F_{i-1})$ and (i) imply the first assertion. Write $F_j/F_{j-1} = \mathbf{Z}w$ and $F'_j/F_{j-1} = \mathbf{Z}x$, and let $\bar{x}$ be the component of x orthogonal to w. Then one has $l_{j+1}(\mathfrak{F})^2 = q(\bar{x}) \leq q(x) = l_j(\mathfrak{F}')^2$, and $l_j(\mathfrak{F}')^2 = q(x) < q(w) = l_j(\mathfrak{F})^2$. This proves the first two inequalities. The last two follow from these and the equality $l_j(\mathfrak{F})l_{j+1}(\mathfrak{F}) = d(F_{j+1})/d(F_{j-1}) = l_j(\mathfrak{F}')l_{j+1}(\mathfrak{F}')$. □

Note in particular that $l_1(\mathfrak{F}') \leq l_1(\mathfrak{F})$ in the situation of proposition 7.2, and, dually, $l_n(\mathfrak{F}') \geq l_n(\mathfrak{F})$.

Proposition 7.3. *Let L be a lattice and let $\mathfrak{F}$ be a flag of L. Let j be an integer with $0 < j < \operatorname{rk} L$ and c a real number with $c \geq \frac{4}{3}$. Suppose that one has $l_{j+1}(\mathfrak{F})^2 < c^{-1} \cdot l_j(\mathfrak{F})^2$. Then there is a step $\mathfrak{F} \rightarrow \mathfrak{F}'$ in $\Gamma(L)$ with colour j and length greater than $\frac{4c}{c+4}$.*

Proof. Write $F_j/F_{j-1} = \mathbf{Z}w$, and choose $x \in F_{j+1}/F_{j-1}$ with $q(x)$ minimal subject to the condition $\mathbf{Z}x + \mathbf{Z}w = F_{j+1}/F_{j-1}$. With $\bar{x}$ as in the previous proof, we have $x = \bar{x} + rw$ with $|r| \leq \frac{1}{2}$, so $q(x) = q(\bar{x}) + r^2 q(w) = l_{j+1}(\mathfrak{F})^2 + r^2 l_j(\mathfrak{F})^2 < (\frac{1}{c} + \frac{1}{4}) l_j(\mathfrak{F})^2 = \frac{c+4}{4c} q(w) \leq q(w)$. The proposition follows, with $\mathfrak{F} = (F'_i)_{i=0}^{\operatorname{rk} L}$ defined by (i) and $F'_j/F_{j-1} = \mathbf{Z}x$. □

The expression $\frac{4c}{c+4}$ will reappear in the next section. It is increasing as a function of c, equal to 1 for $c = \frac{4}{3}$, and it tends to 4 for $c \to \infty$. The parameter y that appears in [3] may be viewed as the inverse of $\frac{4c}{c+4}$. A popular choice is $c = 2$, $\frac{4c}{c+4} = \frac{4}{3}$, $y = \frac{3}{4}$.

8. Paths in the Reduction Graph

Let L be a lattice, and put $n = \operatorname{rk} L$. A *path* in $\Gamma(L)$ is a finite sequence $\mathfrak{F}_1 \to \mathfrak{F}_2 \to \cdots \to \mathfrak{F}_t$ of steps $\mathfrak{F}_i \to \mathfrak{F}_{i+1}$ $(1 \le i < t)$ in $\Gamma(L)$; more properly, one could call this a 'directed path', but for 'undirected paths' —which would turn $\Gamma(L)$ into a *connected* graph— we have no use.

Let c be a real number with $c \ge \frac{4}{3}$. Proposition 7.3 leads to the following procedure for transforming a given flag $\mathfrak{F}$ of a lattice into a flag $\mathfrak{F}'$ satisfying the inequalities $l_{i+1}(\mathfrak{F}')^2 \ge c^{-1} \cdot l_i(\mathfrak{F}')^2$ $(0 < i < n)$ from proposition 6.1. If $\mathfrak{F}$ itself does not satisfy these inequalities, then by proposition 7.3 one can take a step of length greater than $\frac{4c}{c+4}$ from $\mathfrak{F}$, and iterate. Since the number of flags of size smaller than $s(\mathfrak{F})$ is finite, this 'flag reduction algorithm' must terminate with a flag $\mathfrak{F}'$ with the required property. In particular, taking $c = \frac{4}{3}$, we see that we proved the statement made at the end of section 6.

A good upper bound for the number of steps to be taken is of obvious interest for the analysis of actual algorithms that may be based on the procedure just described. Such a bound is easy to obtain in the case $c > \frac{4}{3}$. Namely, in that case we have $\frac{4c}{c+4} > 1$, and since the square of the size of the flag decreases by a factor greater than $\frac{4c}{c+4}$ in each step, the number of steps in the path $\mathfrak{F} \to \cdots \to \mathfrak{F}'$ is at most

$$2 \cdot \frac{\log\bigl(s(\mathfrak{F})/s(\mathfrak{F}')\bigr)}{\log\bigl(\frac{4c}{c+4}\bigr)} .$$

This is a satisfactory bound if a good lower bound for $s(\mathfrak{F}')$ is available, which is often the case; for example, if the lattice L is such that $\langle x, y\rangle \in \mathbf{Z}$ for all $x, y \in L$, then one has $d(K)^2 \in \mathbf{Z}$ for all sublattices K of L, so $s(\mathfrak{F}')^2$ is an integer, and $s(\mathfrak{F}') \ge 1$. In general, one has $s(\mathfrak{F}') \ge \prod_{i=1}^{n}(l/i)^{i/2}$ if l is as in proposition 3.1, by the second remark after proposition 4.3.

The argument just given fails in the case $c = \frac{4}{3}$, and more generally if we allow steps of length arbitrarily close to 1. It is, for fixed rank, nevertheless possible to prove a similar logarithmic upper bound for the length of any path $\mathfrak{F} \to \cdots \to \mathfrak{F}'$ in $\Gamma(L)$, as we shall see in proposition 8.2. We first prove an auxiliary result on paths that consist of 'short' steps only.

Proposition 8.1. *For each integer $n \geq 0$ and each real number $c > \frac{4}{3}$ there is a positive integer $A = A(n,c)$ with the following property. Let L be a lattice of rank n, and let $\mathfrak{F}_1 \to \cdots \to \mathfrak{F}_t$ be a path in $\Gamma(L)$ such that each step $\mathfrak{F}_i \to \mathfrak{F}_{i+1}$ has length at most $\frac{4c}{c+4}$. Then one has $t \leq A$.*

Proof. The proof is by induction on n, the case $n \leq 1$ being trivial. Suppose that $n \geq 2$, and consider a path as in proposition 8.1. We first show that there exists $m \in \{1, 2, \ldots, n\}$ with the following two properties:

(i) $l_{i+1}(\mathfrak{F}_1)^2 \geq c^{-1} \cdot l_i(\mathfrak{F}_1)^2$ for $0 < i < m$;

(ii) none of the steps $\mathfrak{F}_j \to \mathfrak{F}_{j+1}$ in the path has colour m.

If all $i = 1, \ldots, n-1$ satisfy the inequality in (i) then we can clearly take $m = n$. Now suppose that $i \in \{1, \ldots, n-1\}$ is such that $l_{i+1}(\mathfrak{F}_1)^2 < c^{-1} \cdot l_i(\mathfrak{F}_1)^2$. Then by proposition 7.3, there is a step $\mathfrak{F}_1 \to \mathfrak{F}'$ of colour i and length greater than $\frac{4c}{c+4}$, so *any* step of colour i starting at $\mathfrak{F}_1$ has length greater than $\frac{4c}{c+4}$. By hypothesis, $\mathfrak{F}_1 \to \mathfrak{F}_2$ has length at most $\frac{4c}{c+4}$, so it does not have colour i. Therefore proposition 7.2 implies $l_{i+1}(\mathfrak{F}_2) \leq l_{i+1}(\mathfrak{F}_1)$ and $l_i(\mathfrak{F}_2) \geq l_i(\mathfrak{F}_1)$. It follows that the inequality $l_{i+1}(\mathfrak{F})^2 < c^{-1} \cdot l_i(\mathfrak{F})^2$, which is satisfied for $\mathfrak{F} = \mathfrak{F}_1$, is likewise satisfied for $\mathfrak{F} = \mathfrak{F}_2$; by induction on j one now deduces that all steps $\mathfrak{F}_j \to \mathfrak{F}_{j+1}$ in the path have colour different from i and that all $\mathfrak{F} = \mathfrak{F}_j$ satisfy the inequality just stated. Hence $m = i$ satisfies (ii). If we take for m the *least* value of i violating the inequality in (i), then (i) is satisfied as well.

Write F_{j1} for the rank 1 lattice belonging to $\mathfrak{F}_j$. We claim that *among $F_{11}, \ldots, F_{t1}$ there are at most $\big((1 + 2c^{(n-1)/2})^n - 1\big)/2$ different rank 1 lattices.* To prove this, let m be as above. By (ii), the rank m sublattice belonging to $\mathfrak{F}_j$ is the same for all j; let this lattice be called K. The lattices of rank at most m belonging to $\mathfrak{F}_1$ form a flag $\mathfrak{C}$ of K, and by (i) we have $l_{i+1}(\mathfrak{C}) \geq c^{-1} \cdot l_i(\mathfrak{C})^2$ for $0 < i < m$. Applying proposition 6.1 to K and $\mathfrak{C}$ we see that any nonzero $x \in K$ satisfies $q(x)^2 \geq l = c^{1-m} \cdot l_1(\mathfrak{F}_1)^2$. By proposition 3.1, the number of $x \in K$ with $q(x) \leq l_1(\mathfrak{F}_1)^2$ is at most $(1 + 2c^{(m-1)/2})^m$. Since x and $-x$ generate the same lattice, it follows that K has at most $\big((1 + 2c^{(m-1)/2})^m - 1\big)/2$ sublattices M of rank 1 that satisfy $d(M) \leq l_1(\mathfrak{F}_1)$. Each F_{j1} is such an M, and $m \leq n$, so the claim follows.

In our path, the rank 1 sublattice changes only at steps of colour 1, and at each such step the determinant of that sublattice decreases. Hence our claim implies that we can write the path $\mathfrak{F}_1 \to \cdots \to \mathfrak{F}_t$ as the union of at most $\big((1 + 2c^{(n-1)/2})^n - 1\big)/2$ subpaths connected by steps of colour 1, such that in each of the subpaths the rank 1 sublattice is held fixed. But when the rank 1 sublattice is held equal to M (say), one is really considering flags of the rank $n-1$ lattice L/M and paths in $\Gamma(L/M)$. Application of the induction hypothesis on n now leads in a straightforward way to the inequality in proposition 8.1, with $A(n,c) = A(n-1,c) \cdot [\big((1 + 2c^{(n-1)/2})^n - 1\big)/2]$. □

We can now formulate and prove our main result.

Proposition 8.2. *For each non-negative integer n there exists a positive integer $B = B(n)$ with the following property. If L is a lattice of rank n, and $\mathfrak{F}$, $\mathfrak{F}'$ are flags of L, then every path from $\mathfrak{F}$ to $\mathfrak{F}'$ in $\Gamma(L)$ contains at most $B \cdot \bigl(1 + \log(s(\mathfrak{F})/s(\mathfrak{F}'))\bigr)$ flags.*

Proof. Fix a real number c with $c > \frac{4}{3}$, and call a step $\mathfrak{F}_1 \to \mathfrak{F}_2$ in $\Gamma(L)$ *long* if it has length greater than $\frac{4c}{4+c}$, and *short* otherwise.

Consider any path $\mathfrak{F} \to \cdots \to \mathfrak{F}'$. If k is the number of long steps, then one has $s(\mathfrak{F}')^2 \le \left(\frac{4+c}{4c}\right)^k \cdot s(\mathfrak{F})^2$, so $k \cdot \log \frac{4c}{4+c} \le 2\log\bigl(s(\mathfrak{F})/s(\mathfrak{F}')\bigr)$. Hence the path is the union of at most $1 + 2\bigl(\log(s(\mathfrak{F})/s(\mathfrak{F}'))\bigr)/\log\frac{4c}{4+c}$ subpaths connected by long steps, such that each of the subpaths consists of short steps only. By proposition 8.1, the number of flags occurring in each of the subpaths is bounded by a function of the rank. The result follows. □

The proposition just proved is useful in the analysis of algorithms that involve lattices of fixed rank. When the rank varies, it becomes important to express $B(n)$ as an explicit function of n; in particular, if one wishes such an algorithm to run in polynomial time, one may want to bound $B(n)$ by a polynomial function of n. I do not know whether this is possible. I do know the following much weaker result.

Proposition 8.3. *The numbers $B(n)$ in proposition 8.2 can be chosen such that in addition one has $B(n) = (4/3)^{n^3/(12+o(1))}$ for $n \to \infty$.*

Proof. Making the proof of proposition 8.2 explicit, one finds a value for $B(n)$ that is a function of c. One may choose c as a function of n that tends to $\frac{4}{3}$ for $n \to \infty$ sufficiently slowly for the factor $\log\frac{4c}{4+c}$ to be $\left(\frac{4}{3}\right)^{o(n)}$; for example, one may take $c = \frac{4}{3} + \frac{1}{n}$. This yields the result of proposition 8.3, but with 6 instead of 12. To achieve 12, one starts by improving proposition 8.1. In the proof of proposition 8.1, we saw that the flags in a path $\mathfrak{F}_1 \to \cdots \to \mathfrak{F}_t$ consisting of short steps only comprise at most $\bigl((1+2c^{(n-1)/2})^n - 1\bigr)/2$ different sublattices M of rank 1. One now notes that, by duality, they also comprise at most $\bigl((1+2c^{(n-1)/2})^n - 1\bigr)/2$ different sublattices N of rank $n-1$. It follows that there are at most $(1+2c^{(n-1)/2})^n - 3$ steps of colour 1 or $n-1$, and that the path is the union of at most $(1+2c^{(n-1)/2})^n - 2$ subpaths, connected by steps of colours 1 and $n-1$, such that in each of the subpaths both N and M are fixed; it is then really a path in $\Gamma(N/M)$, where N/M has rank $n-2$. In this manner, one proves that one may take $A(n,c) = A(n-2,c) \cdot \bigl[(1+2c^{(n-1)/2})^n - 2\bigr]$ in proposition 8.1. This improved bound leads to proposition 8.3. □

References

[1] K. I. Aardal, *Lattice basis reduction and integer programming*, rapport UU-CS-1999-37, Informatica Instituut, Universiteit Utrecht, 1999.

[2] H. Cohen, *A course in computational algebraic number theory*, Springer-Verlag, Berlin, 1993.

[3] A. K. Lenstra, H. W. Lenstra, Jr. and L. Lovász, *Factoring polynomials with rational coefficients*, Math. Ann. **261** (1982), 515–534.

[4] H. W. Lenstra, Jr., *Integer programming with a fixed number of variables*, report 81-03, Mathematisch Instituut, Universiteit van Amsterdam, April, 1981.

[5] H. W. Lenstra, Jr., *Integer programming with a fixed number of variables*, Math. Oper. Res. **8** (1983), 538–548.

Mathematisch Instituut,
Universiteit Leiden,
Postbus 9512,
2300 RA Leiden, The Netherlands
E-mail address: hwl@math.leidenuniv.nl

Department of Mathematics # 3840,
University of California,
Berkeley, CA 94720-3840, U. S. A.