

A MULTIDIMENSIONAL CONTINUED FRACTION BASED ON A HIGH-ORDER RECURRENCE RELATION

YVES TOURIGNY AND NIGEL P. SMART

ABSTRACT. The paper describes and studies an iterative algorithm for finding small values of a set of linear forms over vectors of integers. The algorithm uses a linear recurrence relation to generate a vector sequence, the basic idea being to choose the integral coefficients in the recurrence relation in such a way that the linear forms take small values, subject to the requirement that the integers should not become too large. The problem of choosing good coefficients for the recurrence relation is thus related to the problem of finding a good approximation of a given vector by a vector in a certain one-parameter family of lattices; the novel feature of our approach is that practical formulae for the coefficients are obtained by considering the limit as the parameter tends to zero. The paper discusses two rounding procedures to solve the underlying inhomogeneous Diophantine approximation problem: the first, which we call “naive rounding” leads to a multidimensional continued fraction algorithm with suboptimal asymptotic convergence properties; in particular, when it is applied to the familiar problem of simultaneous rational approximation, the algorithm reduces to the classical Jacobi–Perron algorithm. The second rounding procedure is Babai’s nearest-plane procedure. We compare the two rounding procedures numerically; our experiments suggest that the multidimensional continued fraction corresponding to nearest-plane rounding converges at an optimal asymptotic rate.

1. INTRODUCTION

The classical continued fraction algorithm is closely connected with Euclid’s algorithm for finding the greatest common divisor of two integers, and so it has a very long history. The present paper describes a generalisation of the algorithm applicable to a set of linear forms. Many such generalisations have been proposed in the last 150 years; the purpose of this introductory section is to explain in broad terms how our own generalisation differs from others, and why it is of interest.

1.1. The classical continued fraction algorithm. Let \mathbb{S} be an interval of unit length and let $[\cdot] : \mathbb{R} \rightarrow \mathbb{Z}$ denote some function such that, for every $\xi \in \mathbb{R}$,

$$(1.1) \quad \{\xi\} := \xi - [\xi] \in \mathbb{S}.$$

Received by the editor August 22, 2006.

2000 *Mathematics Subject Classification.* Primary 11J70.

Key words and phrases. Multidimensional continued fraction, recurrence relation.

The first author acknowledges the support of the Engineering and Physical Sciences Research Council (United Kingdom) under a Discipline Hopping Award.

Two examples are: (1) the *floor* function, i.e. $[\xi]$ is the largest integer not exceeding ξ ; in that case, $\mathbb{S} = [0, 1]$. (2) $[\xi]$ is the integer nearest to ξ ; in that case, $\mathbb{S} = [-1/2, 1/2]$.

It is well-known that every real number $\theta \in \mathbb{S}$ may be expanded as a continued fraction. More precisely, one can write

$$\theta = \frac{1}{c_0 + \frac{1}{c_1 + \cdots + \frac{1}{c_{n-1} + \varrho_n}}},$$

where the coefficients c_n are integers, and the “remainders” ϱ_n are real numbers in \mathbb{S} . Both the coefficients and the remainders are obtained by successive iterates of a map defined by

$$(1.2) \quad \mathcal{G} : \mathbb{S} \rightarrow \mathbb{S}, \quad \mathcal{G}(\xi) = \begin{cases} \left\{ \frac{1}{\xi} \right\} & \text{if } \xi \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$(1.3) \quad \varrho_n = \mathcal{G}^n(\theta) \quad \text{and} \quad c_n = \left[\frac{1}{\varrho_n} \right].$$

If one neglects the remainder ϱ_n , then one obtains a rational number

$$\frac{p_n}{q_n} := \frac{1}{c_0 + \frac{1}{c_1 + \cdots + \frac{1}{c_{n-1}}}}$$

that approximates θ . The continued fraction expansion is analogous to the more familiar binary and decimal expansions but, unlike these, does not rely on the choice of a particular number as “base”. For this reason, it has been a very useful tool in the study of the real numbers.

Some of the generic features of the expansion can be obtained by studying the ergodic properties of the map \mathcal{G} and its associated invariant measure. The case where $[\cdot]$ is the floor function and $\mathbb{S} = [0, 1]$ is the most familiar; see [4] for a brief account. The case where $[\cdot]$ rounds to the nearest integer and $\mathbb{S} = [-1/2, 1/2]$ is qualitatively similar; see [11].

Other remarkable properties may be deduced from first principles: for example, both p_n and q_n satisfy the recurrence relation

$$(1.4) \quad a_{n+1} = c_n a_n + a_{n-1}, \quad n \in \mathbb{N}.$$

With the choice of starting values $p_0 = q_{-1} = 0$ and $p_{-1} = q_0 = 1$, the *unimodularity property*

$$(1.5) \quad \det \begin{pmatrix} p_n & q_n \\ p_{n-1} & q_{n-1} \end{pmatrix} = \pm 1$$

follows. Furthermore, it can be shown by induction that

$$(1.6) \quad \theta = \frac{p_{n-1}\varrho_n + p_n}{q_{n-1}\varrho_n + q_n}.$$

Hence

$$(1.7) \quad \left| \theta - \frac{p_n}{q_n} \right| \leq \frac{C}{q_n^2}$$

for some constant $C > 0$ which depends on the particular choice of the function $[\cdot]$. The significance of this result is that the continued fraction algorithm produces rational approximations of the real number θ which converge at a rate that is essentially optimal as the denominators grow. Indeed, it is a standard result of metric number theory that, for every $\varepsilon > 0$ and $C > 0$, the set of real numbers θ such that the inequality

$$\left| \theta - \frac{p}{q} \right| < \frac{C}{q^{2+\varepsilon}}$$

holds for infinitely many pairs $(p, q) \in \mathbb{Z} \times \mathbb{N}$ has Lebesgue measure zero.

1.2. A more general problem of Diophantine approximation. We consider the following generalisation of the problem of finding good rational approximations of a real number. For $d, k \in \mathbb{N}$ with $k \leq d$, set

$$k' := d + 1 - k.$$

Suppose that we are given a $k \times (d + 1)$ matrix L such that

$$(1.8) \quad \Delta_k := \det \begin{pmatrix} L^{(1,1)} & \dots & L^{(1,k)} \\ \vdots & \dots & \vdots \\ L^{(k,1)} & \dots & L^{(k,k)} \end{pmatrix} \neq 0.$$

Define a set of linear forms

$$L^{(i)} : \mathbb{Z}^k \times \mathbb{Z}^{k'} \rightarrow \mathbb{R}, \quad 1 \leq i \leq k,$$

by

$$(1.9) \quad L^{(i)}(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^k L^{(i,j)} p^{(j)} + \sum_{j=1}^{k'} L^{(i,k+j)} q^{(j)}.$$

Definition 1.1. The *Roth exponent* of the pair $\mathbf{p} \in \mathbb{Z}^k$ and $\mathbf{q} \in \mathbb{Z}^{k'}$ is the number

$$\gamma_L(\mathbf{p}, \mathbf{q}) := -\frac{\log \|\mathbf{L}(\mathbf{p}, \mathbf{q})\|_k}{\log \|\mathbf{q}\|_{k'}},$$

where $\|\cdot\|_m$ denotes the Euclidean norm on \mathbb{R}^m .

Definition 1.2. The *best approximation exponent* of the set of linear forms corresponding to L is the number

$$\gamma_L := \limsup_{\|\mathbf{q}\|_{k'} \rightarrow \infty} \gamma_L(\mathbf{p}, \mathbf{q}).$$

We associate with the set of linear forms corresponding to the matrix L the following approximation problem: *Find sequences $\{\mathbf{p}_n\}_{n \in \mathbb{N}} \subset \mathbb{Z}^k$, $\{\mathbf{q}_n\}_{n \in \mathbb{N}} \subset \mathbb{Z}^{k'}$ such that*

$$(1.10) \quad \lim_{n \rightarrow \infty} \gamma_L(\mathbf{p}_n, \mathbf{q}_n) = \gamma_L.$$

Example 1.3. Let $\theta \in \mathbb{S}^d$. Take $k = d$ and set

$$L = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & -\theta^{(1)} \\ 0 & 1 & 0 & \dots & 0 & -\theta^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\theta^{(d)} \end{pmatrix}.$$

Then

$$\gamma_L(\mathbf{p}, q) = -\frac{\log \|\mathbf{p} - q\theta\|_d}{\log |q|},$$

and our approximation problem is equivalent to that of constructing unbounded sequences of \mathbf{p}_n and q_n such that

$$\max_{1 \leq i \leq d} |q_n \theta^{(i)} - p_n^{(i)}| = O(q_n^{-\gamma}) \quad \text{as } q_n \rightarrow \infty,$$

with $\gamma > 0$ as large as possible. This is the well-known problem of simultaneous rational approximation. Dirichlet proved that, for every θ with irrational entries,

$$(1.11) \quad \gamma_L \geq \frac{1}{d}.$$

It is also known that the set of the θ such that the inequality is strict has Lebesgue measure zero.

The Jacobi–Perron algorithm is one of the earliest and best-known algorithms proposed for the solution of this problem; it has been studied extensively over the last hundred years (see [2, 3, 9, 5, 13] and the references therein). Although the Jacobi–Perron algorithm actually *fails* to solve the problem when $d > 1$ (see below), we shall nevertheless describe it briefly because it is closely related to the algorithm which we shall develop later in this paper.

We use a starting scheme such that

$$(1.12) \quad \theta = \frac{\mathbf{p}_0 + \theta^{(1)}\mathbf{p}_{-1} + \dots + \theta^{(d-1)}\mathbf{p}_{-d+1} + \theta^{(d)}\mathbf{p}_{-d}}{q_0 + \theta^{(1)}q_{-1} + \dots + \theta^{(d-1)}q_{-d+1} + \theta^{(d)}q_{-d}}.$$

Let us suppose that, for some $n \in \mathbb{N}$, the pairs

$$(\mathbf{p}_n, q_n), \dots, (\mathbf{p}_{n-d}, q_{n-d})$$

are available, and that

$$(1.13) \quad \theta = \frac{\mathbf{p}_n + \varrho_n^{(1)}\mathbf{p}_{n-1} + \dots + \varrho_n^{(d_n-1)}\mathbf{p}_{n-d_n+1} + \varrho_n^{(d_n)}\mathbf{p}_{n-d_n}}{q_n + \varrho_n^{(1)}q_{n-1} + \dots + \varrho_n^{(d_n-1)}q_{n-d_n+1} + \varrho_n^{(d_n)}q_{n-d_n}}$$

for some known vector $\varrho_n \in \mathbb{S}^d$, where

$$d_n := \max \left\{ 1 \leq j \leq d : \varrho_n^{(j)} \neq 0 \right\}.$$

Equation (1.12) shows that this is true for $n = 0$ and that

$$\varrho_0 = \theta.$$

We proceed to show how to define $(\mathbf{p}_{n+1}, q_{n+1})$ and ϱ_{n+1} so that this also holds for $n + 1$. We divide the numerator and the denominator in equation (1.13) by $\varrho_n^{(d_n)}$ so that

$$\theta = \frac{\frac{1}{\varrho_n^{(d_n)}}\mathbf{p}_n + \frac{\varrho_n^{(1)}}{\varrho_n^{(d_n)}}\mathbf{p}_{n-1} + \dots + \frac{\varrho_n^{(d_n-1)}}{\varrho_n^{(d_n)}}\mathbf{p}_{n-d_n+1} + \mathbf{p}_{n-d_n}}{\frac{1}{\varrho_n^{(d_n)}}q_n + \frac{\varrho_n^{(1)}}{\varrho_n^{(d_n)}}q_{n-1} + \dots + \frac{\varrho_n^{(d_n-1)}}{\varrho_n^{(d_n)}}q_{n-d_n+1} + q_{n-d_n}}.$$

Now, set

$$c_n^{(i)} = \begin{cases} \left[\frac{1}{\varrho_n^{(d_n)}} \right] & \text{if } i = 1, \\ \left[\frac{\varrho_n^{(i-1)}}{\varrho_n^{(d_n)}} \right] & \text{if } 1 < i \leq d_n. \end{cases}$$

Define the next pair $(\mathbf{p}_{n+1}, q_{n+1})$ via the recurrence relations

$$(1.14) \quad \mathbf{p}_{n+1} = \sum_{i=1}^{d_n} c_n^{(i)} \mathbf{p}_{n-i} + \mathbf{p}_{n-d_n}, \quad q_{n+1} = \sum_{i=1}^{d_n} c_n^{(i)} q_{n-i} + q_{n-d_n}.$$

Upon substitution in the equation above, we obtain

$$\theta = \frac{\mathbf{p}_{n+1} + \left\{ \frac{1}{\varrho_n^{(d_n)}} \right\} \mathbf{p}_n + \left\{ \frac{\varrho_n^{(1)}}{\varrho_n^{(d_n)}} \right\} \mathbf{p}_{n-1} + \cdots + \left\{ \frac{\varrho_n^{(d_n-1)}}{\varrho_n^{(d_n)}} \right\} \mathbf{p}_{n-d_n+1}}{q_{n+1} + \left\{ \frac{1}{\varrho_n^{(d_n)}} \right\} q_n + \left\{ \frac{\varrho_n^{(1)}}{\varrho_n^{(d_n)}} \right\} q_{n-1} + \cdots + \left\{ \frac{\varrho_n^{(d_n-1)}}{\varrho_n^{(d_n)}} \right\} q_{n-d_n+1}}.$$

This shows that equation (1.13) holds when n is replaced by $n + 1$, with

$$\varrho_{n+1} = \mathcal{J}(\varrho_n),$$

where $\mathcal{J} : \mathbb{S}^d \rightarrow \mathbb{S}^d$ is the operator defined by

$$(1.15) \quad \mathcal{J}^{(i)}(\varrho) = \begin{cases} \left\{ \frac{1}{\varrho^{(d_\varrho)}} \right\} & \text{if } i = 1 \text{ and } d_\varrho \neq 0, \\ \left\{ \frac{\varrho^{(i-1)}}{\varrho^{(d_\varrho)}} \right\} & \text{if } 1 < i \leq d_\varrho, \\ 0 & \text{otherwise} \end{cases}$$

and

$$d_\varrho = \max \{ 1 \leq j \leq d : \varrho^{(j)} \neq 0 \}.$$

Thus, the new pair $(\mathbf{p}_{n+1}, q_{n+1})$ is entirely determined by the pairs

$$(\mathbf{p}_{n-k+1}, q_{n-k+1}), \quad 1 \leq k \leq d + 1,$$

and the n th iterate of a map from \mathbb{S}^d to itself. Lagarias called such algorithms *Markovian* and studied their metrical properties in [9] for the case where $\mathbb{S} = [0, 1]$ and $[\cdot]$ is the floor function. For the Jacobi–Perron algorithm, he showed that the left-hand side of equation (1.10) is strictly positive for almost all θ , but conjectured (on the basis of overwhelming numerical evidence) that it falls short of the best possible value $1/d$. One goal of the present paper is to present an algorithm that does achieve this best possible value; as we shall see however, our algorithm is not Markovian. Following Lagarias, it seems doubtful whether *any* Markovian algorithm can solve the best simultaneous rational approximation problem when $d > 1$.

Example 1.4. Let $\theta \in \mathbb{S}$, take $k = 1$ and set

$$L = (1 \quad \theta \quad \dots \quad \theta^d).$$

Then our approximation problem is equivalent to that of finding polynomials

$$\psi_n(t) := p_n + q_n^{(1)}t + \cdots + q_n^{(d)}t^d$$

of degree d , with integer coefficients, such that

$$\psi_n(\theta) = O\left(|\psi_n|^{-\gamma}\right) \quad \text{as } |\psi_n| \rightarrow \infty,$$

with $\gamma > 0$ as large as possible. In this expression, $|\psi|$ denotes the *height* of the polynomial ψ , i.e.

$$|\psi| := \max_{0 \leq j \leq d} |\psi^{(j)}|, \quad \text{where } \psi(t) = \psi^{(0)} + \psi^{(1)}t + \dots + \psi^{(d)}t^d.$$

This problem is, in some sense, the dual of the approximation problem considered in the previous example. By a well-known transference principle,

$$(1.16) \quad \gamma_L = d \quad \text{for almost every } \theta \in \mathbb{S}.$$

In the next subsection, we shall describe a class of algorithms that solve this problem.

1.3. Lattices and Diophantine approximation. One approach to solving the Diophantine problem is to introduce a one-parameter family of lattices, and recast it as the problem of finding a small vector in the lattice—the (small) parameter being used to control the accuracy of the approximation.

Recall that a lattice \mathbb{L} in \mathbb{R}^{d+1} is a set of the form

$$\mathbb{L} = \left\{ \sum_{j=1}^{d+1} x^{(j)} \boldsymbol{\beta}^{(j)} : \mathbf{x} \in \mathbb{Z}^{d+1} \right\},$$

where the $\boldsymbol{\beta}^{(i)}$ are vectors in \mathbb{R}^{d+1} . If the $\boldsymbol{\beta}^{(i)}$ are linearly independent, then we say that they form a *basis* for the lattice. We represent the basis as a matrix whose columns are the basis vectors:

$$B := \left(\boldsymbol{\beta}^{(1)} \quad \dots \quad \boldsymbol{\beta}^{(d+1)} \right).$$

Now, if \tilde{B} is another basis for the same lattice, then there exists a unimodular matrix P with integer entries such that

$$B = \tilde{B}P.$$

It follows in particular that the number

$$\text{vol}(\mathbb{L}) := |\det(B)|$$

is independent of the choice of basis; this number is called the *volume* of the lattice. An ordered basis is said to be *reduced* if its vectors satisfy certain requirements. These requirements vary somewhat from one reduction definition to another, but, from our point of view, the most important is that there exist a constant $C_d > 0$ —depending possibly on d , but otherwise independent of the lattice—such that the first vector in any reduced basis B of the lattice satisfies the inequality

$$(1.17) \quad \|\boldsymbol{\beta}^{(1)}\|_{d+1} \leq C_d \text{vol}(\mathbb{L})^{\frac{1}{d+1}}.$$

We give two examples:

- (1) In his paper, Lagarias [10] emphasizes a variant of the reduction method developed by Minkowski where the first basis vector is the smallest non-zero vector in the lattice. Equation (1.17) then holds with C_d equal to the so-called *Hermite constant* for a lattice of dimension $d+1$; this constant is bounded above by $d+1$.

- (2) A very convenient alternative definition of reduction was proposed by Lenstra, Lenstra and Lovász [8]; importantly, they developed a practical algorithm for finding a reduced basis and showed, among other things, that equation (1.17) holds with $C_d = 2^{d/2}$.

Returning to Diophantine approximation, let $0 < \varepsilon \in \mathbb{R}$ and denote by \mathbb{L}_ε the lattice whose basis is the $(d + 1) \times (d + 1)$ matrix

$$(1.18) \quad \tilde{B}_\varepsilon := \begin{pmatrix} L^{(1,1)} & \dots & L^{(1,k)} & L^{(1,k+1)} & L^{(1,k+2)} & \dots & L^{(1,d+1)} \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ L^{(k,1)} & \dots & L^{(k,k)} & L^{(k,k+1)} & L^{(k,k+2)} & \dots & L^{(k,d+1)} \\ 0 & \dots & 0 & \varepsilon & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \varepsilon & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & & \ddots & \\ 0 & \dots & 0 & 0 & 0 & \dots & \varepsilon \end{pmatrix}.$$

Now, suppose that B_ε is a reduced basis for \mathbb{L}_ε , and denote by P_ε the integral unimodular matrix such that

$$B_\varepsilon = \tilde{B}_\varepsilon P_\varepsilon.$$

We write

$$\mathbf{p}_\varepsilon := \begin{pmatrix} p_\varepsilon^{(1,1)} \\ \vdots \\ p_\varepsilon^{(k,1)} \end{pmatrix} \quad \text{and} \quad \mathbf{q}_\varepsilon := \begin{pmatrix} p_\varepsilon^{(k+1,1)} \\ \vdots \\ p_\varepsilon^{(d+1,1)} \end{pmatrix}.$$

The first vector of the reduced basis is then

$$\beta_\varepsilon^{(1)} = \begin{pmatrix} \mathbf{L}(\mathbf{p}_\varepsilon, \mathbf{q}_\varepsilon) \\ \varepsilon \mathbf{q}_\varepsilon \end{pmatrix}.$$

By invoking equation (1.17), it is very easy to show that

$$\|\mathbf{L}(\mathbf{p}_\varepsilon, \mathbf{q}_\varepsilon)\|_k \xrightarrow{\varepsilon \rightarrow 0} 0$$

and

$$(1.19) \quad \|\mathbf{q}_\varepsilon\|_{k'}^{\frac{k'}{k}} \|\mathbf{L}(\mathbf{p}_\varepsilon, \mathbf{q}_\varepsilon)\|_k \leq C_d^{\frac{d+1}{k}} |\Delta_k|^{\frac{1}{k}}.$$

Note that, for the particular cases in Examples 1.3 and 1.4, the exponent of $\|\mathbf{q}_\varepsilon\|_{k'}$ on the left-hand side of the inequality is (generically) as large as possible in the limit $\varepsilon \rightarrow 0$.

Typically, as the small parameter ε tends to zero, it goes through a sequence of critical values at which the reduced basis changes. A *geodesic multidimensional continued fraction* algorithm then solves the Diophantine approximation problem by (1) choosing a particular definition of reduction, (2) choosing a null sequence from which to pick successive values of ε and (3) using an algorithm for computing the unimodular matrix that maps the given basis to a reduced one.

It is fair to say that the particular geodesic multidimensional continued fraction algorithm devised by Lagarias [10] is not really practical because, for the particular choice of reduction theory, no efficient algorithm is known for finding the reduced basis when d is large. Likewise, the approach studied in [7], which emphasises Siegel sets rather than strict basis reduction, seems to be a purely theoretical tool. By contrast, the approach described in [8], based on the availability of the powerful *LLL reduction* algorithm, is eminently suitable for practical computations. Even so,

in cases where a long subsequence of the Diophantine pairs $(\mathbf{p}_n, \mathbf{q}_n)$ are required, one would like a convenient way of finding the critical values of the parameter ε at which the reduced basis changes, and a means of updating the reduced basis as the parameter changes, rather than having to compute it “from scratch”. The algorithms of Just [6] and Rössner & Schnorr [12] address this issue successfully for the particular problem of simultaneous rational approximation considered in Example 1.3. It is unclear whether these algorithms admit a natural extension to the general Diophantine problem which we consider here.

1.4. Our contribution. The purpose of this paper is to describe a new multidimensional version of the classical continued fraction algorithm that is, like those of Just [6] and Rössner & Schnorr [12], motivated by the lattice-based approach—yet *parameter free*. As mentioned earlier, one feature of the ordinary continued fraction algorithm is that the numerator p_n and the denominator q_n of the n th convergent of the number θ both satisfy the second-order recurrence relation (1.4), where the c_n are the coefficients in the expansion of θ . We shall use the high-order generalisation

$$(1.20) \quad a_{n+1} := c_n^{(1)} a_n + \cdots + c_n^{(d)} a_{n-d+1} + a_{n-d},$$

where the coefficients $c_n^{(j)}$ are integers. Given some starting values, the object is then to choose these coefficients to produce the next Diophantine approximation. While the recurrence relation itself is of the same form as that used in the Jacobi–Perron algorithm (see equation (1.14)), our strategy for choosing the coefficients is different. To describe it, let us introduce a small parameter ε and define some “residual” vectors, say $\mathcal{L}_n(\varepsilon) \in \mathbb{R}^{d+1}$. As in the geodesic approach, we think of this parameter as one that controls the quality of the approximation—the smaller the parameter, the smaller the error of approximation—and that acts also as a “penalty” parameter, constraining the growth of the Diophantine integers. By construction, these residuals are linear combinations of the Diophantine integers; therefore, they also satisfy the recurrence relation (1.20). So we seek coefficients $c_n^{(j)}$ that make the new residual small. This is equivalent to the *inhomogeneous* Diophantine approximation problem of finding a vector in the lattice generated by the $\mathcal{L}_{n+1-i}(\varepsilon)$, $1 \leq i \leq d$, that is near to the vector $-\mathcal{L}_{n-d}(\varepsilon)$.

Babai [1] describes two methods for solving such a problem: the first consists of expressing (the projection of) the approximated vector as a linear combination of the lattice’s basis vectors, and then rounding the coefficients; we shall refer to this method as *the naive rounding procedure*. The second method is analogous, but expresses the (projection of the) approximated vector as a linear combination of the *Gram–Schmidt basis* vectors obtained from the lattice’s basis, and rounds these coefficients; we shall call it *Babai’s nearest-plane rounding procedure*. As it turns out, this procedure is an important component of the LLL algorithm. Using one or the other of these rounding procedures for $\varepsilon > 0$, we then show that coefficients $c_n^{(i)}$ *independent of the parameter ε* can be easily calculated by taking the limit $\varepsilon \rightarrow 0$. We emphasise this essential difference between our approach and geodesic continued fractions: The new algorithm that we shall describe involves no “small parameter”; the sequence of approximation is obtained by recurrence via (1.20)—the natural generalisation of the second-order recurrence (1.4) satisfied by the numerator and denominator of the classical continued fraction. For the particular problem of Example 1.3, the main difference between our algorithm and those of Just [6] and Rössner & Schnorr [12] is that, in our case, the ordering of the residuals is purely

temporal: roughly speaking, one may say that our algorithm implements only the size-reduction component of the LLL algorithm—not the swapping component. As a result, the successive bases that our algorithm generates are *not* LLL-reduced and, because of this, we have not been able to develop a rigorous convergence theory analogous to that of Rössner & Schnorr [12].

Our main results are as follows: when the naive rounding procedure is used for the particular case of simultaneous rational approximation, i.e. Example 1.3, we show that the classical Jacobi–Perron algorithm is obtained. For the particular case of algebraic approximation, i.e. Example 1.4, we show experimentally that, for $d > 3$, the use of naive rounding produces a sequence of the $(\mathbf{p}_n, \mathbf{q}_n)$ that is eventually periodic, and so fails to solve the approximation problem. By contrast, when the nearest-plane procedure is used, we observe experimentally that, for both examples, the continued fraction algorithm produces pairs $(\mathbf{p}_n, \mathbf{q}_n)$ with optimal asymptotic properties. We also point out that, in the case of simultaneous rational approximation, the resulting algorithm is *not Markovian*.

The remainder of the paper is as follows. In §2, we describe the rounding procedures which we shall use to find good coefficients for the recurrence relation. The algorithm for the solution of our general Diophantine problem is described in §3. The last two sections are devoted to numerical experiments where the two rounding procedures are compared: in §4, we consider the problem of constructing good simultaneous rational approximations of a given real vector; in §5, we consider the problem of finding a good algebraic approximation of a given real number.

2. INHOMOGENEOUS DIOPHANTINE APPROXIMATION

Given a lattice \mathbb{L} in \mathbb{R}^{d+1} , and a lattice basis

$$B := \left(\boldsymbol{\beta}^{(1)} \dots \boldsymbol{\beta}^{(d_{\mathbb{L}})} \right),$$

we shall denote by

$$\hat{B} := \left(\hat{\boldsymbol{\beta}}^{(1)} \dots \hat{\boldsymbol{\beta}}^{(d_{\mathbb{L}})} \right)$$

the matrix made of orthogonal vectors obtained by applying the Gram–Schmidt algorithm to B . More precisely,

$$(2.1) \quad \hat{\boldsymbol{\beta}}^{(i)} := \boldsymbol{\beta}^{(i)} - \sum_{j=1}^{i-1} \mu^{(i,j)} \hat{\boldsymbol{\beta}}^{(j)}, \quad \mu^{(i,j)} := \begin{cases} \frac{(\boldsymbol{\beta}^{(i)}, \hat{\boldsymbol{\beta}}^{(j)})}{(\hat{\boldsymbol{\beta}}^{(j)}, \hat{\boldsymbol{\beta}}^{(j)})} & \text{if } j < i, \\ 1 & \text{if } i = j. \end{cases}$$

Babai [1] considers the following problem: given $\boldsymbol{\xi} \in \mathbb{R}^{d+1}$, find an approximation $\boldsymbol{\xi}_{\mathbb{L}} \in \mathbb{L}$ of $\boldsymbol{\xi}$. Following Babai, we shall describe two procedures for its solution. We write

$$\mathbb{V} := \text{span} \left\{ \boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(d_{\mathbb{L}})} \right\}$$

and

$$\mathbf{P}_{\mathbb{V}} : \mathbb{R}^{d+1} \rightarrow \mathbb{V}$$

for the operator that projects onto the vector space \mathbb{V} . Set

$$\mu^{(d_{\mathbb{L}}+1,j)} := \frac{(\boldsymbol{\xi}, \hat{\boldsymbol{\beta}}^{(j)})}{(\hat{\boldsymbol{\beta}}^{(j)}, \hat{\boldsymbol{\beta}}^{(j)})}, \quad 1 \leq j \leq d_{\mathbb{L}}.$$

First, we shall find a convenient expression for the projection of ξ in terms of the basis B . Let us write

$$\mathbf{P}_V \xi =: \sum_{j=1}^{d_L} \xi_B^{(j)} \beta^{(j)}.$$

In order to determine the $\xi_B^{(j)}$, we first express the projection in terms of the Gram-Schmidt basis:

$$\mathbf{P}_V \xi = \sum_{j=1}^{d_L} \mu^{(d_L+1,j)} \hat{\beta}^{(j)}.$$

Then, by (2.1), we easily recover the $\xi_B^{(j)}$ via the recurrence relation

$$(2.2) \quad \xi_B^{(j)} = \mu^{(d_L+1,j)} - \sum_{i=j+1}^{d_L} \mu^{(i,j)} \xi_B^{(i)}, \quad j = d_L, \dots, 1.$$

Babai proposes two different choices for the approximation $\xi_L \in \mathbb{L}$:

- (1) The first choice—which we call the *naive rounding* approximation—is simply

$$\xi_L := \sum_{j=1}^{d_L} \left[\xi_B^{(j)} \right] \beta^{(j)},$$

where the $\xi_B^{(j)}$ are obtained via the recurrence relation (2.2).

- (2) The second choice—called the *nearest-plane rounding* approximation—is

$$\xi_L := \sum_{j=1}^{d_L} \left[\hat{\xi}_B^{(j)} \right] \beta^{(j)},$$

where the $\hat{\xi}_B^{(j)}$ are obtained via the recurrence relation

$$(2.3) \quad \hat{\xi}_B^{(j)} = \mu^{(d_L+1,j)} - \sum_{i=j+1}^{d_L} \mu^{(i,j)} \left[\hat{\xi}_B^{(i)} \right], \quad j = d_L, \dots, 1.$$

Comparing equations (2.2) and (2.3), we see that the difference between the two choices is subtle. In particular, if $d_L = 1$, then the two procedures coincide. In general, however, they do differ, and the following concept is helpful in distinguishing between them.

Definition 2.1. Given a basis B of V , we say that a vector $\eta \in \mathbb{R}^{d+1}$ is *B-reduced* if

$$\mathbf{P}_V \eta = \sum_{j=1}^{d_L} \eta_B^{(j)} \beta^{(j)}, \quad \text{where } \eta_B^{(j)} \in \mathbb{S} \text{ for every } 1 \leq j \leq d_L.$$

Lemma 2.2. *The vector $\xi - \xi_L$ is B-reduced if the naive rounding procedure is used; it is \hat{B} -reduced if the nearest-plane rounding procedure is used.*

Proof. If the naive rounding procedure is used, then

$$\mathbf{P}_V (\xi - \xi_L) = \mathbf{P}_V \xi - \xi_L = \sum_{j=1}^{d_L} \left\{ \xi_B^{(j)} \right\} \beta^{(j)},$$

and so the first statement is obvious.

If the nearest-plane rounding procedure is used, then, by making use of equation (2.1), we can write

$$\mathbf{P}_V \boldsymbol{\xi} - \boldsymbol{\xi}_L = \sum_{j=1}^{d_L} \mu^{(d_L+1,j)} \hat{\boldsymbol{\beta}}^{(j)} - \sum_{j=1}^{d_L} \sum_{i=j}^{d_L} \left[\hat{\xi}_B^{(i)} \right] \mu^{(i,j)} \hat{\boldsymbol{\beta}}^{(j)}.$$

So, after taking the inner product with $\hat{\boldsymbol{\beta}}^{(j)}$, we obtain

$$\begin{aligned} \frac{\left(\mathbf{P}_V \boldsymbol{\xi} - \boldsymbol{\xi}_L, \hat{\boldsymbol{\beta}}^{(j)} \right)}{\left(\hat{\boldsymbol{\beta}}^{(j)}, \hat{\boldsymbol{\beta}}^{(j)} \right)} &= \mu^{(d_L+1,j)} - \sum_{i=j}^{d_L} \left[\hat{\xi}_B^{(i)} \right] \mu^{(i,j)} \\ &= \mu^{(d_L+1,j)} - \sum_{i=j+1}^{d_L} \left[\hat{\xi}_B^{(i)} \right] \mu^{(i,j)} - \left[\hat{\xi}_B^{(j)} \right] = \hat{\xi}_B^{(j)} - \left[\hat{\xi}_B^{(j)} \right] = \left\{ \hat{\xi}_B^{(j)} \right\} \end{aligned}$$

by definition of $\hat{\xi}_B^{(j)}$. □

3. THE ALGORITHMS

The algorithms that we shall present in this section use the recurrence relations

$$(3.1) \quad \mathbf{p}_{n+1} = \sum_{j=1}^{d_n} c_n^{(j)} \mathbf{p}_{n-j+1} + \mathbf{p}_{n-d_n}, \quad \mathbf{q}_{n+1} = \sum_{j=1}^{d_n} c_n^{(j)} \mathbf{q}_{n-j+1} + \mathbf{q}_{n-d_n},$$

to generate vectors of integers

$$\mathbf{p}_n \in \mathbb{Z}^k \text{ and } \mathbf{q}_n \in \mathbb{Z}^{k'}, \quad -d \leq n \in \mathbb{Z},$$

for the approximate solution of the Diophantine problem of §1.2. They differ in the choice of the integer coefficients in the recurrence relation (see §3.2). To describe them completely, we also need to specify the *order* d_n of the recurrence relation, as well as a *starting scheme*. However, for the sake of readability, we feel that it is preferable, in the first instance, to define the order and the starting scheme in the context of some specific Diophantine problems where the choice is “obvious”. It will then be clear how to define the order in the general case. The dependence of the asymptotic behaviour of the algorithm on the starting scheme is an interesting question, but we shall not address it. We begin this section by providing some motivation for the choices that will be made.

3.1. Motivation. Given $\mathbf{p}_n \in \mathbb{Z}^k$ and $\mathbf{q}_n \in \mathbb{Z}^{k'}$, we shall use the convenient notation

$$\mathbf{L}_n := \mathbf{L}(\mathbf{p}_n, \mathbf{q}_n).$$

Given a suitable starting scheme, the coefficients $c_n^{(j)}$ in the recurrence relation (3.1) should be such that, on the one hand, the new residual \mathbf{L}_{n+1} is small while, on the other hand, the vector \mathbf{q}_{n+1} is not too large. One possible strategy—motivated by the lattice approach described in §1.3—to control the size of \mathbf{q}_{n+1} is to introduce a small parameter $\varepsilon > 0$ and define an extended “residual” by

$$(3.2) \quad \mathcal{L}_n(\varepsilon) := \begin{pmatrix} \mathbf{L}_n \\ \varepsilon \mathbf{q}_n \end{pmatrix}.$$

We denote by $\mathbb{L}_n(\varepsilon)$ the lattice generated by the basis

$$(3.3) \quad B_n(\varepsilon) := (\mathcal{L}_n(\varepsilon) \quad \dots \quad \mathcal{L}_{n-d_n+1}(\varepsilon))$$

and by $\mathbb{V}_n(\varepsilon)$ the corresponding vector space. The operator of projection onto $\mathbb{V}_n(\varepsilon)$ will be denoted $\mathbf{P}_n(\varepsilon)$, and we shall write $\hat{B}_n(\varepsilon)$ for the Gram-Schmidt basis obtained from $B_n(\varepsilon)$.

Now, let us assume that, for some $n \geq 0$, the basis $B_n(\varepsilon)$ is available, and write

$$(3.4) \quad \mu_n^{(i,j)}(\varepsilon) := \begin{cases} \frac{(\mathcal{L}_{n+1-i}(\varepsilon), \hat{\beta}_n^{(j)}(\varepsilon))}{(\hat{\beta}_n^{(j)}(\varepsilon), \hat{\beta}_n^{(j)}(\varepsilon))} & \text{if } j < i, \\ 1 & \text{if } j = i, \end{cases} \quad 1 \leq i \leq d_n + 1,$$

for the Gram-Schmidt coefficients. Since the extended residual is a linear combination of the \mathbf{p}_n and \mathbf{q}_n , it satisfies the same recurrence relation, i.e.

$$(3.5) \quad \mathcal{L}_{n+1}(\varepsilon) = c_n^{(1)} \mathcal{L}_n(\varepsilon) + \dots + c_n^{(d_n)} \mathcal{L}_{n-d_n+1}(\varepsilon) + \mathcal{L}_{n-d_n}(\varepsilon).$$

The problem of finding good coefficients is then the same as that of finding a good approximation of $-\mathcal{L}_{n-d_n}(\varepsilon)$ in the lattice $\mathbb{L}_n(\varepsilon)$. So one could use one of the two procedures described in §2 with

$$\boldsymbol{\xi} := -\mathcal{L}_{n-d_n}(\varepsilon), \quad B := B_n(\varepsilon)$$

to find an element

$$\boldsymbol{\xi}_{\mathbb{L}(\varepsilon)} = \sum_{j=1}^{d_n} c_n^{(j)}(\varepsilon) \mathcal{L}_{n-i}(\varepsilon) \in \mathbb{L}_n(\varepsilon)$$

close to $\boldsymbol{\xi}$. Using these $c_n^{(j)}(\varepsilon)$ in the relation (3.5) would produce a new residual that is $B_n(\varepsilon)$ -reduced if the naive rounding procedure is used, and $\hat{B}_n(\varepsilon)$ -reduced if the nearest-plane rounding procedure is used. Now, the projection of $-\mathcal{L}_{n-d}(\varepsilon)$ onto $\mathbb{V}_n(\varepsilon)$ is

$$-\mathbf{P}_n(\varepsilon) \mathcal{L}_{n-d}(\varepsilon) = -\sum_{i=1}^{d_n} \mu_n^{(d_n+1,i)}(\varepsilon) \mathcal{L}_{n-i}(\varepsilon).$$

So, for example, the nearest-plane procedure described in §2 would then yield the approximation

$$\boldsymbol{\xi}_{\mathbb{L}_n(\varepsilon)} := \sum_{j=1}^{d_n} c_n^{(j)}(\varepsilon) \mathcal{L}_{n-j+1}(\varepsilon), \quad c_n^{(j)}(\varepsilon) := \left[\hat{\xi}_n^{(j)}(\varepsilon) \right],$$

where

$$(3.6) \quad \hat{\xi}_n^{(j)}(\varepsilon) = -\mu_n^{(d_n+1,j)}(\varepsilon) - \sum_{i=j+1}^{d_n} \left[\hat{\xi}_n^{(i)}(\varepsilon) \right] \mu_n^{(i,j)}(\varepsilon), \quad j = d_n, \dots, 1.$$

The drawback of this approach is that these integers are ε -dependent; the presence of such a parameter seems undesirable. We remark that, in the case of the geodesic approach described in §1.3, the parameter ε is destined to tend to zero. Now, $B_n(0)$ contains at most k linearly independent vectors, so that the Gram-Schmidt procedure applied directly to $B_n(0)$ breaks down prematurely unless $k = d$. It turns out, however, that if one orthogonalises $B_n(\varepsilon)$, $\varepsilon \neq 0$, before taking the limit as $\varepsilon \rightarrow 0$, then the Gram-Schmidt coefficients have limiting values that can be computed easily, even if $k < d$. Our strategy for choosing the coefficients is based on that simple observation.

In order to describe the asymptotic behaviour of $\hat{B}_n(\varepsilon)$, we define vectors $\hat{\mathbf{L}}_n^{(i)}$ and $\hat{\mathbf{q}}_n^{(i)}$ for $n \geq 0$ and $1 \leq i \leq d+1$ by recurrence on i as follows:

$$(3.7) \quad \hat{\mathbf{L}}_n^{(i)} := \mathbf{L}_{n-i+1} - \sum_{j=1}^{i-1} \nu_n^{(i,j)} \hat{\mathbf{L}}_n^{(j)}, \quad \hat{\mathbf{q}}_n^{(i)} := \mathbf{q}_{n-i+1} - \sum_{j=1}^{i-1} \nu_n^{(i,j)} \hat{\mathbf{q}}_n^{(j)},$$

where

$$(3.8) \quad \nu_n^{(i,j)} := \begin{cases} \frac{(\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(j)})}{(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)})} & \text{if } \hat{\mathbf{L}}_n^{(j)} \neq \mathbf{0}, \\ \frac{\left(\mathbf{q}_{n-i+1} - \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^{j-1} \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j)} \right)}{(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)})} & \text{if } \hat{\mathbf{L}}_n^{(j)} = \mathbf{0} \text{ and } \hat{\mathbf{q}}_n^{(j)} \neq \mathbf{0}, \\ 0 & \text{otherwise.} \end{cases}$$

It will be convenient to use the notation

$$\mathbf{O}_m(\varepsilon^\gamma) := \begin{pmatrix} O(\varepsilon^\gamma) \\ \vdots \\ O(\varepsilon^\gamma) \end{pmatrix} \in \mathbb{R}^m, \quad m \in \mathbb{N}.$$

Theorem 3.1.

$$\hat{\beta}_n^{(i)}(\varepsilon) = \begin{pmatrix} \hat{\mathbf{L}}_n^{(i)} \\ \varepsilon \hat{\mathbf{q}}_n^{(i)} \end{pmatrix} - \varepsilon^2 \sum_{\substack{j=1 \\ \hat{\mathbf{L}}_n^{(j)} \neq \mathbf{0}}}^{i-1} \frac{(\hat{\mathbf{q}}_n^{(i)}, \hat{\mathbf{q}}_n^{(j)})}{(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)})} \begin{pmatrix} \hat{\mathbf{L}}_n^{(j)} \\ \varepsilon \hat{\mathbf{q}}_n^{(j)} \end{pmatrix} + \begin{pmatrix} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{pmatrix}$$

and

$$\mu_n^{(i,j)}(\varepsilon) = \nu_n^{(i,j)} + O(\varepsilon^2) \quad \text{as } \varepsilon \rightarrow 0.$$

Proof. See Appendix A. □

Corollary 3.2. Let $\mathbf{p} \in \mathbb{R}^k$ and $\mathbf{q} \in \mathbb{R}^{k'}$ and set

$$\mathcal{L}(\varepsilon) := \begin{pmatrix} \mathbf{L}(\mathbf{p}, \mathbf{q}) \\ \varepsilon \mathbf{q} \end{pmatrix}, \quad \text{where } \mathbf{L} := \mathbf{L}(\mathbf{p}, \mathbf{q}).$$

Then

$$\frac{(\mathcal{L}(\varepsilon), \hat{\beta}_n^{(j)}(\varepsilon))}{(\hat{\beta}_n^{(j)}(\varepsilon), \hat{\beta}_n^{(j)}(\varepsilon))} \xrightarrow{\varepsilon \rightarrow 0} \begin{cases} \frac{(\mathbf{L}(\mathbf{p}, \mathbf{q}), \hat{\mathbf{L}}_n^{(j)})}{(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)})} & \text{if } \hat{\mathbf{L}}_n^{(j)} \neq \mathbf{0}, \\ \frac{\left(\mathbf{q} - \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^{j-1} \frac{(\mathbf{L}(\mathbf{p}, \mathbf{q}), \hat{\mathbf{L}}_n^{(\ell)})}{(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)})} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j)} \right)}{(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)})} & \text{if } \hat{\mathbf{L}}_n^{(j)} = \mathbf{0} \text{ and } \hat{\mathbf{q}}_n^{(j)} \neq \mathbf{0}. \end{cases}$$

3.2. Choosing the coefficients in the recurrence relation. We shall consider two choices for the coefficients in equation (3.1); although they coincide when $d = 1$, we emphasise that, in higher dimension, they lead to two distinct algorithms with, as we shall see, very different asymptotic behaviours. The first is based on the naive rounding procedure:

$$(3.9) \quad c_n^{(j)} = \left[\eta_n^{(j)} \right] \quad (\text{Naive rounding})$$

where

$$(3.10) \quad \eta_n^{(j)} = -\nu_n^{(d_n+1,j)} - \sum_{i=j+1}^{d_n} \eta_n^{(i)} \nu_n^{(i,j)}, \quad j = d_n, \dots, 1.$$

The second is based on Babai's nearest-plane rounding procedure:

$$(3.11) \quad c_n^{(j)} = \left[\hat{\eta}_n^{(j)} \right] \quad (\text{Nearest-plane rounding})$$

where

$$(3.12) \quad \hat{\eta}_n^{(j)} = -\nu_n^{(d_n+1,j)} - \sum_{i=j+1}^{d_n} c_n^{(i)} \nu_n^{(i,j)}, \quad j = d_n, \dots, 1.$$

4. SIMULTANEOUS RATIONAL APPROXIMATION

We return to the problem considered in Example 1.3; the matrix L is then

$$(4.1) \quad L = \begin{pmatrix} 1 & 0 & \dots & 0 & -\theta^{(1)} \\ 0 & 1 & \dots & 0 & -\theta^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & 1 & -\theta^{(d)} \end{pmatrix}.$$

First, we choose some starting values for the \mathbf{p}_n and the q_n . Let $\mathbf{e}^{(j)}$ be the j th vector in the canonical basis for \mathbb{R}^d . We set

$$\mathbf{p}_n := \begin{cases} \mathbf{e}^{(-n)} & \text{if } -d \leq n \leq -1, \\ \mathbf{0} & \text{if } n = 0 \end{cases}, \quad \text{and} \quad q_n := \begin{cases} 0 & \text{if } -d \leq n \leq -1, \\ 1 & \text{if } n = 0. \end{cases}$$

Because $k = d$ in this case, the formulae in §3 take on a simpler form. In particular, the $\hat{\mathbf{L}}_n^{(j)}$ are simply the orthogonal vectors obtained when the Gram–Schmidt procedure is applied to the vectors \mathbf{L}_{n+1-j} , where

$$\mathbf{L}_n := \mathbf{p}_n - q_n \boldsymbol{\theta} \in \mathbb{R}^d.$$

The $\nu_n^{(i,j)}$ are the corresponding Gram–Schmidt coefficients. We set

$$(4.2) \quad d_n = \max \{ 1 \leq i \leq d : \mathbf{L}_{n+1-j}, 1 \leq j \leq i, \text{ are linearly independent} \}.$$

We shall write

$$B_n := (\mathbf{L}_n \quad \dots \quad \mathbf{L}_{n-d_n+1})$$

and denote by \mathbb{V}_n the vector space with basis B_n . The corresponding Gram–Schmidt basis will be denoted by

$$\hat{B}_n := (\hat{\mathbf{L}}_n^{(1)} \quad \dots \quad \hat{\mathbf{L}}_n^{(d_n)}).$$

Note that the starting scheme is such that $B_{-1} = \hat{B}_{-1}$ is the canonical basis for \mathbb{R}^d . Furthermore

$$(4.3) \quad -\mathbf{L}_0 := \theta^{(1)}\mathbf{L}_{-1} + \cdots + \theta^{(d)}\mathbf{L}_{-d} = \theta^{(1)}\hat{\mathbf{L}}_{-1} + \cdots + \theta^{(d)}\hat{\mathbf{L}}_{-d}.$$

The significance of this identity will soon be apparent.

When representing vectors in \mathbb{V}_n , we shall sometimes need to go from one basis to another; the following lemma will therefore be helpful.

Lemma 4.1. *We have*

$$\mathbf{L}_{n-i+1} = \sum_{j=1}^i \nu_n^{(i,j)} \hat{\mathbf{L}}_n^{(j)}, \quad \hat{\mathbf{L}}_n^{(i)} = \sum_{j=1}^i \lambda_n^{(i,j)} \mathbf{L}_{n-j+1}, \quad 1 \leq i \leq d_n,$$

where

$$\nu_n^{(i,j)} = \frac{\left(\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(j)}\right)}{\left(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)}\right)}, \quad 1 \leq j < i \leq d_n + 1,$$

and

$$\lambda_n^{(i,j)} := \begin{cases} \sum_{\ell=1}^{i-j} (-1)^\ell \nu_n^{(i,j+\ell-1)} \prod_{m=1}^{\ell-1} \nu_n^{(j+m,j+m-1)} & \text{if } i > j, \\ 1 & \text{if } i = j. \end{cases}$$

Furthermore, if

$$\sum_{i=1}^{d_n} \sigma_{B_n}^{(i)} \mathbf{L}_{n-i+1} = \sum_{i=1}^{d_n} \sigma_{\hat{B}_n}^{(i)} \hat{\mathbf{L}}_n^{(i)},$$

then

$$\sigma_{\hat{B}_n}^{(j)} = \sum_{i=j}^{d_n} \nu_n^{(i,j)} \sigma_{B_n}^{(i)} \quad \text{and} \quad \sigma_{B_n}^{(j)} = \sum_{i=j}^{d_n} \lambda_n^{(i,j)} \sigma_{\hat{B}_n}^{(i)}, \quad 1 \leq j \leq d_n.$$

Let $\eta \in \mathbb{R}^{d_n}$ be defined as in equation (3.10). We have

Lemma 4.2.

$$-\mathbf{L}_{n-d_n} = \eta_n^{(1)} \mathbf{L}_n + \cdots + \eta_n^{(d_n)} \mathbf{L}_{n-d_n+1}.$$

Proof. By equation (4.2), the vectors

$$\mathbf{L}_n, \dots, \mathbf{L}_{n-d_n+1},$$

are linearly independent whilst the vectors

$$\mathbf{L}_n, \dots, \mathbf{L}_{n-d_n+1}, \mathbf{L}_{n-d_n},$$

are *dependent*. Hence, \mathbf{L}_{n-d_n} can be expressed as a linear combination of the \mathbf{L}_{n-i+1} , $1 \leq i \leq d_n$. Clearly, by definition of the Gram-Schmidt coefficients,

$$\mathbf{L}_{n-d_n} = \sum_{i=1}^{d_n} \nu_n^{(d_n+1,i)} \hat{\mathbf{L}}_n^{(i)}.$$

The result then follows from Lemma 4.1 and equation (3.10). □

It is instructive to compare the algorithm based on naive rounding with the one that uses nearest-plane rounding. Recall that the only difference between the two algorithms is in the choice of the coefficients $c_n^{(j)}$ that are used in the recurrence relation (3.1); see §3.2.

4.1. Naive rounding. In this subsection, we shall show that the use of naive rounding produces the same \mathbf{p}_n and q_n as the classical Jacobi–Perron algorithm described in the introduction.

Definition 4.3. We define the sequence $\{\varrho_n\}_{n \in \mathbb{N}} \subset \mathbb{S}^d$ of the (*naive rounding*) remainders as follows:

$$\varrho_0 := \boldsymbol{\theta} \quad \text{and} \quad \varrho_{n+1}^{(j)} := \begin{cases} \{\eta_n^{(j)}\} & \text{if } 1 \leq j \leq d_n, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 4.4. For every $n \in \mathbb{N}$,

$$-\mathbf{L}_n = \varrho_n^{(1)} \mathbf{L}_{n-1} + \cdots + \varrho_n^{(d_{n-1})} \mathbf{L}_{n-d_{n-1}}.$$

Proof. For $n = 0$, this is a consequence of the starting scheme; see equation (4.3). Now, let $n \geq 0$. By construction,

$$-\mathbf{L}_{n+1} = -c_n^{(1)} \mathbf{L}_n - \cdots - c_n^{(d_n)} \mathbf{L}_{n-d_{n+1}} - \mathbf{L}_{n-d_n},$$

where (see equation (3.9))

$$c_n^{(j)} = \left[\eta_n^{(j)} \right], \quad 1 \leq j \leq d_n.$$

The result then follows from Lemma 4.2 and the definition of ϱ_n . \square

The lemma brings out an obvious relationship between the remainder and the order of the recurrence relation:

$$(4.4) \quad d_n := \begin{cases} 0 & \text{if } \varrho_n = \mathbf{0}, \\ \max \{1 \leq j \leq d : \varrho_n^{(j)} \neq 0\} & \text{otherwise.} \end{cases}$$

In particular, $d_0 = d$ unless $\theta^{(d)} = 0$. The sequence $\{d_n\}_{n \in \mathbb{N}}$ is monotonically decreasing and so its terms must eventually equal the limit, say d_∞ . We conjecture that

$$d_\infty = \text{rank}(\boldsymbol{\theta}),$$

where

$$\text{rank}(\boldsymbol{\theta}) = \dim \text{span}_{\mathbb{Q}} \{1, \theta^{(1)}, \dots, \theta^{(d)}\}.$$

Corollary 4.5. If $d_n = d$, then

$$\text{vol}(\mathbb{L}_n) = \left| \varrho_n^{(d)} \right| \text{vol}(\mathbb{L}_{n-1}).$$

Proof. We have

$$\begin{aligned} \text{vol}(\mathbb{L}_n) &= |\det(\mathbf{L}_n, \mathbf{L}_{n-1}, \dots, \mathbf{L}_{n-d+1})| \\ &= |\det(\mathbf{L}_{n-1}, \dots, \mathbf{L}_{n-d+1}, \mathbf{L}_n)| \\ &= \left| -\det \left(\mathbf{L}_{n-1}, \dots, \mathbf{L}_{n-d+1}, \sum_{j=1}^d \varrho_n^{(j)} \mathbf{L}_{n-j} \right) \right| \\ &= \left| \varrho_n^{(d)} \det(\mathbf{L}_{n-1}, \dots, \mathbf{L}_{n-d+1}, \mathbf{L}_{n-d}) \right| = \left| \varrho_n^{(d)} \right| \text{vol}(\mathbb{L}_{n-1}). \end{aligned}$$

\square

Theorem 4.6. For every $n \in \mathbb{N}$,

$$\varrho_{n+1} = \mathcal{J}(\varrho_n).$$

Proof. Lemma 4.4 yields

$$-\mathbf{L}_{n-d_n} = \frac{1}{\varrho_n^{(d_n)}} \left(\mathbf{L}_n + \varrho_n^{(1)} \mathbf{L}_{n-1} + \cdots + \varrho_n^{(d_n-1)} \mathbf{L}_{n-d_n+1} \right).$$

Hence, by using Lemma 4.2 and equating coefficients, we find

$$(4.5) \quad \eta_n^{(j)} = \begin{cases} \frac{1}{\varrho_n^{(d_n)}} & \text{if } j = 1, \\ \frac{\varrho_n^{(j-1)}}{\varrho_n^{(d_n)}} & \text{if } 1 < j \leq d_n. \end{cases}$$

The result then follows from Definition 4.3. □

Corollary 4.7. When naive rounding is used, the classical Jacobi–Perron algorithm is obtained.

4.2. Nearest-plane rounding. We now turn to the algorithm obtained when nearest-plane rounding is used. The concept of remainder, which we encountered when studying the Jacobi–Perron algorithm, has a natural counterpart:

Definition 4.8. We define the sequence $\{\hat{\varrho}_n\}_{n \in \mathbb{N}} \subset \mathbb{S}^d$ of the (*nearest-plane rounding*) remainders by

$$\hat{\varrho}_0 := \boldsymbol{\theta} \quad \text{and} \quad \hat{\varrho}_{n+1}^{(j)} := \begin{cases} \hat{\eta}_n^{(j)} & \text{if } 1 \leq j \leq d_n, \\ 0 & \text{otherwise.} \end{cases}$$

The nearest-plane counterpart of Lemma 4.4 is

Lemma 4.9.

$$-\mathbf{L}_n = \hat{\varrho}_n^{(1)} \hat{\mathbf{L}}_{n-1}^{(1)} + \cdots + \hat{\varrho}_n^{(d_n-1)} \hat{\mathbf{L}}_{n-1}^{(d_n-1)}.$$

Its proof proceeds along the same lines. It is also straightforward to show that

$$d_n = \max \left\{ 1 \leq j \leq d : \hat{\varrho}_n^{(j)} \neq 0 \right\}$$

and, if $d_n = d$,

$$(4.6) \quad \text{vol}(\mathbb{L}_n) = \left| \hat{\varrho}_n^{(d)} \right| \text{vol}(\mathbb{L}_{n-1}).$$

Next, we proceed to elaborate the relationship between successive remainders. Let $\boldsymbol{\eta}_n \in \mathbb{R}^{d_n}$ be the vector defined via the recurrence relation (3.10). We have

Lemma 4.10.

$$\eta_n^{(1)} = \frac{1}{\hat{\varrho}_n^{(d_n)}}, \quad \eta_n^{(j+1)} = \frac{1}{\hat{\varrho}_n^{(d_n)}} \sum_{i=j}^{d_n} \lambda_{n-1}^{(i,j)} \hat{\varrho}_n^{(i)}, \quad 1 \leq j < d_n.$$

Proof. By Lemma 4.9 and Lemma 4.1, we can write

$$\mathbf{L}_n = - \sum_{j=1}^{d_n-1} \sum_{i=j}^{d_n-1} \lambda_{n-1}^{(i,j)} \hat{\varrho}_n^{(i)} \mathbf{L}_{n-j}.$$

The result then follows from Lemma 4.2 after equating the coefficients. □

Proposition 4.11.

$$\hat{\eta}_n^{(j)} = - \sum_{i=j+1}^{d_n} \lambda_n^{(i,j)} \hat{\varrho}_{n+1}^{(i)} + \frac{1}{\hat{\varrho}_n^{(d_n)}} \begin{cases} 1 & \text{if } j = 1, \\ \sum_{i=j-1}^{d_n} \lambda_{n-1}^{(i,j-1)} \hat{\varrho}_n^{(i)} & \text{if } 1 < j \leq d_n. \end{cases}$$

Proof. By equation (3.12) and Definition 4.8,

$$\sum_{i=j}^{d_n} \nu_n^{(i,j)} \left[\hat{\eta}_n^{(i)} \right] = -\nu_n^{(d_n+1,j)} - \hat{\varrho}_n^{(j)}, \quad 1 \leq j \leq d_n.$$

On the other hand, by equation (3.10),

$$\sum_{i=j}^{d_n} \nu_n^{(i,j)} \eta_n^{(i)} = -\nu_n^{(d_n+1,j)}, \quad 1 \leq j \leq d_n.$$

Upon subtraction, we find

$$\sum_{i=j}^{d_n} \nu_n^{(i,j)} \left(\eta_n^{(i)} - \left[\hat{\eta}_n^{(i)} \right] \right) = \hat{\varrho}_{n+1}^{(j)}, \quad 1 \leq j \leq d_n.$$

Hence, by Lemma 4.1,

$$\eta_n^{(j)} - \left[\hat{\eta}_n^{(j)} \right] = \sum_{i=j}^{d_n} \lambda_n^{(i,j)} \hat{\varrho}_{n+1}^{(i)}, \quad 1 \leq j \leq d_n.$$

The result then follows from Lemma 4.10. \square

The proposition implies that, in the case $d > 1$, the remainders cannot be expressed as the successive iterates of a map. Indeed,

$$\hat{\varrho}_{n+1}^{(d_n)} = \left\{ \frac{\hat{\varrho}_n^{(d_{n-1})}}{\hat{\varrho}_n^{(d_n)}} + \lambda_{n-1}^{(d_n, d_{n-1})} \right\}.$$

Now, it is clear that the coefficient

$$\lambda_{n-1}^{(d_n, d_{n-1})} = -\nu_{n-1}^{(d_n, d_{n-1})}$$

does vary with n . In other words, when nearest-plane rounding is used, the resulting algorithm is *not* Markovian.

4.3. Convergence. In this subsection, we examine the convergence properties of the algorithm based on the use of nearest-plane rounding. Although the motivation for the algorithm is quite compelling, we have not succeeded in developing a rigorous theory. Equation (4.6) implies that

$$\text{vol}(\mathbb{L}_n) = O(2^{-n}) \quad \text{as } n \rightarrow \infty.$$

One would hope that some estimate for the residual \mathbf{L}_n could be deduced from this, but experiments show that the angles between successive residuals can be arbitrarily small.

Another possible line of attack is to take Lemma 4.9 as a starting point. Then, by the orthogonality of the $\hat{\mathbf{L}}_n^{(j)}$, we obtain

$$(4.7) \quad \|\mathbf{L}_{n+1}\|^2 = \sum_{j=1}^{d_n} \left| \hat{\varrho}_{n+1}^{(j)} \right|^2 \left\| \hat{\mathbf{L}}_n^{(j)} \right\|^2 \leq \frac{1}{4} \left(\|\mathbf{L}_n\|^2 + \cdots + \|\mathbf{L}_{n-d+1}\|^2 \right).$$

TABLE 1. The Roth exponent γ_n , as defined by equation (4.8). The columns headed (n-r) and (n-p-r) correspond to naive rounding and nearest-plane rounding respectively.

n	$d = 1$	$d = 2$		$d = 3$		$d = 4$		$d = 5$	
		n-r	n-p-r	n-r	n-p-r	n-r	n-p-r	n-r	n-p-r
2	1.81	0.37	0.94	0.70	1.29	0.37	1.54	0.37	0.38
4	1.39	0.74	0.90	0.42	0.69	0.13	0.12	0.24	0.16
8	1.18	0.34	0.55	0.20	0.37	0.20	0.26	0.11	0.20
16	1.04	0.38	0.53	0.21	0.37	0.21	0.28	0.16	0.22
32	1.07	0.34	0.52	0.20	0.34	0.19	0.25	0.12	0.20
64	1.03	0.41	0.51	0.22	0.34	0.14	0.27	0.09	0.19
128	1.00	0.40	0.51	0.19	0.34	0.15	0.25	0.11	0.20
256	1.00	0.41	0.50	0.22	0.33	0.14	0.25	0.10	0.20
512	1.00	0.38	0.50	0.22	0.34	0.14	0.25	0.11	0.20
1024	1.00	0.40	0.50	0.23	0.33	0.14	0.25	0.10	0.20

So (the square of the norm of) the residual is dominated by a solution of the recurrence relation

$$4a_{n+1} = a_n + \dots + a_{n-d+1}.$$

The characteristic polynomial associated with this recurrence relation is

$$\varphi_d(t) = 4t^d - t^{d-1} - \dots - t - 1.$$

An explicit calculation shows that, when $d \leq 3$, the roots of φ_d lie within the complex unit disk. Hence,

$$\|\mathbf{L}_n\| \xrightarrow{n \rightarrow \infty} 0 \quad \text{if } d = 2 \text{ or } 3.$$

Example 4.12. Let $\boldsymbol{\theta} \in \mathbb{R}^d$ have random entries that are independent and uniformly distributed in \mathbb{S} . Table 1 shows the Roth exponents

$$(4.8) \quad \gamma_n := -\frac{\ln \|\mathbf{P}_n - q_n \boldsymbol{\theta}\|}{\ln |q_n|}$$

for various values of n and a typical realisation of $\boldsymbol{\theta}$. We have used the well-known MAPLE symbolic algebra package to carry out the calculations; these were done in floating-point arithmetic, but with enough precision (up to 2000 significant digits in some cases) to ensure that the results are accurate. It is apparent that, when nearest-plane rounding is used,

$$\gamma_n \xrightarrow{n \rightarrow \infty} \frac{1}{d}.$$

Thus, we conjecture that the approximations produced by the algorithm converge at an optimal rate.

By contrast, the naive rounding procedure leads to an algorithm that produces a sequence with a suboptimal rate of convergence. This is a known shortcoming of the classical Jacobi–Perron algorithm [3].

5. ALGEBRAIC APPROXIMATION

We return to the problem considered in Example 1.4, where

$$(5.1) \quad L = (1 \quad \theta \quad \cdots \quad \theta^d)$$

and $\theta \in \mathbb{S}$ is given. Here $k = 1$, so we write L_n instead of \mathbf{L}_n . The formulae of §3.1 simplify considerably: assuming that $L_n \neq 0$, we have

$$\hat{L}_n^{(1)} = L_n \quad \text{and} \quad \nu_n^{(i,1)} = \frac{L_{n-i+1}}{L_n}.$$

So, by induction on j , equation (3.7) yields

$$\hat{L}_n^{(j)} = 0 \quad \text{for every } j \geq 2.$$

Equation (3.8) then becomes

$$\nu_n^{(i,j)} = \begin{cases} \frac{L_{n-i+1}}{L_n} & \text{if } j = 1, \\ \frac{(\mathbf{q}_{n-i+1} - \nu_n^{(i,1)} \hat{\mathbf{q}}_n^{(1)}, \hat{\mathbf{q}}_n^{(j)})}{(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)})} & \text{if } j > 1 \text{ and } \hat{\mathbf{q}}_n^{(j)} \neq \mathbf{0}, \\ 0 & \text{otherwise} \end{cases}$$

where

$$\hat{\mathbf{q}}_n^{(i)} = \mathbf{q}_{n-i+1} - \sum_{j=1}^{i-1} \nu_n^{(i,j)} \hat{\mathbf{q}}_n^{(j)}.$$

It remains to define the order of the recurrence relations (3.1), and the starting values. We set

$$d_n := \begin{cases} 0 & \text{if } L_n = 0, \\ 1 & \text{if } L_n \neq 0 \text{ and } \hat{\mathbf{q}}_n^{(2)} = \mathbf{0}, \\ \min \{2 \leq i \leq d : \hat{\mathbf{q}}_n^{(i)} \neq \mathbf{0}\} & \text{otherwise} \end{cases}$$

and

$$(5.2) \quad (p_n, \mathbf{q}_n) := \begin{cases} (1, \mathbf{0}) & \text{if } n = -d, \\ (0, \mathbf{e}^{(d+n)}) & \text{if } -d < n \leq 0, \end{cases}$$

where $\mathbf{e}^{(i)}$ is the i th canonical basis vector in \mathbb{R}^d .

Example 5.1. Let θ be a random variable uniformly distributed in \mathbb{S} and let θ_n denote the solution of the polynomial equation

$$p_n + q_n^{(1)}\theta_n + \cdots + q_n^{(d)}\theta_n^d = 0$$

that is nearest to θ . We postulate the existence of numbers α and β such that, as $n \rightarrow \infty$,

$$(5.3) \quad e_n := \theta - \theta_n \sim c \exp(-n\alpha) \quad \text{and} \quad |\psi_n| \sim c' \exp(n\beta),$$

where

$$\psi_n(t) := p_n + q_n^{(1)}t + \cdots + q_n^{(d)}t^d.$$

Then the numbers

$$(5.4) \quad \alpha_n := -\frac{\ln |\theta - \theta_n|}{n} \quad \text{and} \quad \beta_n := \frac{\ln |\psi_n|}{n}$$

TABLE 2. Estimates of the exponents α_n and β_n , as defined by equation (5.4) and of the Roth exponent γ_n when naive rounding is used.

n	$d = 2$			$d = 3$			$d = 4$			$d = 5$		
	α_n	β_n	γ_n									
8	2.10	.78	1.39	.94	.14	7.88	.32	.09	4.73	.32	.09	4.73
16	1.41	.47	1.90	.47	.14	3.94	.47	.07	7.88	.47	.07	7.88
32	1.17	.38	2.06	.24	.10	4.86	.24	.06	5.48	.24	.06	5.48
64	1.00	.34	1.94	.59	.14	3.11	—	—	—	—	—	—
128	0.99	.33	2.01	.60	.14	3.06	—	—	—	—	—	—
256	1.05	.35	2.00	.57	.14	3.04	—	—	—	—	—	—
512	1.03	.34	2.00	.55	.14	3.00	—	—	—	—	—	—
1024	1.01	.34	2.00	.54	.13	3.02	—	—	—	—	—	—

should approximate the numers α and β respectively for n large. The results of a numerical experiment are displayed in Tables 2 and 3. The tables also show the Roth exponent

$$\gamma_n := -\frac{\ln |L_n|}{\ln |\psi_n|}$$

for various values of n and d .

Table 2 corresponds to the case where naive rounding is used. It is apparent that

$$\gamma_n \xrightarrow{n \rightarrow \infty} d, \quad d = 2, 3.$$

However, for $d = 4, 5$, the algorithm produces consecutive coefficient vectors \mathbf{c}_n that vanish. As a consequence, the (p_n, \mathbf{q}_n) eventually form a periodic sequence, and so there is no convergence.

Table 3 corresponds to the case where nearest-plane rounding is used. The results suggest that

$$\gamma_n \xrightarrow{n \rightarrow \infty} d, \quad d = 2, 3, 4, 5.$$

We have also carried out calculations up to $d = 8$; they are not shown in the table, but the results are qualitatively similar, namely that an optimal rate of convergence is obtained. Our experience also suggests that the nearest-plane procedure *never* produces trivial \mathbf{c}_n .

APPENDIX A. PROOF OF THEOREM 3.1

We shall need two lemmata.

Lemma A.1.

$$\left(\hat{\mathbf{L}}_n^{(i)}, \hat{\mathbf{L}}_n^{(j)} \right) = 0 \quad \text{for every } 1 \leq j < i \leq d_n.$$

Proof. By induction on i : For $i = 2$, we have

$$\frac{\left(\hat{\mathbf{L}}_n^{(2)}, \hat{\mathbf{L}}_n^{(1)} \right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)} \right)} = \frac{\left(\mathbf{L}_{n-1}, \hat{\mathbf{L}}_n^{(1)} \right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)} \right)} - \nu_n^{(2,1)} = 0.$$

TABLE 3. Estimates of the exponents α_n and β_n , as defined by equation (5.4) and of the Roth exponent γ_n when nearest-plane rounding is used.

n	$d = 2$			$d = 3$			$d = 4$			$d = 5$		
	α_n	β_n	γ_n									
8	3.15	.96	2.24	2.67	.63	3.20	2.24	.37	5.02	1.87	.29	5.59
16	2.56	.82	2.17	2.44	.59	3.18	2.33	.45	4.24	1.66	.22	6.37
32	2.37	.75	2.16	2.58	.62	3.19	2.29	.43	4.32	1.95	.34	4.85
64	2.37	.78	2.06	2.46	.63	2.93	2.30	.54	3.40	2.04	.33	5.11
128	2.37	.78	2.02	2.26	.55	3.10	2.30	.46	4.06	2.02	.34	5.04
256	2.34	.78	2.01	2.28	.57	3.02	2.28	.46	3.93	2.14	.38	4.58
512	2.38	.79	2.01	2.25	.56	2.99	2.20	.44	4.00	2.09	.35	5.01
1024	2.46	.82	2.00	2.26	.56	3.00	2.14	.43	4.00	2.08	.35	5.01

Next, suppose that the result holds up to and including some $2 \leq i < d_n$. We shall show that

$$\left(\hat{\mathbf{L}}_n^{(i+1)}, \hat{\mathbf{L}}_n^{(j)}\right) = 0 \quad \text{for every } 1 \leq j \leq i.$$

If $\hat{\mathbf{L}}_n^{(j)} = \mathbf{0}$, this is trivial. Otherwise,

$$\begin{aligned} \frac{\left(\hat{\mathbf{L}}_n^{(i+1)}, \hat{\mathbf{L}}_n^{(j)}\right)}{\left(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)}\right)} &= \frac{\left(\mathbf{L}_{n-i}, \hat{\mathbf{L}}_n^{(j)}\right)}{\left(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)}\right)} - \sum_{\ell=1}^i \nu_n^{(i+1, \ell)} \frac{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(j)}\right)}{\left(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)}\right)} \\ &= \frac{\left(\mathbf{L}_{n-i}, \hat{\mathbf{L}}_n^{(j)}\right)}{\left(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)}\right)} - \sum_{\ell=j}^i \nu_n^{(i+1, \ell)} \frac{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(j)}\right)}{\left(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)}\right)} \quad \text{by the induction hypothesis} \\ &= - \sum_{\ell=j+1}^i \nu_n^{(i+1, \ell)} \frac{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(j)}\right)}{\left(\hat{\mathbf{L}}_n^{(j)}, \hat{\mathbf{L}}_n^{(j)}\right)} \quad \text{by definition of } \nu_n^{(i+1, j)} \\ &= 0 \quad \text{by the induction hypothesis, since } j < \ell \leq i. \end{aligned}$$

□

Lemma A.2. For every $2 \leq j < d_n$,

$$\hat{\mathbf{L}}_n^{(j)} = \mathbf{0} \quad \implies \quad \left(\hat{\mathbf{q}}_n^{(i)}, \hat{\mathbf{q}}_n^{(j)}\right) = 0 \quad \text{for every } j < i \leq d_n.$$

Proof. Suppose that $\hat{\mathbf{L}}_n^{(j)} = \mathbf{0}$, and let $i > j$. By definition of $\hat{\mathbf{q}}_n^{(i)}$,

$$(A.1) \quad \frac{\left(\hat{\mathbf{q}}_n^{(i)}, \hat{\mathbf{q}}_n^{(j)}\right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)}\right)} = \frac{\left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^{i-1} \nu_n^{(i, \ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j)}\right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)}\right)}.$$

On the other hand, by definition of $\nu_n^{(i,j)}$, since $\hat{\mathbf{L}}_n^{(j)} = \mathbf{0}$,

$$\nu_n^{(i,j)} = \frac{\left(\mathbf{q}_{n-i+1} - \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^{i-1} \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j)} \right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)} \right)}.$$

Subtract this equation from equation (A.1) to obtain

$$(A.2) \quad \frac{\left(\hat{\mathbf{q}}_n^{(i)}, \hat{\mathbf{q}}_n^{(j)} \right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)} \right)} = - \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} = \mathbf{0}}}^{j-1} \nu_n^{(i,\ell)} \frac{\left(\hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j)} \right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)} \right)} - \sum_{\ell=j+1}^{i-1} \nu_n^{(i,\ell)} \frac{\left(\hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j)} \right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)} \right)}.$$

We use induction on j : For $j = 2$, this gives, since $\hat{\mathbf{L}}_n^{(1)} \neq \mathbf{0}$,

$$\frac{\left(\hat{\mathbf{q}}_n^{(i)}, \hat{\mathbf{q}}_n^{(2)} \right)}{\left(\hat{\mathbf{q}}_n^{(2)}, \hat{\mathbf{q}}_n^{(2)} \right)} = - \sum_{\ell=3}^{i-1} \nu_n^{(i,\ell)} \frac{\left(\hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(2)} \right)}{\left(\hat{\mathbf{q}}_n^{(2)}, \hat{\mathbf{q}}_n^{(2)} \right)}.$$

It is then straightforward to use induction (on i) to show that the right-hand side is zero for every $2 \leq i \leq d_n$.

Next, we make the induction hypothesis (on j), i.e. we suppose that the result contained in the lemma holds for every $2 \leq j \leq j'$, for some $2 \geq j' < d_n - 1$. If $\hat{\mathbf{L}}_n^{(j'+1)} = \mathbf{0}$, then we can set $j := j' + 1$ in equation (A.2) and use the induction hypothesis to obtain

$$\frac{\left(\hat{\mathbf{q}}_n^{(i)}, \hat{\mathbf{q}}_n^{(j)} \right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)} \right)} = - \sum_{\ell=j+1}^{i-1} \nu_n^{(i,\ell)} \frac{\left(\hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j)} \right)}{\left(\hat{\mathbf{q}}_n^{(j)}, \hat{\mathbf{q}}_n^{(j)} \right)}.$$

For $i = j + 1$, the sum is empty and so the right-hand side is zero. It then follows easily by induction on i that it is zero for every $i > j + 1$. This completes the induction on j , and the proof is done. \square

We are now in a position to prove the theorem. We have

$$\hat{\beta}_n^{(1)}(\varepsilon) = \mathcal{L}_n(\varepsilon) = \begin{pmatrix} \hat{\mathbf{L}}_n^{(1)} \\ \varepsilon \hat{\mathbf{q}}_n^{(1)} \end{pmatrix}$$

and so

$$\frac{\left(\mathcal{L}_{n-i+1}(\varepsilon), \hat{\beta}_n^{(1)}(\varepsilon) \right)}{\left(\hat{\beta}_n^{(1)}(\varepsilon), \hat{\beta}_n^{(1)}(\varepsilon) \right)} = \frac{\left(\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(1)} \right) + \varepsilon^2 \left(\mathbf{q}_{n-i+1}, \hat{\mathbf{q}}_n^{(1)} \right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)} \right) + \varepsilon^2 \left(\hat{\mathbf{q}}_n^{(1)}, \hat{\mathbf{q}}_n^{(1)} \right)}.$$

Hence, for $1 \leq i \leq d_n + 1$,

$$\begin{aligned} \mu_n^{(i,1)}(\varepsilon) &= \frac{1}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)}\right)} \frac{\left(\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(1)}\right) + \varepsilon^2 \left(\mathbf{q}_{n-i+1}, \hat{\mathbf{q}}_n^{(1)}\right)}{1 + \varepsilon^2 \frac{\left(\hat{\mathbf{q}}_n^{(1)}, \hat{\mathbf{q}}_n^{(1)}\right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)}\right)}} \\ &= \frac{1}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)}\right)} \left(\left(\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(1)}\right) + \varepsilon^2 \left(\mathbf{q}_{n-i+1}, \hat{\mathbf{q}}_n^{(1)}\right) \right) \\ &\quad \times \left(1 - \varepsilon^2 \frac{\left(\hat{\mathbf{q}}_n^{(1)}, \hat{\mathbf{q}}_n^{(1)}\right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)}\right)} + O(\varepsilon^4) \right), \end{aligned}$$

and so, after a little re-arrangement,

$$(A.3) \quad \mu_n^{(i,1)}(\varepsilon) = \nu_n^{(i,1)} + \varepsilon^2 \frac{\left(\mathbf{q}_{n-i+1} - \nu_n^{(i,1)} \hat{\mathbf{q}}_n^{(1)}, \hat{\mathbf{q}}_n^{(1)}\right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)}\right)} + O(\varepsilon^4) \quad \text{as } \varepsilon \rightarrow 0.$$

Next, we shall use induction on j to show that, for $1 \leq j \leq d_n$ and $j < i \leq d_n + 1$,

$$(A.4) \quad \begin{aligned} \mathcal{L}_{n-i+1}(\varepsilon) - \sum_{\ell=1}^j \mu_n^{(i,\ell)}(\varepsilon) \hat{\beta}_n^{(\ell)}(\varepsilon) &= \left(\begin{array}{c} \mathbf{L}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)} \right) \end{array} \right) \\ - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{\left(\mathbf{q}_{n-i+1} - \sum_{m=1}^j \nu_n^{(i,m)} \hat{\mathbf{q}}_n^{(m)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \left(\begin{array}{c} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \hat{\mathbf{q}}_n^{(\ell)} \end{array} \right) &+ \left(\begin{array}{c} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{array} \right) \quad \text{as } \varepsilon \rightarrow 0. \end{aligned}$$

By making use of equation (A.3), we have, for $j = 1$,

$$\begin{aligned} \mathcal{L}_{n-i+1}(\varepsilon) - \mu_n^{(i,1)}(\varepsilon) \hat{\beta}_n^{(1)}(\varepsilon) &= \left(\begin{array}{c} \mathbf{L}_{n-i+1} \\ \varepsilon \mathbf{q}_{n-i+1} \end{array} \right) \\ &\quad - \left(\nu_n^{(i,1)} + \varepsilon^2 \frac{\left(\mathbf{q}_{n-i+1} - \nu_n^{(i,1)} \hat{\mathbf{q}}_n^{(1)}, \hat{\mathbf{q}}_n^{(1)}\right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)}\right)} + O(\varepsilon^4) \right) \left(\begin{array}{c} \hat{\mathbf{L}}_n^{(1)} \\ \varepsilon \hat{\mathbf{q}}_n^{(1)} \end{array} \right) \\ &= \left(\begin{array}{c} \mathbf{L}_{n-i+1} - \nu_n^{(i,1)} \hat{\mathbf{L}}_n^{(1)} \\ \varepsilon \left(\mathbf{q}_{n-i+1} - \nu_n^{(i,1)} \hat{\mathbf{q}}_n^{(1)} \right) \end{array} \right) - \varepsilon^2 \frac{\left(\mathbf{q}_{n-i+1} - \nu_n^{(i,1)} \hat{\mathbf{q}}_n^{(1)}, \hat{\mathbf{q}}_n^{(1)}\right)}{\left(\hat{\mathbf{L}}_n^{(1)}, \hat{\mathbf{L}}_n^{(1)}\right)} \left(\begin{array}{c} \hat{\mathbf{L}}_n^{(1)} \\ \varepsilon \hat{\mathbf{q}}_n^{(1)} \end{array} \right) + \left(\begin{array}{c} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{array} \right) \end{aligned}$$

as $\varepsilon \rightarrow 0$. So equation (A.4) holds for $j = 1$.

Next, we make the induction hypothesis, namely that equation (A.4) holds for some $1 \leq j < d_n$. For $i = j + 1$, this yields

$$(A.5) \quad \hat{\beta}_n^{(j+1)}(\varepsilon) = \left(\begin{array}{c} \hat{\mathbf{L}}_n^{(j+1)} \\ \varepsilon \hat{\mathbf{q}}_n^{(j+1)} \end{array} \right) - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{\left(\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \left(\begin{array}{c} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \hat{\mathbf{q}}_n^{(\ell)} \end{array} \right) + \left(\begin{array}{c} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{array} \right)$$

as $\varepsilon \rightarrow 0$. Then

$$\begin{aligned}
 (\mathcal{L}_{n-i+1}(\varepsilon), \hat{\beta}_n^{(j+1)}(\varepsilon)) &= (\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(j+1)}) + \varepsilon^2 (\mathbf{q}_{n-i+1}, \hat{\mathbf{q}}_n^{(j+1)}) \\
 &\quad - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{(\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(\ell)})}{(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)})} (\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(\ell)}) + O(\varepsilon^4) \\
 &= (\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(j+1)}) + \varepsilon^2 \left(\mathbf{q}_{n-i+1} - \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j+1)} \right) + O(\varepsilon^4) \\
 &= (\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(j+1)}) + \varepsilon^2 \left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j+1)} \right) + O(\varepsilon^4) \quad \text{as } \varepsilon \rightarrow 0,
 \end{aligned}$$

by virtue of Lemma A.2. Also,

$$\begin{aligned}
 (\hat{\beta}_n^{(j+1)}(\varepsilon), \hat{\beta}_n^{(j+1)}(\varepsilon)) &= (\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(j+1)}) \\
 &\quad + \varepsilon^2 (\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(j+1)}) - 2\varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{(\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(\ell)})}{(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)})} (\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(\ell)}) + O(\varepsilon^4) \\
 &= (\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(j+1)}) + \varepsilon^2 (\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(j+1)}) + O(\varepsilon^4) \quad \text{as } \varepsilon \rightarrow 0,
 \end{aligned}$$

by virtue of Lemma A.1. It follows that

$$\begin{aligned}
 \text{(A.6)} \quad \mu_n^{(i,j+1)}(\varepsilon) &= \frac{(\mathcal{L}_{n-i+1}(\varepsilon), \hat{\beta}_n^{(j+1)}(\varepsilon))}{(\hat{\beta}_n^{(j+1)}(\varepsilon), \hat{\beta}_n^{(j+1)}(\varepsilon))} \\
 &= \frac{(\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(j+1)}) + \varepsilon^2 (\mathbf{q}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j+1)}) + O(\varepsilon^4)}{(\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(j+1)}) + \varepsilon^2 (\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(j+1)}) + O(\varepsilon^4)}
 \end{aligned}$$

as $\varepsilon \rightarrow 0$.

We need to consider the cases

$$\text{(a) } \hat{\mathbf{L}}_n^{(j+1)} \neq \mathbf{0} \quad \text{and} \quad \text{(b) } \hat{\mathbf{L}}_n^{(j+1)} = \mathbf{0}$$

separately. Suppose that (a) holds. Then equation (A.6) leads to

$$\begin{aligned}
 \mu_n^{(i,j+1)}(\varepsilon) &= \frac{(\mathbf{L}_{n-i+1}, \hat{\mathbf{L}}_n^{(j+1)}) + \varepsilon^2 (\mathbf{q}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j+1)}) + O(\varepsilon^4)}{(\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(j+1)})} \\
 &\quad \times \left(1 - \varepsilon^2 \frac{(\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(j+1)})}{(\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(j+1)})} + O(\varepsilon^4) \right)
 \end{aligned}$$

and so

$$(A.7) \quad \mu_n^{(i,j+1)}(\varepsilon) = \nu_n^{(i,j+1)} + \varepsilon^2 \frac{\left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j+1)}\right)}{\left(\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(j+1)}\right)} + O(\varepsilon^4)$$

as $\varepsilon \rightarrow 0$. Combining this with equation (A.5), we deduce that

$$(A.8) \quad \begin{aligned} \mu_n^{(i,j+1)}(\varepsilon) \hat{\beta}_n^{(j+1)}(\varepsilon) &= \nu_n^{(i,j+1)} \left(\frac{\hat{\mathbf{L}}_n^{(j+1)}}{\varepsilon \hat{\mathbf{q}}_n^{(j+1)}} \right) \\ &\quad - \nu_n^{(i,j+1)} \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{\left(\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \left(\frac{\hat{\mathbf{L}}_n^{(\ell)}}{\varepsilon \hat{\mathbf{q}}_n^{(\ell)}} \right) \\ &\quad + \varepsilon^2 \frac{\left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)}, \hat{\mathbf{q}}_n^{(j+1)}\right)}{\left(\hat{\mathbf{L}}_n^{(j+1)}, \hat{\mathbf{L}}_n^{(j+1)}\right)} \left(\frac{\hat{\mathbf{L}}_n^{(j+1)}}{\varepsilon \hat{\mathbf{q}}_n^{(j+1)}} \right) + \begin{pmatrix} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{pmatrix} \quad \text{as } \varepsilon \rightarrow 0. \end{aligned}$$

Now,

$$\begin{aligned} \mathcal{L}_{n-i+1}(\varepsilon) - \sum_{\ell=1}^{j+1} \mu_n^{(i,\ell)}(\varepsilon) \hat{\beta}_n^{(\ell)}(\varepsilon) &= \mathcal{L}_{n-i+1}(\varepsilon) - \sum_{\ell=1}^j \mu_n^{(i,\ell)}(\varepsilon) \hat{\beta}_n^{(\ell)}(\varepsilon) - \mu_n^{(i,j+1)}(\varepsilon) \hat{\beta}_n^{(j+1)}(\varepsilon) \\ &= \begin{pmatrix} \mathbf{L}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)} \right) \end{pmatrix} \\ &\quad - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{\left(\mathbf{q}_{n-i+1} - \sum_{m=1}^j \nu_n^{(i,m)} \hat{\mathbf{q}}_n^{(m)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \left(\frac{\hat{\mathbf{L}}_n^{(\ell)}}{\varepsilon \hat{\mathbf{q}}_n^{(\ell)}} \right) \\ &\quad - \mu_n^{(i,j+1)}(\varepsilon) \hat{\beta}_n^{(j+1)}(\varepsilon) + \begin{pmatrix} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{pmatrix} \quad \text{as } \varepsilon \rightarrow 0, \end{aligned}$$

by the induction hypothesis. So, by equation (A.8), this yields

$$\begin{aligned} \mathcal{L}_{n-i+1}(\varepsilon) - \sum_{\ell=1}^{j+1} \mu_n^{(i,\ell)}(\varepsilon) \hat{\beta}_n^{(\ell)}(\varepsilon) &= \begin{pmatrix} \mathbf{L}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)} \hat{\mathbf{q}}_n^{(\ell)} \right) \end{pmatrix} \\ &\quad - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^{j+1} \frac{\left(\mathbf{q}_{n-i+1} - \sum_{m=1}^{j+1} \nu_n^{(i,m)} \hat{\mathbf{q}}_n^{(m)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \left(\frac{\hat{\mathbf{L}}_n^{(\ell)}}{\varepsilon \hat{\mathbf{q}}_n^{(\ell)}} \right) + \begin{pmatrix} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{pmatrix} \quad \text{as } \varepsilon \rightarrow 0. \end{aligned}$$

This is equation (A.4) with j replaced by $j+1$, and so we have finished with the case where $\mathbf{L}_n^{(j+1)} \neq \mathbf{0}$.

It remains to consider the case where $\mathbf{L}_n^{(j+1)} = \mathbf{0}$. In this case, equation (A.6) gives

$$(A.9) \quad \mu_n^{(i,j+1)}(\varepsilon) = \nu_n^{(i,j+1)} + O(\varepsilon^2) \quad \text{as } \varepsilon \rightarrow 0.$$

Combining this with equation (A.5), we obtain

$$(A.10) \quad \mu_n^{(i,j+1)}(\varepsilon)\hat{\beta}_n^{(j+1)}(\varepsilon) = \nu_n^{(i,j+1)} \begin{pmatrix} \hat{\mathbf{L}}_n^{(j+1)} \\ \varepsilon\hat{\mathbf{q}}_n^{(j+1)} \end{pmatrix} - \nu_n^{(i,j+1)}\varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{\left(\hat{\mathbf{q}}_n^{(j+1)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \begin{pmatrix} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon\hat{\mathbf{q}}_n^{(\ell)} \end{pmatrix} + \begin{pmatrix} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{pmatrix} \quad \text{as } \varepsilon \rightarrow 0,$$

since $\hat{\mathbf{L}}_n^{(j+1)} = \mathbf{0}$. The induction hypothesis then gives

$$\begin{aligned} \mathcal{L}_{n-i+1}(\varepsilon) - \sum_{\ell=1}^{j+1} \mu_n^{(i,\ell)}(\varepsilon)\hat{\beta}_n^{(\ell)}(\varepsilon) &= \begin{pmatrix} \mathbf{L}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)}\hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^j \nu_n^{(i,\ell)}\hat{\mathbf{q}}_n^{(\ell)}\right) \end{pmatrix} \\ &\quad - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{\left(\mathbf{q}_{n-i+1} - \sum_{m=1}^j \nu_n^{(i,m)}\hat{\mathbf{q}}_n^{(m)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \begin{pmatrix} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon\hat{\mathbf{q}}_n^{(\ell)} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{L}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)}\hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)}\hat{\mathbf{q}}_n^{(\ell)}\right) \end{pmatrix} \\ &\quad - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^j \frac{\left(\mathbf{q}_{n-i+1} - \sum_{m=1}^{j+1} \nu_n^{(i,m)}\hat{\mathbf{q}}_n^{(m)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \begin{pmatrix} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon\hat{\mathbf{q}}_n^{(\ell)} \end{pmatrix} + \begin{pmatrix} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{L}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)}\hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon \left(\mathbf{q}_{n-i+1} - \sum_{\ell=1}^{j+1} \nu_n^{(i,\ell)}\hat{\mathbf{q}}_n^{(\ell)}\right) \end{pmatrix} \\ &\quad - \varepsilon^2 \sum_{\substack{\ell=1 \\ \hat{\mathbf{L}}_n^{(\ell)} \neq \mathbf{0}}}^{j+1} \frac{\left(\mathbf{q}_{n-i+1} - \sum_{m=1}^{j+1} \nu_n^{(i,m)}\hat{\mathbf{q}}_n^{(m)}, \hat{\mathbf{q}}_n^{(\ell)}\right)}{\left(\hat{\mathbf{L}}_n^{(\ell)}, \hat{\mathbf{L}}_n^{(\ell)}\right)} \begin{pmatrix} \hat{\mathbf{L}}_n^{(\ell)} \\ \varepsilon\hat{\mathbf{q}}_n^{(\ell)} \end{pmatrix} + \begin{pmatrix} \mathbf{O}_k(\varepsilon^4) \\ \mathbf{O}_{k'}(\varepsilon^3) \end{pmatrix} \end{aligned}$$

as $\varepsilon \rightarrow 0$, since $\hat{\mathbf{L}}_n^{(j+1)} = \mathbf{0}$. This is, again, equation (A.4) with j replaced by $j + 1$, and thus we have shown by induction that equation (A.4) holds for every $1 \leq j \leq d_n$. The proof of the theorem follows easily from equations (A.3), (A.5), (A.7) and (A.9).

REFERENCES

1. L. Babai, On Lovász' lattice reduction and the nearest lattice point problem, *Combinatorica* **6**, 1-13 (1986). MR856638 (88a:68049)
2. P. Baldwin, A multidimensional continued fraction and its statistical properties, *J. Stat. Phys.* **66**, 1463-1505 (1992). MR1156411 (93d:11082)
3. P. Baldwin, A convergence exponent for multidimensional continued fraction algorithms, *J. Stat. Phys.* **66**, 1507-1526 (1992). MR1156412 (93d:11083)
4. P. Billingsley, *Ergodic Theory and Information*, Wiley, New York, 1965. MR0192027 (33:254)

5. D. M. Hardcastle and K. Khanin, Continued fractions and the d -dimensional Gauss transformation, *Comm. Math. Phys.* **215**, 487-515 (2001). MR1810942 (2002a:11090)
6. B. Just, Generalizing the continued fraction algorithm to arbitrary dimensions, *SIAM J. Comput.* **21**, 909-926 (1992). MR1181407 (93k:11065)
7. K. Khanin, J. L. Lopes and J. Marklof, Multidimensional continued fractions, dynamical renormalization and KAM theory, *Comm. Math. Phys.* **270** (2007), 197-231. MR2276445
8. A. K. Lenstra, H. W. Lenstra & L. Lovász, Factoring polynomials with rational coefficients, *Math. Ann.* **261**, 515-534 (1982). MR682664 (84a:12002)
9. J. C. Lagarias, The quality of the Diophantine approximations found by the Jacobi-Perron algorithm and related algorithms, *Mh. Math.* **115**, 299-328 (1993). MR1230366 (94j:11074)
10. J. C. Lagarias, Geodesic multidimensional continued fractions, *Proc. London Math. Soc.* **69**, 464-488 (1994). MR1289860 (95j:11066)
11. G. J. Rieger, Ein Gauss-Kuzmin-Levy-Satz für Kettenbrüche nach nächsten Ganzen, *Manuscripta Math.* **24**, 437-438 (1978). MR0568143 (58:27875)
12. C. Rössner & C. Schnorr, An optimal, stable continued fraction algorithm for arbitrary dimension, *ECCC TR96-020*, (1996). <http://www.eccc.uni-trier.de/eccc>
13. F. Schweiger, *Multidimensional Continued Fractions*, Oxford University Press, Oxford, 2000. MR2121855 (2005i:11090)

SCHOOL OF MATHEMATICS, UNIVERSITY OF BRISTOL, BRISTOL BS8 1TW, UNITED KINGDOM
E-mail address: `y.tourigny@bristol.ac.uk`

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF BRISTOL, BRISTOL, UNITED KINGDOM
E-mail address: `n.p.smart@bristol.ac.uk`