JOURNAL de Théorie des Nombres de BORDEAUX

anciennement Séminaire de Théorie des Nombres de Bordeaux

Nickolas ANDERSEN et William DUKE

The Minkowski chain and Diophantine approximation

Tome 32, n° 2 (2020), p. 503-523.

<http://jtnb.centre-mersenne.org/item?id=JTNB_2020__32_2_503_0>

Journal de Théorie des Nombres
de Bordeaux **32** (2020), 503–523

The Minkowski chain and Diophantine approximation

par Nickolas ANDERSEN et William DUKE

Résumé. La chaîne de Hurwitz donne une suite de paires d'approximations de Farey d'un nombre réel irrationnel. Minkowski a donné un critère d'algébraicité d'un nombre en utilisant une certaine généralisation de la chaîne de Hurwitz. Nous appliquons cette généralisation (la chaîne de Minkowski) pour donner des critères pour qu'une forme linéaire réelle soit mal approchable ou singulière. Les preuves reposent sur des propriétés des minima successifs et des bases réduites de réseaux.

Abstract. The Hurwitz chain gives a sequence of pairs of Farey approximations to an irrational real number. Minkowski gave a criterion for a number to be algebraic by using a certain generalization of the Hurwitz chain. We apply Minkowski's generalization (the Minkowski chain) to give criteria for a real linear form to be either badly approximable or singular. The proofs rely on properties of successive minima and reduced bases of lattices.

1. Introduction

Every irrational $\alpha \in \mathbb{R}$ has a unique expansion as an infinite regular continued fraction

$$\alpha = a_0 + \frac{1}{a_1+} \, \frac{1}{a_2+} \, \frac{1}{a_3+} \cdots$$

where a_j are integers called the partial quotients of α with $a_j > 0$ for $j \geq 1$. A striking result of elementary number theory, going back to Euler and Lagrange, is that α is algebraic of degree two over $\mathbb{Q}$ if and only if this expansion is eventually periodic.

More generally, suppose that $\alpha \in \mathbb{R}$ is such that $\{\alpha^n, \alpha^{n-1}, \ldots, \alpha, 1\}$ are linearly independent over $\mathbb{Q}$. The $n = 1$ case above leads naturally to the following problem. Find an algorithm, like the regular continued fraction, which provides a criterion for α to be algebraic of degree $\ell = n + 1$ over $\mathbb{Q}$. Since Jacobi [13], most investigations of multi-dimensional generalizations of continued fractions, as applied to characterizing algebraic numbers, have concentrated on periodicity. This approach has had only limited success.

Manuscrit reçu le 15 janvier 2020, révisé le 13 mai 2020, accepté le 3 juin 2020.
2020 *Mathematics Subject Classification.* 11J13, 11J70.
Mots-clefs. Minkowski chain, Diophantine approximation.
Supported by NSF grant DMS 1701638. The second author was also supported by the Simons Foundation: Award Number 554649.

However, already in 1899 Minkowski [19][1] found such an algorithm that produces a sequence of nonsingular $\ell \times \ell$ integral matrices, the Minkowski chain, which characterizes algebraic α not through periodicity but rather a certain finiteness condition. The Minkowski chain generalizes the Hurwitz chain, itself a refinement of the regular continued fraction. In a speech appearing as the preface to Minkowski's collected papers,[2] Hilbert said that "Der Minkowskische Algorithmus ist nicht ganz einfach...". One goal of our paper is to revive interest in the Minkowski chain and its applications. In particular, Minkowsi's criterion for an algebraic number has not received the attention we think it deserves. Another goal is to supplement Minkowsi's criterion by characterizing badly approximable and singular real linear forms in several variables in terms of the Minkowski chain.

In the next section we recall the definitions of the Hurwitz and Minkowski chains, formulate their relationships to each other and to the regular continued fraction and state Minkowski's criterion. We also give some illustrative examples. Then in Section 3 we state our results on Diophantine approximations by linear forms. The remainder of the paper contains the proofs. We have tried to make the presentation as self-contained as is feasible and we provide proofs of all numbered theorems, corollaries and lemmas.

2. The Minkowski chain

Suppose that $\alpha \in (0,1)$ is irrational. A natural way to approximate α by rational numbers, while controlling the size of the denominators, is to use Farey fractions. For $m \in \mathbb{Z}^+$ let $\mathcal{F}_m$ be the m^{th} Farey set, which consists of all rational numbers in $[0,1]$ in increasing order whose denominators are at most m. Thus

$$\mathcal{F}_1 = \{\tfrac{0}{1}, \tfrac{1}{1}\}, \quad \mathcal{F}_2 = \{\tfrac{0}{1}, \tfrac{1}{2}, \tfrac{1}{1}\}, \quad \mathcal{F}_3 = \{\tfrac{0}{1}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{1}{1}\},$$
$$\mathcal{F}_4 = \{\tfrac{0}{1}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{1}{1}\}, \dots$$

For a fixed m let $(\frac{p}{q}, \frac{p'}{q'})$ be the unique pair of successive Farey fractions in $\mathcal{F}_m$ with $\frac{p}{q} < \alpha < \frac{p'}{q'}$. After $m = 2$ the pair of surrounding fractions might not change as m increases to $m+1$, but when it does one fraction will remain and the new one will be $\frac{p+p'}{q+q'}$. This process was studied in some detail by Hurwitz [12] in 1894 and the sequence of (distinct) Farey pairs is called the Hurwitz chain for α by Philippon in [23].

We can encode the Hurwitz chain of an irrational $\alpha \in (0,1)$ by a unique infinite word in the letters R and L. We label a pair with R if within the pair the old fraction is to the right of the new one and L if it is to the left.

[1] A translation (with additions) of this paper into English is given in Vol. 1 Chap. IX of [11].
[2] See p. XV. of Vol. I of the Gesammelte Abhandlungen.

We label the first pair $(\frac{0}{1}, \frac{1}{1})$ with L and the next with R if it is $(\frac{1}{2}, \frac{1}{1})$ and with L if it is $(\frac{0}{1}, \frac{1}{2})$.

For example, the Hurwitz chain for $\alpha = \frac{1}{2}(-1+\sqrt{5})$ begins

$$(\tfrac{0}{1},\ \tfrac{1}{1}),\ (\tfrac{1}{2}, \tfrac{1}{1}),\ (\tfrac{1}{2}, \tfrac{2}{3}),\ (\tfrac{3}{5}, \tfrac{2}{3}),\ (\tfrac{3}{5}, \tfrac{5}{8}),\ (\tfrac{8}{13}, \tfrac{5}{8}), \dots \tag{2.1}$$

with corresponding word $LRLRLR\dots$.

The word corresponding to the Hurwitz chain for $\alpha \in (0,1)$ determines the partial quotients a_j in the regular continued fraction

$$\alpha = \frac{1}{a_1+}\ \frac{1}{a_2+}\ \frac{1}{a_3+}\cdots. \tag{2.2}$$

It follows from standard properties of the convergents of the continued fraction that a_j is given by the number of successive L's or R's in the j^{th} block of the word. Thus the partial quotients for $\alpha = \frac{1}{2}(-1+\sqrt{5})$ are $a_j = 1$ for all j. Clearly α is quadratic over $\mathbb{Q}$ if and only if the word associated to the Hurwitz chain for α is eventually periodic.

Minkowski discovered that to detect algebraic numbers of degree greater than two it is better to abandon periodicity. His algorithm is readily described. We give it in a slightly generalized form that we need later. Suppose that $(\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{R}^n$ is such that $\{\alpha_1, \dots, \alpha_n, 1\}$ are linearly independent over $\mathbb{Q}$. Set

$$\ell = n + 1.$$

Define for any real matrix $A = (a_{i,j})$ the norm $\|A\|_\infty = \max(|a_{i,j}|)$. For $\ell \geq 2$ and $m \in \mathbb{Z}^+$ let $\mathcal{A}_m$ consist of all integral $\ell \times \ell$ matrices A with $\det A \neq 0$ and $\|A\|_\infty \leq m$. Write

$$A(\alpha_1, \dots, \alpha_n, 1)^\top = (\beta_1, \beta_2, \dots, \beta_\ell)^\top. \tag{2.3}$$

Let $\mathcal{A}_{m,1} \subset \mathcal{A}_m$ be those $A \in \mathcal{A}_m$ that minimize $\|A(\alpha_1, \dots, \alpha_n, 1)^\top\|_\infty$ and for which the minimum is $|\beta_1|$. This fixes the first row of A by the linear independence assumption, provided we make some sign convention, for example that the first non-zero entry in the first row is positive. Next let $\mathcal{A}_{m,2} \subset \mathcal{A}_{m,1}$ be those $A \in \mathcal{A}_{m,1}$ for which $|\beta_2|$ gives the minimal value thereby with the corresponding convention fixing the second row of A. Continue this process of defining rows of A. Then $\mathcal{A}_{m,\ell}$ contains exactly one element which we will call A_m. The matrices A_m need not change as m goes to $m+1$. Let $B_k = A_{m_k}$, where $k = 1, 2, \dots$, define the subsequence of distinct matrices starting with $B_1 = A_1$. The sequence $\{B_1, B_2, \dots\}$ of matrices is what we will call the *Minkowski chain* for $(\alpha_1, \dots, \alpha_n)$.

When $n = 1$ and $\alpha \in (0,1)$ the Minkowski chain corresponds to the Hurwitz chain for α. More precisely, we have the following result.

Theorem 1. *Let the k^{th} matrix in the Minkowski chain for an irrational $\alpha \in (0,1)$ be*

$$B_k = \begin{pmatrix} q & -p \\ q' & -p' \end{pmatrix}.$$

Then the k^{th} pair in the Hurwitz chain is either $(\frac{p}{q}, \frac{p'}{q'})$ or $(\frac{p'}{q'}, \frac{p}{q})$.

An immediate corollary is the following fact which, as far as we know, need not hold in general for $n > 1$.

Corollary 1. *When $n = 1$ we have that $|\det B_k| = 1$ for all k.*

Let any $\ell \times \ell$ matrix $B = (b_{i,j})$ act on an n-tuple $(x_1, x_2, \dots, x_n)$ projectively as a linear fractional map:

$$B[(x_1, \dots, x_n)] = \left(\frac{\sum_{j=1}^n b_{1,j}x_j + b_{1,\ell}}{\sum_{j=1}^n b_{\ell,j}x_j + b_{\ell,\ell}}, \dots, \frac{\sum_{j=1}^n b_{n,j}x_j + b_{n,\ell}}{\sum_{j=1}^n b_{\ell,j}x_j + b_{\ell,\ell}} \right).$$

For each $k \in \mathbb{Z}^+$ set

$$B_k[(\alpha_1, \dots, \alpha_n)] = (\alpha_{k,1}, \dots, \alpha_{k,n}) = \left(\frac{\beta_{k,1}}{\beta_{k,\ell}}, \dots, \frac{\beta_{k,n}}{\beta_{k,\ell}} \right) \tag{2.4}$$

where B_k is the k^{th} matrix in the Minkowski chain for $(\alpha_1, \dots, \alpha_n)$ and where

$$B_k(\alpha_1, \dots, \alpha_n, 1)^\top = (\beta_{k,1}, \beta_{k,2}, \dots, \beta_{k,\ell})^\top.$$

Clearly we have that

$$0 < |\alpha_{k,1}| < |\alpha_{k,2}| < \cdots < |\alpha_{k,n}| < 1.$$

We are mostly interested in properties of the sequence $\{(\alpha_{k,1}, \dots, \alpha_{k,n})\}_{k \geq 1}$ of n-tuples attached to $(\alpha_1, \dots, \alpha_n)$. Minkowski realized that it is the finiteness of the set of n-tuples $B_k[(\alpha^n, \dots, \alpha)]$, rather than periodicity determined by the chain, which characterizes algebraic α of degree ℓ.

Theorem 2 (Minkowski [19])**.** *Suppose that $\alpha \in \mathbb{R}$ and that $\{\alpha^n, \alpha^{n-1}, \dots, \alpha, 1\}$ are linearly independent over $\mathbb{Q}$. Then α is algebraic of degree $\ell = n+1$ over $\mathbb{Q}$ if and only if the sequence $\{B_k[(\alpha^n, \alpha^{n-1}, \dots, \alpha)]\}_{k \geq 1}$ contains only finitely many different n-tuples.*

Actually, Minkowski's formulation allows α to be complex. He also did not assume that $\{\alpha^n, \dots, \alpha, 1\}$ are linearly independent over $\mathbb{Q}$, but by using the algorithm with smaller n we may assume this without any loss and with uniqueness of the expansion. In another paper [20] he considered when there can exist subsequences of $\{B_{k+1}B_k^{-1}\}$ that are eventually periodic and found that for real α this is possible only when $\mathbb{Q}(\alpha)$ is quadratic or real cubic with negative discriminant.

Examples.

(i) The Minkowski chain for $\alpha = \frac{-1+\sqrt{5}}{2}$ is

$$B_1 = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right), \quad B_2 = \left(\begin{smallmatrix} 2 & -1 \\ 1 & -1 \end{smallmatrix}\right), \quad B_3 = \left(\begin{smallmatrix} 3 & -2 \\ 2 & -1 \end{smallmatrix}\right), \ldots, B_k = \left(\begin{smallmatrix} F_{k+1} & -F_k \\ F_k & -F_{k-1} \end{smallmatrix}\right), \ldots,$$

which corresponds to (2.1). Here F_k is the k^{th} Fibonnaci number and for each k

$$B_k[(\tfrac{-1+\sqrt{5}}{2})] = \tfrac{1-\sqrt{5}}{2}.$$

(ii) Let $\theta = 2\cos\left(\frac{2\pi}{7}\right)$ so that $\mathbb{Q}(\theta)$ is the real cubic field of discriminant 49, i.e. the splitting field of $x^3 + x^2 - 2x - 1$. The Minkowski chain for (θ^2, θ) begins

$$B_1 = \begin{pmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & -1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & -2 & 1 \\ 2 & -1 & -2 \\ 0 & 1 & -1 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 1 & -2 & 1 \\ 3 & -3 & -1 \\ 2 & 0 & -3 \end{pmatrix},$$

$$B_4 = \begin{pmatrix} 1 & 2 & -4 \\ 1 & -2 & 1 \\ 3 & -3 & -1 \end{pmatrix}, \ldots.$$

By Theorem 2 we know that the set of values $\{B_k[(\theta^2, \theta)]\}$ is finite. Among the first 30 terms there are only six distinct pairs up to sign, namely

$$(0.15883\ldots, 0.64310\ldots), \quad (0.24698\ldots, 0.55496\ldots), \quad (0.35690\ldots, 0.44504\ldots),$$
$$(0.44504\ldots, 0.80194\ldots), \quad (0.55496\ldots, 0.69202\ldots), \quad (0.64310\ldots, 0.80194\ldots).$$

(iii) Suppose that α is transcendental, so $\{\alpha^n, \ldots, \alpha, 1\}$ are linearly independent over $\mathbb{Q}$ for any positive integer n. For a fixed n let $B_k[(\alpha^n, \ldots, \alpha)] = (\alpha_{k,1}, \ldots, \alpha_{k,n})$ come from the Minkowski chain for $(\alpha^n, \ldots, \alpha)$ as above. By Theorem 2 we know that

$$\{(\alpha_{k,1}, \ldots, \alpha_{k,n})\}_{k \geq 1}$$

is an infinite set.

Recall that $\alpha \in \mathbb{R}$ is a Liouville number if, for every positive integer m, there exist infinitely many relatively prime integers p, q with $q > 0$ such that

$$0 < |\alpha - \tfrac{p}{q}| < q^{-m}.$$

Liouville's theorem on Diophantine approximation implies that a Liouville number α is transcendental. If α is a Liouville number and $n \in \mathbb{Z}^+$ is fixed, Theorem 3 below implies that not only is $\{(\alpha_{k,1}, \ldots, \alpha_{k,n})\}_{k \geq 1}$ infinite, but also $|\alpha_{k,1}|$ gets arbitrarily close to zero as $k \to \infty$. For the Liouville constant

$$\lambda = \sum_{m \geq 1} 10^{-m!} = 0.11000100000000000000000100\ldots$$

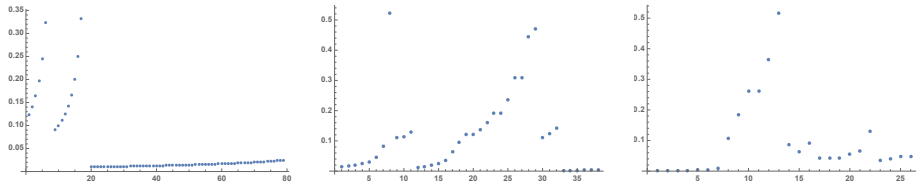

FIGURE 2.1. The sequences $|\lambda_{k,1}|$ for $n = 1, 2, 3$.

and the cases $n = 1, 2, 3$, the behavior of $|\lambda_{k,1}|$ is shown in Figure 2.1.

3. Applications to Diophantine approximation

We now turn to the application of the Minkowski chain to Diophantine approximation by badly approximable and by singular real linear forms in two or more variables.

Associate to any $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ the linear form[3]

$$L_\alpha(x) = \alpha_1 x_1 + \cdots + \alpha_n x_n.$$

Those $\{\alpha_1, \dots, \alpha_n, 1\}$ that give a basis over $\mathbb{Q}$ for a real number field have the following well-known Diophantine approximation property [25, Thm. 4A p. 42]. There is a constant $c = c_\alpha > 0$ so that for any non-zero $q = (q_1, \dots, q_n) \in \mathbb{Z}^n$

$$\|L_\alpha(q)\| \geq c\|q\|_\infty^{-n}. \tag{3.1}$$

Here $\|t\|$ denotes the distance from a real t to the nearest integer. For any $\alpha \in \mathbb{R}^n$ if the form $L_\alpha(x)$ satisfies (3.1) then L_α is said to be *badly approximable*. For simplicity we shall also sometimes say that α is badly approximable. It is known that the set of all badly approximable $\alpha \in \mathbb{R}^n$ has Lebesgue measure zero [15] yet has full Hausdorff dimension n, hence includes α for which $\{\alpha_1, \dots, \alpha_n, 1\}$ does not give a $\mathbb{Q}$-basis for a number field [24]. For more on the history of these results see [25] and its references.

A natural problem presents itself; can we formulate a criterion for a real linear form in n variables to be badly approximable using the Minkowski chain?

Theorem 3. *Suppose that $\{\alpha_1, \dots, \alpha_n, 1\}$ are linearly independent over $\mathbb{Q}$ and that $\alpha_{k,1}$ is given by the Minkowski chain for α. Then the form L_α is badly approximable if and only if $|\alpha_{k,1}|$ is bounded away from 0.*

[3] Our abuse of notation in using α as an n-tuple of real numbers and as a number, depending on the context, is convenient and should not cause confusion.

Theorem 3 generalizes the well-known characterization of badly approximable numbers in case $n = 1$ that an irrational $\alpha \in \mathbb{R}$ is badly approximable if and only if the partial quotients in its regular continued fraction expansion are bounded (see Lemma 4.2). For the standard direct proof see [25, Thm. 5F p. 22].

For a Liouville number α it is well-known that $L_{(\alpha^n,\ldots,\alpha)}$ is not badly approximable for any n (see e.g. [1]). The claim made in example (iii) from the previous section, that for any fixed positive integer n the value $|\alpha_{k,1}|$ gets arbitrarily close to zero as $k \to \infty$, is thus a consequence of Theorem 3.

An important property of any badly approximable L_α, discovered by Davenport and Schmidt, is that Dirichlet's approximation theorem can be improved for it (see [9]). By this we mean that there exists a $\mu < 1$ such that for every sufficiently large Q there exists a $q \in \mathbb{Z}^n$ such that

$$0 < \|q\|_\infty \leq Q \quad \text{and} \quad \|L_\alpha(q)\| \leq \mu Q^{-n}.$$

When $n = 1$ every irrational α for which Dirichlet's theorem can be improved is badly approximable. However, for $n > 1$ there exist L_α with $\{\alpha_1, \ldots, \alpha_n, 1\}$ linearly independent over $\mathbb{Q}$ that are not badly approximable but for which Dirichlet's theorem can be improved. In fact, Dirichlet's theorem can sometimes be "infinitely improved". More precisely say L_α (or α) is *singular* if for any $\varepsilon > 0$ there is a Q_ε so that if $Q \geq Q_\varepsilon$ there is a $q \in \mathbb{Z}^n$ with

$$0 < \|q\|_\infty \leq Q \quad \text{such that} \quad \|L_\alpha(q)\| \leq \varepsilon\, Q^{-n}.$$

Such forms are clearly not badly approximable. Starting with work of Khinchine [14] it is known that singular L_α with $\{\alpha_1, \ldots, \alpha_n, 1\}$ linearly independent over $\mathbb{Q}$ exist if $n > 1$. It has recently been shown that when $n > 1$ the set of singular $\alpha \in \mathbb{R}^n$ has Hausdorff dimension $\frac{n^2}{n+1}$ (see [3], [4], and [6]). The Minkowski chain also gives a criterion for L_α to be singular.

Theorem 4. *Suppose that $\{\alpha_1, \ldots, \alpha_n, 1\}$ are linearly independent over $\mathbb{Q}$. Then the form L_α is singular if and only if $|\alpha_{k,1}| \to 0$ as $k \to \infty$.*

Remarks.

(i) Theorems 3 and 4 differ substantially from the dynamical criteria for bad approximability and singularity given (more generally for systems of forms) by Dani [5]. Roughly speaking, he showed that badly approximable systems of forms correspond to certain bounded trajectories in the space of unimodular lattices while singular systems correspond to divergent trajectories. A version of these criteria in the case of a single form is one ingredient in our proofs of Theorems 3 and 4. More generally, the fact that the $\beta_{k,j}$ coming from Minkowski's algorithm are not the successive minima of the natural corresponding convex body must be dealt with. This kind of

problem is familiar in the subject and is treated in the parametric geometry of numbers (see [26]). For completeness we give independent proofs.

(ii) In all cases that we have checked numerically, each matrix in the Minkowski chain has been in $\mathrm{GL}(\ell, \mathbb{Z})$. In connection with this as a possible general property see [17].

(iii) It was shown by Khintchine [14] that the linear form L_α is badly approximable if and only if α is a badly approximable n-tuple. See also [25, p. 100]. Davenport and Schmidt [8, Thm. 2] gave the deeper result that Dirichlet's theorem can be improved for the linear form L_α if and only if it can be improved in the form of simultaneous approximation of α by n rationals.

4. The Hurwitz chain

In this section we prove Theorem 1 and a lemma relating the partial quotients of $\alpha \in (0,1)$ to the Minkowski chain of α when $n = 1$. We require an elementary lemma about Farey fractions. We always assume that rational fractions are in lowest form.

Lemma 4.1. *Suppose that $\frac{p}{q} < \frac{p'}{q'}$ is a pair of successive fractions in $\mathcal{F}_m$ and that $\alpha \in (\frac{p}{q}, \frac{p'}{q'})$ is irrational. Then*

(i) *$|q\alpha - p| < |q'\alpha - p'|$ if and only if $\alpha \in (\frac{p}{q}, \frac{p+p'}{q+q'})$*

(ii) *The fraction $\frac{p+p'}{q+q'}$ is the unique fraction with the smallest denominator greater than m that is closer to α than at least one of $\frac{p}{q}, \frac{p'}{q'}$.*

Proof.

(i). If $\alpha \in (\frac{p}{q}, \frac{p+p'}{q+q'})$ then $|\alpha - \frac{p'}{q'}| > |\frac{p+p'}{q+q'} - \frac{p'}{q'}| = \frac{1}{q'(q+q')}$ so $|q'\alpha - p'| > \frac{1}{(q+q')}$. Similarly $|q\alpha - p| < \frac{1}{(q+q')}$ so $|q\alpha - p| < |q'\alpha - p'|$ in this case. The converse is similar using $\alpha \in (\frac{p+p'}{q+q'}, \frac{p'}{q'})$.

(ii). It is well-known (see e.g. [25, p. 4]) that $\frac{p+p'}{q+q'}$ is the unique fraction with the smallest denominator greater than m that is between $\frac{p}{q}$ and $\frac{p'}{q'}$. Thus we need only show that $\frac{p+p'}{q+q'}$ is closer to α than any other $\frac{p''}{q''}$ with $m < q'' \le q + q'$ and either $\frac{p''}{q''} < \frac{p}{q}$ or $\frac{p''}{q''} > \frac{p'}{q'}$.

Suppose that $\frac{p''}{q''} < \frac{p}{q}$. If $\alpha > \frac{p+p'}{q+q'}$ we are done so assume that

$$\frac{p}{q} < \alpha < \frac{p+p'}{q+q'}. \tag{4.1}$$

Now

$$|\alpha - \tfrac{p''}{q''}| > |\tfrac{p''}{q''} - \tfrac{p}{q}| \ge \tfrac{1}{q''q} \ge \tfrac{1}{q(q+q')}$$

while by (4.1)

$$|\alpha - \tfrac{p+p'}{q+q'}| < |\tfrac{p+p'}{q+q'} - \tfrac{p}{q}| = \tfrac{1}{q(q+q')}.$$

The case $\frac{p''}{q''} > \frac{p'}{q'}$ is similar. □

Proof of Theorem 1. We want to show that if

$$B_k = \begin{pmatrix} q_k & -p_k \\ q'_k & -p'_k \end{pmatrix}$$

then the k^{th} pair in the Hurwitz chain is either $(\frac{p_k}{q_k}, \frac{p'_k}{q'_k})$ or $(\frac{p'_k}{q'_k}, \frac{p_k}{q_k})$.

This follows by induction on k. It holds for $k = 1$. Suppose it holds for some $k \geq 1$. Thus $\frac{p_k}{q_k}, \frac{p'_k}{q'_k}$ or $\frac{p'_k}{q'_k}, \frac{p_k}{q_k}$ are successive Farey fractions in $\mathcal{F}_m$ where $m = \max(q_k, q'_k)$.

By the definition of the Minkowski chain given around (2.3) we know that the first row of B_k (the one with $|q_k\alpha - p_k|$ minimal) must appear in B_{k+1} as either the first row or the second row. Now (i) of Lemma 4.1 implies that the fraction associated to the row retained is the one retained by the Hurwitz chain.

Thus we must show that the new row of B_{k+1}, say $(q'', -p'')$, is precisely $(q_k + q'_k, -(p_k + p'_k))$. By the definition of the Minkowski chain $q'' > m$ and certainly

$$|q''\alpha - p''| < |q'_k\alpha - p'_k|.$$

Thus $|\alpha - \frac{p''}{q''}| < |\alpha - \frac{p'_k}{q'_k}|$ so by (ii) of Lemma 4.1 we know that $q'' \geq q_k + q'_k$. Now

$$|(q_k + q'_k)\alpha - (p_k + p'_k)| = |(q_k\alpha - p_k) + (q'_k\alpha - p'_k)|. \tag{4.2}$$

Also, $\alpha - \frac{p_k}{q_k}$ and $\alpha - \frac{p'_k}{q'_k}$ have different signs hence so do $q_k\alpha - p_k$ and $q'_k\alpha - p'_k$. By construction of B_k we know that $|q_k\alpha - p_k| < |q'_k\alpha - p'_k|$. Therefore by (4.2) we have that

$$|(q_k + q'_k)\alpha - (p_k + p'_k)| < |q'_k\alpha - p'_k|.$$

It follows that $(q'', -p'') = (q_k + q'_k, -(p_k + p'_k))$. This completes the proof of Theorem 1. □

It is easy to give a formula for the k^{th} pair in the Hurwitz chain for $\alpha \in (0, 1)$ in terms of the partial quotients a_j of α. For a fixed $k \in \mathbb{Z}^+$ write $k = a_1 + \cdots + a_j + a$ where $0 \leq a < a_{j+1}$. Set $R = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $L = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and let $A = L$ if j is even and $A = R$ if j is odd. Then the k^{th} pair in the Hurwitz chain for $\alpha \in (0, 1)$ is given by $(\frac{p_k}{q_k}, \frac{p'_k}{q'_k})$, where

$$\begin{pmatrix} p'_k & p_k \\ q'_k & q_k \end{pmatrix} = L^{a_1} R^{a_2} \cdots A^a. \tag{4.3}$$

Let $b = a$ if $a > 0$ and $b = a_j$ otherwise. Then by Theorem 1 we have for B_k from the Minkowski chain the formula $B_k = MB_{k-b}$, where M is either L^b, R^b, $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) L^b$, or $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) R^b$.

The following consequence of these formulas gives the usual characterization of badly approximable irrationals.

Lemma 4.2. *An equivalent criterion for the boundedness of the partial quotients of an irrational $\alpha \in (0,1)$ is that $|\alpha_{k,1}|$ from the Minkowski chain for α is bounded away from zero.*

Proof. If $M = L^b = \left(\begin{smallmatrix} 1 & 0 \\ b & 1 \end{smallmatrix}\right)$ then

$$\alpha_{k,1} = \frac{q_k\alpha - p_k}{q'_k\alpha - p'_k} = \frac{q_{k-b}\alpha - p_{k-b}}{(q_{k-b}\alpha - p_{k-b}) + b(q'_{k-b}\alpha - p'_{k-b})} = \frac{1}{1 + b\alpha_{k-b,1}}.$$

The other three cases are similar. In each case we see that $|\alpha_{k,1}|$ is bounded below if and only if $\sup\{a_j\}$ is finite. □

Remark. The original paper by Hurwitz [12] is still a good reference for the Hurwitz chain. A modern reference is [23], which also details its relation to semi-regular continued fractions. The dynamical properties of the Hurwitz chain are discussed in [16], where it is called the additive continued fraction.

5. Successive minima

In this section we will give what is essentially Minkowski's proof of Theorem 2. A crucial ingredient is his theorem on successive minima in the geometry of numbers.

For a general norm F on $\mathbb{R}^\ell$ and any full lattice $\Lambda \subset \mathbb{R}^\ell$ let

$$\mu_1 \le \mu_2 \le \cdots \le \mu_\ell$$

be the successive minima of Λ with respect to F. This means that μ_j is the infimum over all $\mu > 0$ such that there are j linearly independent points $v \in \Lambda$ with $F(v) \le \mu$. There exist (not necessarily unique) minimizing vectors $w_1, \dots, w_\ell \in \Lambda$, which means that they are linearly independent and satisfy $F(w_j) = \mu_j$ for $j = 1, \dots, \ell$. Note that $\{w_1, \dots, w_\ell\}$ do not necessarily form a $\mathbb{Z}$-basis for Λ.

The following fundamental result was first proved in [22, Kap. V]. Shorter proofs were given by Davenport [7] and Weyl [28]. See also [2].

Theorem (Minkowski's Theorem on Successive Minima)**.** *Suppose that Λ has determinant one. Then*

$$\mathrm{vol}(\mathcal{B})\mu_1 \cdots \mu_\ell \le 2^\ell,$$

where $\mathrm{vol}(\mathcal{B})$ *is the volume of* $\mathcal{B} = \{x \in \mathbb{R}^\ell; F(x) < 1\}$.

We remark that for the proof of Theorem 2 we can get by with a weaker result that replaces 2^ℓ by a larger constant. In fact, Minkowski gives a proof of this result in his paper with the constant $2^\ell \ell!$ in place of 2^ℓ. See also [25, Cor. 2B p. 88] for a proof with the constant $2^\ell \ell^{\frac{\ell}{2}}$, which is based on the case of an ellipsoid and a theorem of Jordan.

For $r = (q_1, \ldots, q_n, p) \in \mathbb{Z}^\ell$ define

$$\xi(r) \stackrel{\text{def}}{=} p + L_\alpha(q_1, \ldots, q_n). \tag{5.1}$$

For $m \in \mathbb{Z}^+$ recall the integral $\ell \times \ell$ matrix $A_m = (a_{i,j})$ defined above. From (2.3) we have for each $i = 1, \ldots, \ell$ that

$$\beta_i(m) = \beta_i = \xi(a_{i,1}, \ldots, a_{i,\ell}). \tag{5.2}$$

Note that we will often suppress in the notation the dependence of β_i on m (or on k).

Lemma 5.1. *Fix $m \in \mathbb{Z}^+$ and suppose that $r_1, \ldots, r_\ell \in \mathbb{Z}^\ell$ are linearly independent and satisfy $\|r_i\|_\infty \le m$ for $i = 1, \ldots, \ell$. Let them be ordered so that*

$$|\xi(r_1)| \le |\xi(r_2)| \le \cdots \le |\xi(r_\ell)|. \tag{5.3}$$

Then for each $i = 1, \ldots \ell$ we have that

$$|\beta_i| \le |\xi(r_i)|.$$

Proof. For a fixed $m \in \mathbb{Z}^+$ let $w_j = (a_{i,1}, \ldots a_{i,\ell})$ denote the i^{th} row of A_m, which is the integral vector produced by the Minkowski algorithm. Thus for $j = 1, \ldots, \ell$, we know that $|\beta_j|$ gives the smallest value of $|\xi(w)|$ for any $w \in \mathbb{Z}^\ell$ with $\|w\|_\infty \le m$ that is linearly independent of $\{w_1, \ldots, w_{j-1}\}$.

Note that at least $\ell - 1$ of the $\{r_1, \ldots, r_\ell\}$ are independent of w_1 and so each of those r_k satisfies $|\xi(r_k)| \ge |\beta_2|$. At least $\ell - 2$ of the r_k are independent of $\{w_1, w_2\}$ and so these r_k satisfy $|\xi(r_k)| \ge |\beta_3|$. Continue this process until we have at least one r_k that satisfies $|\xi(r_k)| \ge |\beta_\ell|$. By (5.3) we know that this last set of $r's$ must contain r_ℓ and so $|\xi(r_\ell)| \ge |\beta_\ell|$. Working backward we can finish the proof. □

Lemma 5.2. *Fix $m \in \mathbb{Z}^+$ and let $A_m = (a_{i,j})$ and $\beta_1, \ldots, \beta_\ell$ be from the Minkowski algorithm. Let $\Lambda = \mathbb{Z}^\ell$ and define the norm on $\mathbb{R}^\ell$ by*

$$G_m(x_1, \ldots, x_\ell) = \max\left(|x_1|, \ldots, |x_\ell|, \tfrac{m}{|\beta_\ell|}|L_\alpha(x_1, \ldots, x_n) + x_\ell|\right).$$

Let $\mu_1 \le \mu_2 \le \cdots \le \mu_\ell$ be the successive minima of G_m. Then

$$\mu_\ell \ge m. \tag{5.4}$$

Proof. Note that for $r = (q_1, \ldots, q_n, p) \in \mathbb{Z}^\ell$ we have

$$G_m(r) = \max\left(|q_1|, \ldots, |q_n|, |p|, \tfrac{|\xi(r)|m}{|\beta_\ell|}\right) \tag{5.5}$$

where $\xi(r)$ was defined in (5.1). Suppose that $\{r_1, \dots, r_\ell\}$ are independent and such that for each j we have $G_m(r_j) = \mu_j$. By (5.5) we see that if $\|r_j\|_\infty > m$ for any $j = 1, \dots, \ell$ then $\mu_\ell > m$. Otherwise apply Lemma 5.1 to $\{r_1, \dots, r_\ell\}$ to conclude that $|\xi(r_j)| \geq |\beta_j|$ for each $j = 1, \dots, \ell$. Therefore in particular for $j = \ell$ we get by (5.5) again that $\mu_\ell = G_m(r_\ell) \geq m$. Thus in any case we have (5.4). □

Minkowski only proved the following result for $L_{(\alpha,\dots,\alpha^n)}$ where α is algebraic of degree ℓ, but his proof extends naturally.

Lemma 5.3. *Suppose that $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$. If L_α is badly approximable then there are constants $c, C > 0$ depending only on α such that*

$$cm^{-n} < |\beta_1| < \cdots < |\beta_\ell| < Cm^{-n}. \tag{5.6}$$

Proof. In this proof and those that follow we usually name and keep track of constants that depend only on α, even though it would be cleaner to use the $\ll$ or $\gg$ notation. We do this to help the reader verify inequalities.

Fix $m \in \mathbb{Z}^+$ and let $A_m = (a_{i,j})$ and $\beta_1, \dots, \beta_\ell$ be from the Minkowski algorithm. Let now $\Lambda = \mathbb{Z}^\ell$ and G_m the norm on $\mathbb{R}^\ell$ in Lemma 5.2. The form L_α being badly approximable means that there is a $c > 0$ so that

$$|\xi(r)| > c\|q\|_\infty^{-n} \tag{5.7}$$

for all $r = (q_1, \dots, q_n, p) \in \mathbb{Z}^\ell$, where $q = (q_1, \dots, q_n)$. By the definition of β_1 and (5.7) we have that

$$|\beta_1| = \min_{\|r\|_\infty \leq m} |\xi(r)| \geq \min_{\|q\|_\infty \leq m} |\xi(r)| > \tfrac{c}{m^n}. \tag{5.8}$$

Now $G_m(r_1) = \mu_1$ and so (5.5) implies that

$$|\xi(r_1)| \leq \tfrac{\mu_1 |\beta_\ell|}{m} \tag{5.9}$$

and also that $\|r_1\|_\infty \leq \mu_1$. Thus by (5.7) again we also have that

$$|\xi(r_1)| \geq \tfrac{c}{\mu_1^n}. \tag{5.10}$$

By (5.9) and (5.10) we conclude that

$$\Big(\frac{\mu_1^\ell |\beta_\ell|}{m}\Big)^n \geq c^n, \tag{5.11}$$

which is the form we will need.

A straightforward calculation shows that

$$\mathrm{vol}(\{x \in \mathbb{R}^\ell; G_m(x) < 1\}) \geq V \tfrac{|\beta_\ell|}{m},$$

where $V > 0$ is a constant depending only on α. By Minkowski's theorem on successive minima we have

$$V \tfrac{|\beta_\ell|}{m} \mu_1^n \mu_\ell \leq 2^\ell$$

so that using (5.11) we have for $M = 2^{\ell^2} V^{-\ell}$ that

$$c^n \tfrac{|\beta_\ell|}{m} \mu_\ell^\ell \le \left(\tfrac{|\beta_\ell|}{m} \mu_1^n \mu_\ell \right)^\ell \le M.$$

Thus $|\beta_\ell| \le \frac{Cm}{\mu_\ell^\ell}$ for $C = \frac{M}{c^n}$. Finally, from Lemma 5.2 we derive that $|\beta_\ell| \le \frac{C}{m^n}$. Together with (5.8), this finishes the proof of Lemma 5.3. □

Proof of Theorem 2. The proof of the implication *algebraic implies finite* works the same for the Minkowski chain for any $\mathbb{Q}$-basis $\{\alpha_1, \alpha_2, \dots, \alpha_n, 1\}$ of a real number field K. Recall that for each k we have

$$(\beta_1, \dots, \beta_\ell) = B_k[(\alpha_1, \dots, \alpha_n, 1)],$$

where again we suppress the dependence of β_j on k in the notation. Clearly $\beta_j \in K$ and there is a positive integer b such that $b\beta_j$ is an algebraic integer for any j, k, so we must have $N_{K/\mathbb{Q}}(b\beta_j) \ge 1$. Denote by $\{\beta_i^{(j)}; j = 1, \dots, \ell\}$ the set of Galois conjugates of $\beta_i = \beta_i^{(1)}$. Set $C_1 = \max_{j=1,\dots,\ell}(1 + |\alpha_1^{(j)}| + \cdots + |\alpha_n^{(j)}|)$. Clearly

$$|\beta_i^{(j)}| \le C_1 m, \tag{5.12}$$

where $m = m_k$ from the algorithm.

We know that $L_{(\alpha_1,\dots,\alpha_n)}$ is badly approximable so by Lemma 5.3 we have that for each i

$$|\beta_i| \le C m^{-n}. \tag{5.13}$$

Therefore we have that

$$b^{-\ell} \le |N_{K/\mathbb{Q}}(\beta_i)| = |\prod_{j=1}^{\ell} \beta_i^{(j)}| \le C\, C_1^n. \tag{5.14}$$

From the first inequality of (5.14), (5.12) and (5.13) we get that for $k > 1$

$$|\beta_i^{(k)}| \ge C_2 m \tag{5.15}$$

for some constant $C_2 > 0$ depending only on α. Here we have used (5.13) for the first factor in the product and (5.12) for all of the remaining factors except for the k^{th}. Recall from (2.4) that

$$(\alpha_{k,1}, \dots, \alpha_{k,n}) = (\tfrac{\beta_1}{\beta_\ell}, \dots, \tfrac{\beta_n}{\beta_\ell}) \in K^n.$$

Let $\gamma_{k,i} = b^\ell N_{K/\mathbb{Q}}(\beta_\ell)\alpha_{k,i}$. Then $\gamma_{k,i}$ is an algebraic integer in K. From (5.6), (5.12), (5.14) and (5.15) we have for each i, j, k that $|\gamma_{k,i}^{(j)}| \le C_3$ for some $C_3 > 0$ that depends only on α. It follows that there are only finitely many such $(\gamma_{k,1}, \dots, \gamma_{k,n})$ hence only finitely many $(\alpha_{k,1}, \dots, \alpha_{k,n})$.

For the converse, we need to assume that $(\alpha_1, \alpha_2, \dots, \alpha_n) = (\alpha^n, \dots, \alpha)$ and suppose that there are only finitely many values of the sequence

$\{B_k[(\alpha^n,\dots,\alpha)]\}_{k\geq 1}$. Then for some $k'>k$ we have $B_k[(\alpha^n,\dots,\alpha)] = B_{k'}[(\alpha^n,\dots,\alpha)]$. Hence

$$B_k^{-1}B_{k'}[(\alpha^n,\dots,\alpha)] = (\alpha^n,\dots,\alpha),$$

which implies that

$$B_k^{-1}B_{k'}(\alpha^n,\dots,\alpha,1)^\top = \theta(\alpha^n,\dots,\alpha,1)^\top \tag{5.16}$$

for some $\theta\in\mathbb{R}$ with $|\theta|\leq 1$. Write $B_k^{-1}B_{k'} = (c_{i,j})$ so that for each $k=1,\dots,n$ we can write two successive rows of (5.16) as

$$\begin{aligned} c_{k,1}\alpha^n+\cdots+(c_{k,k}-\theta)\alpha^{n-k+1}+\cdots+c_{k,n}\alpha+c_{k,\ell}=0\\ c_{k+1,1}\alpha^n+\cdots+(c_{k+1,k+1}-\theta)\alpha^{n-k}+\cdots+c_{k+1,n}\alpha+c_{k+1,\ell}=0.\end{aligned}$$

Multiply the first equation by α and subtract rows to get for each $k=1,\dots,n$ that

$$\begin{aligned} c_{k,1}\alpha^\ell+(c_{k,2}-c_{k+1,1})\alpha^n+\cdots+(c_{k,k}-c_{k+1,k+1})\alpha^{n-k}\\ +\cdots+(c_{k,\ell}-c_{k+1,n})\alpha-c_{k+1,\ell}=0.\end{aligned}$$

Unless all of these vanish identically we see that α is algebraic of degree ℓ, upon using that we are assuming that $\{\alpha^n,\dots,\alpha,1\}$ are linearly independent over $\mathbb{Q}$. If they all vanish it follows easily that $c_{i,j}=\delta_{i,j}c$ for some $c\in\mathbb{Q}$. Thus $c=\theta$ and $B_{k'}=\theta B_k$ with $|\theta|<1$, which contradicts that $k'>k$. □

6. Reduced bases

In this section we present several well-known results from the theory of reduced bases that we need. Perhaps the best reference for this material is a set of unpublished notes from a seminar given at IAS in 1949 [29]. Because these notes might not be readily available we have included here all proofs. Another reference is [10].

Let $\Lambda\subset\mathbb{R}^\ell$ be a full lattice and F a norm on $\mathbb{R}^\ell$. The lattice points in Λ taking on the successive minima on F are linearly independent but do not necessarily form a basis for Λ. Minkowski's second theorem implies, with some extra work, a substitute that bounds the product of values of F of the elements of a reduced basis. In case F is a positive definite quadratic form the theory was developed by Minkowski [21].

Suppose that $\{v_1,\dots,v_\ell\}$ is an ordered $\mathbb{Z}$-basis for Λ. Define for $k=1,\dots,\ell$

$$R_k = \{a_1v_1+\cdots+a_\ell v_\ell;\ a_j\in\mathbb{Z}\ \text{ with } \gcd(a_k,a_{k+1},\dots,a_\ell)=1\}\subset\Lambda. \tag{6.1}$$

Note that $v_j\notin R_k$ for $j<k$.

In general, an (ordered) $\mathbb{Z}$-basis $\{v_1, \ldots, v_\ell\}$ for Λ is *reduced* with respect to F if for each $k = 1, \ldots, \ell$ we have that for all $v \in R_k$

$$F(v) \geq F(v_k).$$

It follows that if $\{v_1, \ldots, v_\ell\}$ is reduced with respect to F and $\lambda_k = F(v_k)$ then

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_\ell.$$

Lemma 6.1. *For any norm F and full lattice $\Lambda \subset \mathbb{R}^\ell$ reduced bases $\{v_1, \ldots, v_\ell\}$ exist.*

Proof. The beginning of the proof is similar to the construction of the A_m in Minkowski's algorithm except that now we demand that $A \in \mathrm{GL}(\ell, \mathbb{Z})$. Let $\{u_1, \ldots, u_\ell\}$ be any $\mathbb{Z}$-basis for Λ. We construct $A = (r_1, \ldots, r_\ell) \in \mathrm{GL}(\ell, \mathbb{Z})$ where r_i is a column vector. Choose r_1 so that

$$\lambda_1 = F(v_1) = F\big((u_1, \ldots, u_\ell)r_1\big)$$

is minimal. Note that this exists by convexity. Now choose r_2 to minimize $\lambda_2 = F(v_2) = F((u_1, \ldots, u_\ell)r_2)$. Thus $\lambda_1 \leq \lambda_2$. Continue this process to determine A and the basis $\{v_1, \ldots, v_\ell\}$ where for $\lambda_k = F(v_k)$ we have that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_\ell$. We want to show that $\{v_1, \ldots, v_\ell\}$ is reduced.

Fix k with $1 \leq k \leq \ell$. By construction if s is any column of a matrix in $\mathrm{GL}(\ell, \mathbb{Z})$ that is linearly independent of $\{r_1, \ldots, r_{k-1}\}$ then

$$F\big((u_1, \ldots, u_\ell)s\big) \geq F\big((u_1, \ldots, u_\ell)r_k\big). \tag{6.2}$$

Let $q^\top = (q_1, \ldots, q_\ell) \in \mathbb{Z}^\ell$ be any integral vector such that $\gcd(q_k, \ldots, q_\ell) = 1$. Fix a matrix of the form

$$A' = \begin{pmatrix} I & B \\ 0 & C \end{pmatrix} \in \mathrm{GL}(\ell, \mathbb{Z})$$

where I is the $(k-1) \times (k-1)$ identity matrix and where the k^{th} column of A' is q. This is possible by our assumption on q. Clearly the first $k-1$ columns of AA' coincide with those of A. Hence if s is the k^{th} column of AA' then by (6.2)

$$\begin{aligned} F(v_k) &= F\big((u_1, \ldots, u_\ell)r_k\big) \\ &\leq F\big((u_1, \ldots, u_\ell)s\big) = F\big((u_1, \ldots, u_\ell)Aq_k\big) = F\big((v_1, \ldots, v_\ell)q_k\big). \end{aligned}$$

It follows that $\{v_1, \ldots, v_\ell\}$ is reduced. □

The statement of Part (i) of the following lemma is given in [27, Lem. 2 p. 100] with a different proof than the one we give below. Our proof is adapted from the proof of Part (ii) given in [29].

Lemma 6.2. *Let $F : \mathbb{R}^\ell \to [0,\infty)$ be a norm and $\Lambda \subset \mathbb{R}^\ell$ be a full lattice. Suppose that $v_1, \dots, v_\ell$ is a reduced basis for Λ with respect to F so that for $\lambda_i = F(v_i)$*

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_\ell.$$

(i) *If $u_1, \dots, u_\ell \in \Lambda$ is any linearly independent set in Λ ordered so that for $\nu_j = F(u_j)$*

$$\nu_1 \le \nu_2 \le \cdots \le \nu_\ell$$

then for each $k = 1, \dots, \ell$ we have that $\lambda_k \le (\frac{3}{2})^{k-1}\nu_k$.

(ii) *If $w_1, \dots, w_\ell \in \Lambda$ are minimizing vectors in Λ with successive minima $\mu_j = F(w_j)$*

$$\mu_1 \le \mu_2 \le \cdots \le \mu_\ell$$

then $\lambda_1 = \mu_1$ and for each $k = 2, \dots, \ell$ we have that $\lambda_k \le (\frac{3}{2})^{k-2}\mu_k$.

Proof. Let's first prove (i). There are $a_{i,j} \in \mathbb{Z}$ such that for each i

$$u_i = \sum_{1 \le j \le \ell} a_{i,j} v_j.$$

Fix k with $1 \le k \le \ell$. Since $\{u_1, \dots, u_\ell\}$ are linearly independent there is a j with $1 \le j \le k$ so that

$$a_{j,k}, a_{j,k+1}, \dots, a_{j,\ell}$$

are not all zero. Thus for any such j let $d = \gcd(a_{j,k}, a_{j,k+1}, \dots, a_{j,\ell}) > 0$.

If $d = 1$ then $u_j \in R_k$ and hence

$$\lambda_k \le F(u_j) = \nu_j \le \nu_k.$$

If $d > 1$ define for $m = 1, \dots, k-1$ the integer r_m with $|r_m| \le \frac{d}{2}$ so that

$$a_{j,m} + r_m \equiv 0 \pmod{d}.$$

Then $y = \frac{1}{d}(v_j + r_1 v_1 + \cdots + r_{k-1} v_{k-1}) \in R_k$. Hence we have

$$\begin{aligned} \lambda_k \le & F(y) \le \tfrac{1}{d} F(u_j) + \tfrac{r_1}{d}\lambda_1 + \cdots + \tfrac{r_{k-1}}{d}\lambda_{k-1} \\ \le & \tfrac{\nu_k}{2} + \tfrac{1}{2}(\lambda_1 + \cdots + \lambda_{k-1}). \end{aligned}$$

Therefore in any case for $k = 1, \dots \ell$ we have

$$\lambda_k \le \nu_k + \tfrac{1}{2}(\lambda_1 + \cdots + \lambda_{k-1}). \tag{6.3}$$

Suppose now that for $j = 1, \dots, k-1$

$$\lambda_j \le (\tfrac{3}{2})^{j-1}\nu_j.$$

Then by (6.3) we deduce that

$$\lambda_k \le \nu_k + \tfrac{1}{2}((\tfrac{3}{2})^0 + (\tfrac{3}{2})^1 + \cdots + (\tfrac{3}{2})^{k-2})\nu_k = (\tfrac{3}{2})^{k-1}\nu_k.$$

Since $\lambda_1 \le \nu_1$ the result (i) follows by induction.

The proof of (ii) is similar except that we use the fact that $\lambda_1 = \mu_1$. □

The following result was found independently by Mahler [18] and Weyl [28].

Theorem 5 (First Finiteness Theorem)**.** *Let $\Lambda \subset \mathbb{R}^\ell$ be a full lattice with determinant 1. For a reduced basis $\{v_1, \dots, v_\ell\}$ with $\lambda_k = F(v_k)$ we have*

$$\frac{2^\ell}{\ell!} \le \mathrm{vol}(\mathcal{B})\, \lambda_1 \cdots \lambda_\ell \le 2^\ell \left(\tfrac{3}{2}\right)^{\frac{(\ell-1)(\ell-2)}{2}}, \tag{6.4}$$

where $\mathrm{vol}(\mathcal{B})$ *is the volume of* $\mathcal{B} = \{x \in \mathbb{R}^\ell; F(x) < 1\}$.

Proof. The first inequality is a consequence of the fact that the closure of $\mathcal{B}$ contains the octahedron with vertices at the points

$$\{\pm\tfrac{v_1}{\lambda_1}, \dots, \pm\tfrac{v_\ell}{\lambda_\ell}\}$$

and this octahedron has volume $\frac{2^\ell}{\ell!\lambda_1\cdots\lambda_\ell}$, which is easily found by computing the determinant of the linear transformation that maps the k^{th} standard unit vector to $\frac{v_k}{\lambda_k}$ for each k.

The second inequality is an immediate consequence of (ii) of Lemma 6.2 and Minkowski's Second Theorem. □

We remark that we could also apply (i) of Lemma 6.2 to get the second inequality in (6.4) with the right hand side multiplied by $\frac{3}{2}$, which would be sufficient for our purposes.

7. Criteria for badly approximable and singular forms

In this section we will prove Theorems 3 and 4. We make use of the lattice $\Lambda_t(\alpha) \subset \mathbb{R}^\ell$ defined in terms of α for a fixed parameter $t > 0$ by

$$\Lambda_t = \Lambda_t(\alpha) = (t^{-1}, 0, \dots, 0, \alpha_1 t^n)\mathbb{Z} + \cdots + (0, 0, \dots, t^{-1}, \alpha_n t^n)\mathbb{Z} + (0, 0, \dots, 0, t^n)\mathbb{Z}. \tag{7.1}$$

Clearly $\det(\Lambda_t) = 1$. Consider the norm on $\mathbb{R}^\ell$ given by

$$F_\infty(x_1, \dots, x_n, y) = \|(x_1, \dots, x_n, y)\|_\infty. \tag{7.2}$$

The next lemma follows as a special case from results of [5]. For convenience we give the proof here, which for our case is quite simple.

Lemma 7.1. *For the lattice $\Lambda_t(\alpha)$ defined above let*

$$\lambda_1(t) = \min_{\substack{v \in \Lambda_t(\alpha) \\ v \neq 0}} F_\infty(v).$$

(i) *The form L_α is badly approximable if and only if there is a $c > 0$ depending only on α such that $\lambda_1(t) > c$ for all $t \ge 1$.*

(ii) *The form L_α is singular if and only if $\lambda_1(t) \to 0$ as $t \to \infty$.*

Proof.

Part (i). First suppose that α is badly approximable, so that (3.1) holds with some $c_\alpha > 0$. Fix $t \geq 1$ and $v = (x_1, \dots, x_n, y) \in \Lambda_t(\alpha)$. If $x_1, \dots, x_n = 0$ then yt^{-n} is a non-zero integer so $F_\infty(v) \geq 1$. Thus suppose that $x = (x_1, \dots, x_n) \neq 0$, set $c = c_\alpha^{\frac{1}{\ell}}$ and $\mu = \|x\|_\infty > 0$.

If $\mu \leq c$ then $t^n \|L_\alpha(tx)\| > c$, so that $F_\infty(v) > c$. If $\mu > c$ then again $F_\infty(v) > c$. In either case it follows that $\lambda_1(t) > c$.

For the converse assertion, suppose that $\lambda_1(t) > c$. Fix non-zero $q = (q_1, \dots, q_n) \in \mathbb{Z}^n$. Assume that $c < 1$. Next choose $t = c^{-1}\|q\|_\infty$ so that $t > 1$ and $t^{-1}\|q\|_\infty = c$. For any integer p we have $(t^{-1}q_1, \dots, t^{-1}q_n, t^n(p + L_\alpha(q))) \in \Lambda_t(\alpha)$ hence $t^n|L_\alpha(q) + p| > c$, which implies that

$$\|L_\alpha(q)\| > c^\ell \|q\|_\infty^{-n},$$

so L_α is badly approximable.

Part (ii). Suppose that L_α is singular and $\varepsilon > 0$ is fixed. For sufficiently large t there exists $q \in \mathbb{Z}^n$ and $p \in \mathbb{Z}$ such that $\|q\|_\infty \leq t\varepsilon^{\frac{1}{\ell}}$ and

$$|p + L_\alpha(q)| \leq t^{-n}\varepsilon^{1-\frac{n}{\ell}} = t^{-n}\varepsilon^{\frac{1}{\ell}}.$$

Let $v = (t^{-1}q_1, \dots, t^{-1}q_n, t^n(p + L_\alpha(q))) \in \Lambda_t(\alpha)$; for sufficiently large t we have $F_\infty(v) \leq \varepsilon^{\frac{1}{\ell}}$, proving that $\lambda_1(t) \to 0$ as $t \to \infty$.

The converse is similar and is left to the reader. □

Proof of Theorem 3. We have shown in Lemma 5.3 that if L_α is badly approximable then

$$\frac{|\beta_1|}{|\beta_\ell|} \geq \frac{c}{C}.$$

Now suppose that L_α is not badly approximable. By (i) of Lemma 7.1, for any $\varepsilon > 0$ there exists some $t \geq 1$ and $v \in \Lambda_t(\alpha)$ so that $F(v) < \varepsilon$ where again

$$F(v) = F_\infty(v) = \|v\|_\infty.$$

Let $\{v_1, v_2, \dots, v_\ell\}$ be a reduced basis for $\Lambda_t(\alpha)$ with respect to F and such that for $\lambda_i = F(v_i)$ we have $\lambda_1 \leq \cdots \leq \lambda_\ell$.

Now $v_1 = (t^{-1}q_1, \dots, t^{-1}q_n, t^n\xi(r))$ for some non-zero $r = (q_1, \dots, q_n, p) \in \mathbb{Z}^\ell$, where $\xi(r)$ was defined in (5.1). Clearly we have

$$\lambda_1 = F(v_1) \leq F(v) < \varepsilon \tag{7.3}$$

so from the definition of F

$$t^{-1}|q_j| < \varepsilon \quad \text{for } j = 1, \dots, n \text{ and } \quad t^n|\xi(r)| < \varepsilon. \tag{7.4}$$

Next set $m = \lceil \kappa t\varepsilon \rceil$, where κ is a constant depending only on α chosen to be large enough so that $\max(|q_1|, \dots, |q_n|, |p|) \leq m$, which is possible

by (7.4). For $A_m = (a_{i,j})$ from Minkowski's algorithm let for each $i = 1, \dots, \ell$

$$u_i = (t^{-1}a_{i,1}, \dots, t^{-1}a_{i,n}, t^n\beta_i) \in \Lambda_t(\alpha). \tag{7.5}$$

By the definition of β_1 and (7.4) again we have that

$$|\beta_1| \le |\xi(r)| < t^{-n}\varepsilon. \tag{7.6}$$

By construction

$$t^n|\beta_1| < t^n|\beta_2| < \cdots < t^n|\beta_\ell|, \tag{7.7}$$

but we do not know that necessarily

$$F(u_1) \le F(u_2) \le \cdots \le F(u_\ell).$$

Let $k \in \{1, \dots, \ell\}$ be such that $F(u_k) = \max(F(u_1), \dots, F(u_\ell))$. Since $\{u_1, \dots, u_\ell\}$ are linearly independent in $\Lambda_t(\alpha)$, by (i) of Lemma 6.2 we have that

$$F(u_k) \ge (\tfrac{2}{3})^n F(v_\ell). \tag{7.8}$$

By the first inequality of the First Finiteness Theorem and (7.3) we see that $F(v_\ell) > (\frac{1}{\ell!\varepsilon})^{1/n}$ and therefore by (7.8)

$$F(u_k) > (\tfrac{2}{3})^n (\tfrac{1}{\ell!\varepsilon})^{1/n}. \tag{7.9}$$

Now

$$\max(t^{-1}|a_{k,1}|, \dots, t^{-1}|a_{k,n}|) \le \tfrac{m}{t} \le \kappa\varepsilon + t^{-1} < (\tfrac{2}{3})^n (\tfrac{1}{\ell!\varepsilon})^{1/n}$$

for $\varepsilon > 0$ sufficiently small. Hence by this, (7.9) and (7.7) we have

$$|\beta_\ell| \ge |\beta_k| > t^{-n}(\tfrac{2}{3})^n (\tfrac{1}{\ell!\varepsilon})^{1/n}, \tag{7.10}$$

after referring again to (7.5). By (7.6) we conclude that for sufficiently small ε

$$|\tfrac{\beta_1}{\beta_\ell}| < (\tfrac{3}{2})^n (\ell!\varepsilon)^{1/n}\varepsilon.$$

It follows that if L_α is not badly approximable then $|\frac{\beta_1}{\beta_\ell}|$ can be made arbitrarily small for some m, hence $|\alpha_{k,1}|$ can be made arbitrarily small for some k. □

Proof of Theorem 4. Suppose that α is singular and let $\varepsilon \in (0, 1)$. Then there exists a t_0 such that

$$\lambda_1(t) < \varepsilon \quad \text{for all } t \ge t_0.$$

This is the analogue of (7.3) but now the inequality holds for all sufficiently large t. Let m be any positive integer greater than t_0 and let $t = m/\varepsilon \ge t_0$. If $A_m = (a_{i,j})$ is the m-th matrix from Minkowski's algorithm, then by following the argument between (7.5) and (7.10) above we find that

$$|\beta_1| < \frac{\varepsilon}{t^n} \quad \text{and} \quad |\beta_\ell| > \frac{c}{t^n \varepsilon^{1/n}}$$

for some $c > 0$. It follows that $\left|\frac{\beta_1}{\beta_\ell}\right|$ can be made arbitrarily small for all sufficiently large m, hence $|\alpha_{k,1}|$ can be made arbitrarily small for all sufficiently large k.

Conversely, suppose that α is not singular. Then there exists a $c > 0$ and a sequence $\{Q_j\}$ tending to infinity such that for each j there are infinitely many $q \in \mathbb{Z}^n$ with

$$\|q\|_\infty \leq Q_j \quad \text{ and } \|L_\alpha(q)\| \geq cQ_j^{-n}.$$

Fix one of these Q_j and let $m = Q_j$. Then

$$\beta_1 = \min_{\substack{q \in \mathbb{Z}^n \setminus \{0\} \\ \|q\|_\infty \leq m}} |\xi(q)| \geq \frac{c}{m^n}.$$

This is analogous to (5.8) but now the lower bound only holds for a sequence of m tending to infinity. Note that (5.4) is true for these m, and following (5.9)–(5.11) we find that $|\beta_\ell| \leq \frac{C}{m}$ holds here as well. It follows that for infinitely many m we have $\left|\frac{\beta_1}{\beta_\ell}\right| \geq c'$ for some $c' > 0$ and thus $|\alpha_{k,1}|$ is bounded away from zero for infinitely many k. □

Acknowledgements. We thank the anonymous referee for several helpful comments.

References

[1] Y. Bugeaud, "On simultaneous rational approximation to a real number and its integral powers", *Ann. Inst. Fourier* **60** (2010), no. 6, p. 2165-2182.

[2] J. W. S. Cassels, *An introduction to the geometry of numbers*, Grundlehren der Mathematischen Wissenschaften, vol. 99, Springer, 1959, viii+344 pages.

[3] Y. Cheung, "Hausdorff dimension of the set of singular pairs", *Ann. Math.* **173** (2011), no. 1, p. 127-167.

[4] Y. Cheung & N. Chevallier, "Hausdorff dimension of singular vectors", *Duke Math. J.* **165** (2016), no. 12, p. 2273-2329.

[5] S. G. Dani, "Divergent trajectories of flows on homogeneous spaces and Diophantine approximation", *J. Reine Angew. Math.* **359** (1985), p. 55-89.

[6] T. Das, L. Fishman, D. Simmons & M. Urbański, "A variational principle in the parametric geometry of numbers, with applications to metric Diophantine approximation", *C. R. Math. Acad. Sci. Paris* **355** (2017), no. 8, p. 835-846.

[7] H. Davenport, "Minkowski's inequality for the minima associated with a convex body", *Q. J. Math., Oxf. Ser.* **10** (1939), no. 1, p. 119-121.

[8] H. Davenport & W. M. Schmidt, "Dirichlet's theorem on Diophantine approximation. II", *Acta Arith.* **16** (1969), p. 413-424.

[9] ———, "Dirichlet's theorem on Diophantine approximation", in *Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69)*, Academic Press Inc., 1970, p. 113-132.

[10] P. M. Gruber & C. G. Lekkerkerker, *Geometry of numbers*, North-Holland Mathematical Library, North-Holland, 1987, xvi+732 pages.

[11] H. Hancock, *Development of the Minkowski geometry of numbers. Vol. 1, 2*, Dover Publications, 1964.

[12] A. Hurwitz, "Über die angenäherte Darstellung der Zahlen durch rationale Brüche", *Math. Ann.* **XLIV** (1894), p. 417-436.

[13] C. G. J. Jacobi, "Allgemeine Theorie der kettenbruchähnlichen Algorithmen, in welche jede Zahl aus drei vorhergehenden gebildet wird", *J. Reine Angew. Math.* **69** (1891), p. 29-64.
[14] A. Khintchine, "Über eine Klasse linearer diophantischer Approximationen", *Rend. Circ. Mat. Palermo* **50** (1926), p. 170-195.
[15] ———, "Zur metrischen Theorie der diophantischen Approximationen", *Math. Z.* **24** (1926), p. 706-714.
[16] J. C. Lagarias, "Number theory and dynamical systems", in *The unreasonable effectiveness of number theory (Orono, ME, 1991)*, Proceedings of Symposia in Applied Mathematics, vol. 46, American Mathematical Society, 1991, p. 35-72.
[17] ———, "Geodesic multidimensional continued fractions", *Proc. Lond. Math. Soc.* **69** (1994), no. 3, p. 464-488.
[18] K. Mahler, "On Minkowski's theory of reduction of positive definite quadratic forms", *Q. J. Math., Oxf. Ser.* **9** (1938), p. 259-262.
[19] H. Minkowski, "Ein Kriterium für die algebraischen Zahlen", *Gött. Nachr.* **1899** (1899), p. 64-88.
[20] ———, "Über periodische Approximationen algebraischer Zahlen", *Acta Math.* (1902), p. 333-352.
[21] ———, "Diskontinuitatsbereich fur arithmetische Aquivalenz", *J. Reine Angew. Math.* **129** (1905), p. 220-274.
[22] ———, *Geometrie der Zahlen*, Teubner, 1910.
[23] P. Philippon, "A Farey tail", *Notices Am. Math. Soc.* **59** (2012), no. 6, p. 746-757.
[24] W. M. Schmidt, "Badly approximable systems of linear forms", *J. Number Theory* **1** (1969), p. 139-154.
[25] ———, *Diophantine approximation*, Lecture Notes in Mathematics, Springer, 1980, x+299 pages.
[26] W. M. Schmidt & L. Summerer, "Parametric geometry of numbers and applications", *Acta Arith.* **140** (2009), no. 1, p. 67-91.
[27] C. L. Siegel, *Lectures on the geometry of numbers*, Springer, 1989, x+160 pages.
[28] H. Weyl, "On geometry of numbers", *Proc. Lond. Math. Soc.* **47** (1942), p. 268-289.
[29] H. Weyl, C. L. Siegel & K. Mahler, *Seminar on Geometry of Numbers, IAS*, 1949.

Nickolas Andersen
Brigham Young University
Department Of Mathematics
Provo, UT 84602, USA
E-mail: nick@math.byu.edu

William Duke
UCLA Mathematics Department
Box 951555
Los Angeles, CA 90095-1555, USA
E-mail: wdduke@ucla.edu