# CHAPTER 3. ALGORITHMS FOR DIOPHANTINE APPROXIMATION.

## 3.1. Introduction.

In this section we give details of the computational methods we use to reduce upper bounds for the solutions of diophantine equations. Our starting point will always be a linear form $\Lambda$, that is close to 0 (in the real or p-adic sense, with the word "close" defined explicitly in terms of an inequality involving the unknowns), together with a large but explicitly known upper bound for the absolute values of the unknowns. Our aim is to reduce the upper bound by showing that there are no solutions between the new and the old upper bound.

Let $\vartheta_1, \ldots, \vartheta_n, \beta$ be given numbers, in $\mathbb{R}$, or in $\Omega_p$, for a fixed prime $p$. Let $x_1, \ldots, x_n$ be unknowns in $\mathbb{Z}$. Put

$$\Lambda = \beta + \sum_{i=1}^{n} x_i \cdot \vartheta_i \ .$$

We classify such linear forms according to three criteria:

→ homogeneous if $\beta = 0$, inhomogeneous if $\beta \neq 0$ ;

→ one-dimensional if $n = 2$, multi-dimensional if $n \geq 3$ ;

→ real if all the numbers are in $\mathbb{R}$, p-adic if all the numbers are in $\Omega_p$ .

The reason that the case $n = 2$ is called one-dimensional is that in the homogeneous case the linear form

$$\Lambda = x_1 \cdot \vartheta_1 + x_2 \cdot \vartheta_2$$

leads to studying the simple, one-dimensional continued fraction expansion of $-\vartheta_1/\vartheta_2$. The inhomogeneous case with $n = 1$, viz.

$$\Lambda = \beta + x \cdot \vartheta$$

is not of any interest in the real case, but it is of interest in the p-adic case. We call this the zero-dimensional case.

In the p-adic case we require that the quotients $\vartheta_i/\vartheta_j$ and $\beta/\vartheta_j$ are in

$\mathbb{Q}_p$ itself, whereas the numbers $\vartheta_i$, $\beta$ are allowed to be in some larger subfield of $\Omega_p$.

Let $c$, $\delta$ be positive constants. Put $X = \max|x_i|$. Let $X_0$ be a (large) positive constant. In the real case we shall always assume that

$$|\Lambda| < c\cdot\exp(-\delta\cdot X) , \tag{3.1}$$

$$X \le X_0 . \tag{3.2}$$

Let $c_1$, $c_2$ be real constants, with $c_2 > 0$. In the p-adic case we shall assume that $x_j > 0$ for some index $j \in \{1,\dots,n\}$, and

$$\mathrm{ord}_p(\Lambda) \ge c_1 + c_2\cdot x_j , \tag{3.3}$$

$$X \le X_0 . \tag{3.4}$$

Our aim is to find a constant $X_1$, of the size of $\log X_0$, such that in the real case (3.2) can be replaced by $X \le X_1$, and in the p-adic case the bound $x_j \le X_0$ (a consequence of (3.4)) can be improved to $x_j \le X_1$.

In the forthcoming sections we treat all cases, according to the classification given above. We insert Sections 3.4, 3.5 on the $L^3$-algorithm, which will be our main computational tool, Section 3.6 on finding short vectors in lattices, and Section 3.13 on certain sublattices that are useful for our applications.

## 3.2. Homogeneous one-dimensional approximation in the real case: continued fractions.

We first study the case

$$\Lambda = x_1\cdot\vartheta_1 + x_2\cdot\vartheta_2 .$$

Put $\vartheta = -\vartheta_1/\vartheta_2$. We assume that $\vartheta$ is irrational. Let the continued fraction expansion of $\vartheta$ be given by

$$\vartheta = [\ a_0,\ a_1,\ a_2,\ \dots\ ] ,$$

and let the convergents $p_n/q_n$ for $n = 0, 1, 2, \dots$ be defined by

$$\begin{cases} p_{-1} = 1 , & p_0 = a_0 , \quad p_{n+1} = a_{n+1}\cdot p_n + p_{n-1} \\ q_{-1} = 0 , & q_0 = 1 , \quad q_{n+1} = a_{n+1}\cdot q_n + q_{n-1} \end{cases} .$$

It is well known that the convergents satisfy the inequalities

$$\frac{1}{(a_{n+1}+2)\cdot q_n^2} < | \vartheta - \frac{p_n}{q_n} | < \frac{1}{a_{n+1}\cdot q_n^2} , \tag{3.5}$$

and that if $p/q$ satisfies the inequality

$$| \vartheta - \frac{p}{q} | < \frac{1}{2\cdot q^2} , \tag{3.6}$$

then $p/q$ must be one of the convergents (cf. Hardy and Wright [1979], Theorems 163, 171 and 184).

We may assume without loss of generality that $|\vartheta_1| < |\vartheta_2|$ , that $x_1 > 0$ , and that $(x_1,x_2) = 1$ . From (3.1) it follows that there exists a number $X^*$ such that $X \geq X^*$ implies $X = x_1$ and (3.6) for $(p,q) = (-x_2,x_1)$ . We now have the following criteria.

<u>LEMMA 3.1.</u> (i). *If* (3.1) *and* (3.2) *hold for* $x_1$, $x_2$ *with* $X \geq X^*$ , *then* $(-x_2,x_1) = (p_k,q_k)$ *for an index* $k$ *that satisfies*

$$k \leq -1 + \log\left(\sqrt{5}\cdot X_0+1\right)/\log\left(\tfrac{1}{2}(1+\sqrt{5})\right) . \tag{3.7}$$

*Moreover, the partial quotient* $a_{k+1}$ *satisfies*

$$a_{k+1} > -2 + |\vartheta_2|\cdot c^{-1}\cdot\exp(\delta\cdot q_k)/q_k . \tag{3.8}$$

(ii). *If for some* $k$ *with* $q_k \geq X^*$

$$a_{k+1} > |\vartheta_2|\cdot c^{-1}\cdot\exp(\delta\cdot q_k)/q_k , \tag{3.9}$$

*then* (3.1) *holds for* $(-x_2,x_1) = (p_k,q_k)$ .

<u>Proof.</u> (i). By $X \geq X^*$ and (3.6) it follows that $(-x_2,x_1) = (p_k,q_k)$ for an index $k$ . Since $q_k$ is at least the (k+1) th Fibonacci number, (3.7) follows from $q_k = x_1 = X \leq X_0$ . To prove (3.8), apply (3.1) and the first inequality of (3.5).

(ii). Combine (3.9) with the second inequality of (3.5). □

We may apply Lemma 3.1(i) directly, or as follows.

LEMMA 3.2. *Let*

$$A = \max(a_{k+1}) ,$$

*where the maximum is taken over all indices* $k$ *satisfying* (3.7). *If* (3.1) *and* (3.2) *hold for* $x_1, x_2$ *with* $X \geq X_1$ , *then*

$$X < \frac{1}{\delta}\cdot\log\left(c\cdot(A+2)/|\vartheta_2|\right) + \frac{1}{\delta}\cdot\log X .$$

Remark. From Lemma 3.2 an upper bound for $X$ follows. We can apply Lemma 2.1 here, but Lemma 2.1 is sharp for large $b$ only.

Proof. (3.1) and (3.5) yield

$$(a_{n+1}+2)\cdot q_n^2 > q_n\cdot|\vartheta_2|/|\Lambda| > q_n\cdot|\vartheta_2|\cdot c^{-1}\cdot\exp(\delta\cdot X) .$$

The result follows by applying Lemma 3.1(i). □

In practice it does not often occur that $A$ is large. Therefore this lemma is useful indeed.

Summarizing, this case comes down to computing the continued fraction of a real number to a certain precision, and establishing that it has no extremely large partial quotients. This idea has been applied in practice by Ellison [1971$^b$], by Cijsouw, Korlaar and Tijdeman (appendix to Stroeker and Tijdeman [1982]), and by Hunt and van der Poorten (unpublished) for solving diophantine equations, by Steiner [1977] in connection with the Syracuse ('$3\cdot N + 1$') problem, and by Cherubini and Walliser [1987] (using a small home computer only) for determining all imaginary quadratic number fields with class number 1. We shall use it in Chapters 4 and 5.

## 3.3. Inhomogeneous one-dimensional approximation in the real case: the Davenport lemma.

The next case is when $\Lambda$ has the form

$$\Lambda = \beta + x_1\cdot\vartheta_1 + x_2\cdot\vartheta_2 ,$$

where $\beta \neq 0$. We then use the so-called Davenport lemma, which was introduced by Baker and Davenport [1969]. It is, like the homogeneous one-dimensional case, based on the simple one-dimensional continued fraction algorithm.

Put again $\vartheta = -\vartheta_1/\vartheta_2$, and put $\psi = \beta/\vartheta_2$. Then we have

$$\frac{\Lambda}{\vartheta_2} = \psi - x_1\cdot\vartheta + x_2 .$$

Let $p/q$ be a convergent of $\vartheta$, with $q > X_0$. We now have the following result.

<u>LEMMA 3.3. (Davenport).</u> *Suppose that, in the above notation,*

$$\|q\cdot\psi\| > 2\cdot X_0/q , \tag{3.10}$$

*(by* $\|\cdot\|$ *we denote the distance to the nearest integer). Then the solutions of* (3.1), (3.2) *satisfy*

$$X < \frac{1}{\delta}\cdot\log\left(q^2\cdot c/|\vartheta_2|\cdot X_0\right) . \tag{3.11}$$

<u>Proof.</u> From (3.5) and (3.10) we infer

$$2\cdot X_0/q < \|q\cdot(\psi-x_1\cdot\vartheta+x_2)+x_1\cdot(q\cdot\vartheta-p)\| < q\cdot|\Lambda/\vartheta_2| + |x_1|/q .$$

By (3.1), (3.2), and

$$X_0 < q^2\cdot c\cdot|\vartheta_2^{-1}|\cdot\exp(-\delta\cdot X) ,$$

this leads to (3.11). □

If (3.10) is not true for the first convergent with denominator $> X_0$, then one should try some further convergents. If $q$ is not essentially larger than $X_0$, then (3.11) yields a reduced upper bound for $X$ of size $\log X_0$, as desired. If no $q$ of the size of $X_0$ can be found that also satisfies (3.10) (a situation which is very unlikely to occur, as experiments show), then not all is lost, since then only very few exceptional possible solutions have to be checked. See Baker and Davenport [1969] for details.

Summarizing, we see that in this case the essential idea is that an extremely large solution of (3.1) and (3.2) leads to a large range of convergents $p/q$

of $\vartheta$ for which the values of $\|q\cdot\psi\|$ are all extremely small. In practice it appears to be the case that $q\cdot\psi$ is always far enough from the nearest integer (the values of $\|q\cdot\psi\|$ seem to be distributed randomly over the interval $[0,0.5]$ ). This method has been used in practice by Baker and Davenport [1969] as we already mentioned, by Ellison, Ellison, Pesek, Stahl and Stall [1972], and by Steiner [1986]. We shall use it in Chapter 4.

## 3.4. The $L^3$-lattice basis reduction algorithm, theory.

To deal with linear forms with $n \geq 3$ , a straightforward generalization of the case $n = 2$ would be to study multi-dimensional continued fractions. For a good survey of this field, see Brentjes [1981]. However, the available algorithms in this field seem not to have the desired efficiency and generality. Fortunately, since 1981 there is a useful alternative, which in a sense is also a generalization of the one-dimensional continued fraction algorithm.

In 1981, L. Lovász invented an algorithm, that has since then become known as the $L^3$-algorithm. It has been published in Lenstra, Lenstra and Lovász [1982], Fig. 1, p. 521. Throughout this and the next section we refer to this paper as "$\mathcal{LLL}$". The algorithm computes from an arbitrary basis of a lattice in $\mathbb{R}^n$ another basis of this lattice, a so-called *reduced* basis, which has certain nice properties (its vectors are nearly orthogonal).

The algorithm has many important applications in a variety of mathematical fields, such as the factorization of polynomials ($\mathcal{LLL}$), public-key cryptography (Lagarias and Odlyzko [1985]), and the disproof of the Mertens Conjecture (Odlyzko and te Riele [1985]). Of interest to us are its applications to diophantine approximation, which already had been noticed in $\mathcal{LLL}$, p. 525. The algorithm has a very good theoretical complexity (polynomial-time in the length of the input parameters), and performs also very well in practical computations.

Let $\Gamma \subset \mathbb{R}^n$ be a lattice, that is given by the basis $\underline{b}_1, \ldots, \underline{b}_n$ . We introduce the concept of a *reduced* basis of $\Gamma$ , according to $\mathcal{LLL}$, p.516. The vectors $\underline{b}_i^*$ ( $i = 1, \ldots, n$ ) and the real numbers $\mu_{i,j}$ ( $1 \leq j < i \leq n$ ) are inductively defined by

$$\underline{b}_i^* = \underline{b}_i - \sum_{j=1}^{i-1} \mu_{i,j} \cdot \underline{b}_j^* \ , \quad \mu_{i,j} = (\underline{b}_i, \underline{b}_j^*) \ / \ (\underline{b}_j^*, \underline{b}_j^*) \ .$$

Then $\underline{b}_1^*, \ldots, \underline{b}_n^*$ is an orthogonal basis of $\mathbb{R}^n$ . We call the lattice basis $\underline{b}_1, \ldots, \underline{b}_n$ of $\Gamma$ *reduced* if

$$|\mu_{i,j}| \le \tfrac{1}{2} \quad \text{for} \quad 1 \le j < i \le n \ ,$$

$$|\underline{b}_i^* + \mu_{i,i-1} \cdot \underline{b}_{i-1}^*|^2 \ge \tfrac{3}{4} \cdot |\underline{b}_{i-1}^*|^2 \quad \text{for} \quad 1 < i \le n \ .$$

Hence a reduced basis is nearly orthogonal. For a reduced basis $\underline{b}_1, \ldots, \underline{b}_n$ we have, by $\mathcal{LLL}$ (1.7),

$$|\underline{b}_i^*| \ge 2^{-(n-1)/2} \cdot |\underline{b}_1| \quad \text{for} \quad i = 1, \ldots, n \ . \tag{3.12}$$

We remark that a lattice may have more than one reduced basis, and that the ordering of the basis vectors is not arbitrary. The $L^3$-algorithm accepts as input any basis $\underline{b}_1, \ldots, \underline{b}_n$ of $\Gamma$ , and it computes a reduced basis $\underline{c}_1, \ldots, \underline{c}_n$ of that lattice. The properties of reduced bases that are of most interest to us are the following. Let $\underline{y} \in \mathbb{R}^n$ be a given point, that is not a lattice point. We denote by $\ell(\Gamma)$ the length of the shortest non-zero vector in the lattice, viz.

$$\ell(\Gamma) = \min_{\underline{0} \ne \underline{x} \in \Gamma} |\underline{x}| \ ,$$

and we denote by $\ell(\Gamma, \underline{y})$ the distance from $\underline{y}$ to the lattice point nearest to it, viz.

$$\ell(\Gamma, \underline{y}) = \min_{\underline{x} \in \Gamma} |\underline{x} - \underline{y}| \ .$$

From a reduced basis lower bounds for both $\ell(\Gamma)$ and $\ell(\Gamma, \underline{y})$ can be computed, according to the following results.

LEMMA 3.4. (Lenstra, Lenstra and Lovasz [1982]). *Let* $\underline{c}_1, \ldots, \underline{c}_n$ *be a reduced basis of the lattice* $\Gamma$ . *Then*

$$\ell(\Gamma) \ge 2^{-(n-1)/2} \cdot |\underline{c}_1| \ .$$

Proof. This is Proposition (1.11) from $\mathcal{LLL}$. We recall the proof here. Let $\underline{0} \ne \underline{x} \in \Gamma$ be the lattice point with minimal length $|\underline{x}| = \ell(\Gamma)$ . Write

$$\underline{x} = \sum_{i=1}^{n} r_i \cdot \underline{c}_i = \sum_{i=1}^{n} r_i^* \cdot \underline{b}_i^* ,$$

with $r_i \in \mathbb{Z}$ , $r_i^* \in \mathbb{R}$ . Let $i_0$ be the largest index such that $r_{i_0} \neq 0$ . Then, since $\underline{c}_1, \ldots, \underline{c}_i$ span the same linear space as $\underline{b}_1^*, \ldots, \underline{b}_i^*$ for all $i$ , and $\underline{b}_{i+1}^*$ is the projection of $\underline{c}_{i+1}$ on the orthogonal complement of this linear space, it follows that $r_{i_0} = r_{i_0}^*$ . Hence, by (3.12),

$$\ell(\Gamma)^2 = |\underline{x}|^2 = \sum_{i=1}^{i_0} r_i^{*2} \cdot |\underline{b}_i^*|^2 \geq r_{i_0}^{*2} \cdot |\underline{b}_{i_0}^*|^2 = r_{i_0}^2 \cdot |\underline{b}_{i_0}^*|^2$$

$$\geq |\underline{b}_{i_0}^*|^2 \geq 2^{-(n-1)} \cdot |\underline{c}_1|^2 . \qquad \square$$

LEMMA 3.5. *Let* $\underline{c}_1, \ldots, \underline{c}_n$ *be a reduced basis of the lattice* $\Gamma$ *, and let* $\underline{y} = \sum_{i=1}^{n} s_i \cdot \underline{c}_i$ *for* $s_1, \ldots, s_n \in \mathbb{R}$ *, with not all* $s_i$ *in* $\mathbb{Z}$ *. Let* $i_0$ *be the largest index such that* $s_{i_0} \notin \mathbb{Z}$ *. Then*

$$\ell(\Gamma, \underline{y}) \geq 2^{-(n-1)/2} \cdot \| s_{i_0} \| \cdot |\underline{c}_1| .$$

Proof. The proof of this lemma resembles that of Lemma 3.4. Let $\underline{x} \in \Gamma$ be the lattice point nearest to $\underline{y}$ . So $|\underline{x}-\underline{y}| = \ell(\Gamma, \underline{y})$ . Write

$$\underline{x} = \sum_{i=1}^{n} r_i \cdot \underline{c}_i = \sum_{i=1}^{n} r_i^* \cdot \underline{b}_i^* , \quad \underline{y} = \sum_{i=1}^{n} s_i \cdot \underline{c}_i = \sum_{i=1}^{n} s_i^* \cdot \underline{b}_i^* ,$$

with $r_i \in \mathbb{Z}$ , $r_i^*, s_i, s_i^* \in \mathbb{R}$ . Let $i_1$ be the largest index such that $r_{i_1} \neq s_{i_1}$ . Then, reasoning as in the proof of Lemma 3.4, we find

$$r_{i_1} - s_{i_1} = r_{i_1}^* - s_{i_1}^* .$$

Using (3.12) it follows that

$$\ell(\Gamma, \underline{y})^2 \geq (r_{i_1} - s_{i_1})^2 \cdot |\underline{b}_{i_1}^*|^2 \geq (r_{i_1} - s_{i_1})^2 \cdot 2^{-(n-1)} \cdot |\underline{c}_1|^2 .$$

Obviously, $i_1 \geq i_0$ . If $i_1 = i_0$ the result follows at once. If $i_1 > i_0$ then $s_{i_1} \in \mathbb{Z}$, $s_{i_1} \neq r_{i_1}$ , hence $|r_{i_1} - s_{i_1}| \geq 1$ , and the result follows. □

The above lemma is rather weak in the extraordinary situation that $s_{i_0}$ is extremely close to an integer. If one of the other $s_i$ is not close to an integer, we can apply the following variant.

<u>LEMMA 3.6.</u> *Let* $\underline{c}_1, \ldots, \underline{c}_n$ *be a reduced basis of the lattice* $\Gamma$ , *and let* $\underline{y} = \sum_{i=1}^{n} s_i \cdot \underline{c}_i$ *for* $s_1, \ldots, s_n \in \mathbb{R}$ , *with not all* $s_i$ *in* $\mathbb{Z}$ . *Suppose that there is an index* $i_0$ *and constants* $\delta_1$ , $0 < \delta_2 \leq \frac{1}{2}$ *such that*

$$\|s_i\| \leq \delta_1 \quad \text{for} \quad i = i_0+1, \ldots, n ,$$

$$\|s_{i_0}\| \geq \delta_2 .$$

*Then*

$$\ell(\Gamma,\underline{y}) \geq 2^{-(n-1)/2} \cdot \delta_2 \cdot |\underline{c}_1| - (n-i_0) \cdot \delta_1 \cdot \max_{i>i_0} |\underline{c}_i| .$$

<u>Proof.</u> With notation as in the proof of Lemma 3.5, let $t_i$ be the integer nearest to $s_i$ , for $i \geq i_0 + 1$ , and $t_i = s_i$ for $i \leq i_0$ . Put

$$\underline{z} = \sum_{i=1}^{n} t_i \cdot \underline{c}_i = \sum_{i=1}^{n} t_i^* \cdot \underline{b}_i^*$$

with $t_i^* \in \mathbb{R}$ . Let $i_1$ be the largest index such that $r_{i_1} \neq t_{i_1}$ . Then

$$r_{i_1} - t_{i_1} = r_{i_1}^* - t_{i_1}^* .$$

We have

$$\ell(\Gamma,\underline{y}) = |\underline{x}-\underline{y}| \geq |\underline{x}-\underline{z}| - |\underline{z}-\underline{y}| .$$

Now,

$$|\underline{z}-\underline{y}| \leq \sum_{i=i_0+1}^{n} |s_i-t_i| \cdot |\underline{c}_i| \leq (n-i_0) \cdot \delta_1 \cdot \max_{i>i_0} |\underline{c}_i| ,$$

and, using (3.12),

$$\begin{aligned} |\underline{x}-\underline{z}|^2 = \sum_{i=1}^{n} (r_i^*-t_i^*)^2 \cdot |\underline{b}_i^*|^2 &\geq (r_{i_1}^*-t_{i_1}^*)^2 \cdot |\underline{b}_{i_1}^*|^2 \\ &\geq (r_{i_1}-t_{i_1})^2 \cdot 2^{-(n-1)} \cdot |\underline{c}_1|^2 . \end{aligned}$$

Obviously, $i_1 \geq i_0$ . If $i_1 = i_0$ the result follows at once. If $i_1 > i_0$ then $t_{i_1} \in \mathbb{Z}$, $t_{i_1} \neq r_{i_1}$ , hence $|r_{i_1} - t_{i_1}| \geq 1 > \delta_2$ , and the result follows. □

<u>Remark.</u> Babai [1986] showed that the $L^3$-algorithm can be used to find a lattice point $\underline{x}$ with $|\underline{x}-\underline{y}| \leq c \cdot \ell(\Gamma,\underline{y})$ for a constant $c$ depending on the dimension of the lattice only. This result can also be used instead of Lemma 3.5 or 3.6.

### 3.5. The $L^3$-lattice basis reduction algorithm, practice.

Below we describe the variant of the $L^3$-algorithm that we use in this thesis to solve diophantine equations. This variant has been designed to work with integers only, so that rounding-off errors are avoided completely. In the algorithm as stated in $\mathcal{LLL}$, Fig. 1, p. 521, non-integral rational numbers may occur, even if the input parameters are all integers.

Let $\Gamma \subset \mathbb{Z}^n$ be a lattice with basis vectors $\underline{b}_1, \ldots, \underline{b}_n$ . Define $\underline{b}_i^*$, $\mu_{ij}$, $d_i$ as in $\mathcal{LLL}$ (1.2), (1.3), (1.24), respectively. The $d_i$ can be used as denominators for all numbers that appear in the original algorithm ($\mathcal{LLL}$, p. 523). Thus, put for all relevant indices $i$, $j$

$$\underline{c}_i = d_{i-1} \cdot \underline{b}_i^* ,$$
$$\lambda_{i,j} = d_j \cdot \mu_{i,j} . \tag{3.13}$$

They are integral, by $\mathcal{LLL}$ (1.28), (1.29). Notice that, with $B_i = |\underline{b}_i^*|^2$ ,

$$d_i = d_{i-1} \cdot B_i . \tag{3.14}$$

We can now rewrite the algorithm in terms of $\underline{c}_i$, $d_i$, $\lambda_{i,j}$ in stead of $\underline{b}_i^*$, $B_i$, $\mu_{i,j}$ , thus eliminating all non-integral rationals. We give this variant of the $L^3$-algorithm in Fig. 1. All the lines in this variant are evident from applying (3.13) and (3.14) to the corresponding lines in the original algorithm, except the lines (A), (B) and (C), which will be explained below.

We added a few lines to the algorithm, in order to compute the matrix of the transformation from the initial to the reduced basis. Let $\mathcal{B}$ be the matrix with column vectors $\underline{b}_1, \ldots, \underline{b}_n$ , the initial basis of the lattice $\Gamma$ ,

Figure 1. Variant of the $L^3$-algorithm.

```
         $d_0 := 1$ ;
         $\underline{c}_i := \underline{b}_i$ ;
             $\lambda_{i,j} := (\underline{b}_i,\underline{c}_j)$ ;
(A)          $\underline{c}_i := (d_j\cdot\underline{c}_i-\lambda_{i,j}\cdot\underline{c}_j)/d_{j-1}$
             } for $j=1,\dots,i-1$ ;
         $d_i := (\underline{c}_i,\underline{c}_i)/d_{i-1}$
         } for $i=1,\dots,n$ ;

         $k := 2$ ;
    (1)  perform (*) for $\ell = k-1$ ;
         if $4\cdot d_{k-2}\cdot d_k < 3\cdot d_{k-1}^2 - 4\cdot\lambda_{k,k-1}^2$ go to (2) ;
         perform (*) for $\ell = k-2, \dots, 1$ ;
         if $k = n$ terminate ;
         $k := k+1$ ;  go to (1) ;

    (2)  $\begin{pmatrix}\underline{b}_{k-1}\\ \underline{b}_k\end{pmatrix} := \begin{pmatrix}\underline{b}_k\\ \underline{b}_{k-1}\end{pmatrix}$ ;
         $\begin{pmatrix}\underline{u}_{k-1}\\ \underline{u}_k\end{pmatrix} := \begin{pmatrix}\underline{u}_k\\ \underline{u}_{k-1}\end{pmatrix}$ ;  $\begin{pmatrix}\underline{v}_{k-1}'^{T}\\ \underline{v}_k'^{T}\end{pmatrix} := \begin{pmatrix}\underline{v}_k'^{T}\\ \underline{v}_{k-1}'^{T}\end{pmatrix}$ ;
         $\begin{pmatrix}\lambda_{k-1,j}\\ \lambda_{k,j}\end{pmatrix} := \begin{pmatrix}\lambda_{k,j}\\ \lambda_{k-1,j}\end{pmatrix}$  for $j = 1, \dots, k-2$ ;
(B)      $\begin{pmatrix}\lambda_{i,k-1}\\ \lambda_{i,k}\end{pmatrix} := \left(\lambda_{i,k-1}\cdot\begin{pmatrix}\lambda_{k,k-1}\\ d_k\end{pmatrix} + \lambda_{i,k}\cdot\begin{pmatrix}d_{k-2}\\ -\lambda_{k,k-1}\end{pmatrix}\right) / d_{k-1}$
                                                        for $i = k+1, \dots, n$ ;
(C)      $d_{k-1} := (d_{k-2}\cdot d_k + \lambda_{k,k-1}^2) / d_{k-1}$ ;
         if $k > 2$ then $k := k-1$ ;
         go to (1) ;

    (*)  if $2\cdot|\lambda_{k,\ell}| > d_\ell$ then
         {  $r :=$ integer nearest to $\lambda_{k,\ell}/d_\ell$ ;
            $\underline{b}_k := \underline{b}_k - r\cdot\underline{b}_\ell$ ;  $\underline{u}_k := \underline{u}_k - r\cdot\underline{u}_\ell$ ;  $\underline{v}_\ell'^{T} := \underline{v}_\ell'^{T} + r\cdot\underline{v}_k'^{T}$ ;
            $\lambda_{k,j} := \lambda_{k,j} - r\cdot\lambda_{\ell,j}$  for $j = 1, \dots, \ell-1$ ;
            $\lambda_{k,\ell} := \lambda_{k,\ell} - r\cdot d_\ell$ .
```

which is the input for the algorithm. We say: $\mathcal{B}$ is the matrix *associated to* the basis $\underline{b}_1, \dots, \underline{b}_n$ . Let $\mathcal{C}$ be the matrix associated to the reduced basis $\underline{c}_1, \dots, \underline{c}_n$ , which the algorithm delivers as output. Then we define this transformation matrix $\mathcal{V}$ by

$$\mathcal{C} = \mathcal{B} \cdot \mathcal{V} .$$

More generally, let $\mathcal{U}$ be the matrix of a transformation from some $\mathcal{B}_0$ to $\mathcal{B}$ , so $\mathcal{B} = \mathcal{B}_0 \cdot \mathcal{U}$ . Denote the column vectors of $\mathcal{U}$ by $\underline{u}_1, \dots, \underline{u}_n$ , and the row vectors of $\mathcal{U}^{-1}$ by $\underline{v}_1'^{T}, \dots, \underline{v}_n'^{T}$ . We feed the algorithm with $\mathcal{U}$ and $\mathcal{U}^{-1}$ also. All manipulations in the algorithm done on the $\underline{b}_i$ are also done on the $\underline{u}_i$ , and the $\underline{v}_i'^{T}$ are adjusted accordingly. This does not affect the computation time seriously. The algorithm now gives as output matrices $\mathcal{C}$ , $\mathcal{U}'$ and $\mathcal{U}'^{-1}$ , such that $\mathcal{C}$ is associated to a reduced basis, $\mathcal{C} = \mathcal{B} \cdot \mathcal{V}$ , and $\mathcal{U}' = \mathcal{U} \cdot \mathcal{V}$ . Note that $\mathcal{V}$ is not computed explicitly, unless $\mathcal{U} = \mathcal{I}$ (the unit matrix), in which case $\mathcal{U}' = \mathcal{V}$ . It follows that

$$\mathcal{C} = \mathcal{B} \cdot \mathcal{U}^{-1} \cdot \mathcal{U}' = \mathcal{B}_0 \cdot \mathcal{U}' ,$$

so $\mathcal{U}'$ is the matrix of the transformation from $\mathcal{B}_0$ to $\mathcal{C}$ . Note that if $\mathcal{B}_0^{-1}$ is known, then it is not much extra effort to compute $\mathcal{C}^{-1}$ as well.

We now explain why lines (A), (B) and (C) are correct.

(A): From $\mathcal{LLL}$ (1.2) it follows that

$$\underline{c}_i = d_{i-1} \cdot \underline{b}_i - \sum_{k=1}^{i-1} \frac{d_{i-1}}{d_{k-1} \cdot d_k} \cdot \lambda_{i,k} \cdot \underline{c}_k .$$

Define for $j = 0, 1, \dots, i-1$

$$\underline{c}_i(j) = d_j \cdot \underline{b}_i - \sum_{k=1}^{j} \frac{d_j}{d_{k-1} \cdot d_k} \cdot \lambda_{i,k} \cdot \underline{c}_k .$$

Then $\underline{c}_i(0) = \underline{b}_i$ , and $\underline{c}_i(i-1) = \underline{c}_i$ . The $\underline{c}_i(j)$ is exactly the vector computed in (A) at the $j$ th step, since

$$\frac{d_j \cdot \underline{c}_i(j-1) - \lambda_{i,j} \cdot \underline{c}_j}{d_{j-1}}$$

$$= d_j \cdot \underline{b}_i - \sum_{k=1}^{j-1} \frac{d_j}{d_{k-1} \cdot d_k} \cdot \lambda_{i,k} \cdot \underline{c}_k - \frac{d_j}{d_{j-1} \cdot d_j} \cdot \lambda_{i,j} \cdot \underline{c}_j = \underline{c}_i(j) .$$

This explains the recursive formula in line (A). It remains to show that the occurring vectors $\underline{c}_i(j)$ are integral. This follows from

$$d_j \cdot \sum_{k=1}^{j} \frac{1}{d_{k-1} \cdot d_k} \cdot \lambda_{i,k} \cdot \underline{c}_k = d_j \cdot \sum_{k=1}^{j} \mu_{i,k} \cdot \underline{b}_k^* ,$$

which is integral by *ℒℒℒ* p. 523, *ℓ*. 11.

(B), (C): Notice that the third and fourth line, starting from label (2), in the original algorithm, are independent of the first, second and fifth line. Thus a permutation of these lines is allowed. We rewrite the first, second and fifth line as follows, where we indicate variables that have been changed with a prime sign.

$$B'_{k-1} := B_k + \mu^2_{k,k-1} \cdot B_{k-1} ; \qquad (3.15)$$

$$B'_k := B_{k-1} \cdot B_k / B'_{k-1} ; \qquad (3.16)$$

$$\mu'_{k,k-1} := \mu_{k,k-1} \cdot B_{k-1} / B'_{k-1} ; \qquad (3.17)$$

$$\mu'_{i,k-1} := \mu'_{k,k-1} \cdot \mu_{i,k-1} + (1 - \mu_{k,k-1} \cdot \mu'_{k,k-1}) \cdot \mu_{i,k} ; \qquad (3.18)$$

$$\mu'_{i,k} := \mu_{i,k-1} - \mu_{k,k-1} \cdot \mu_{i,k} ; \qquad (3.19)$$

where (3.18) and (3.19) hold for $i = k+1, \ldots, n$. The $d_i$ remain unchanged for $i = 0, 1, \ldots, k-2$, and by (3.16) also for $i = k$. Now, (3.15) is equivalent to

$$\frac{d'_{k-1}}{d_{k-2}} = \frac{d_k}{d_{k-1}} + \frac{\lambda^2_{k,k-1}}{d^2_{k-1}} \cdot \frac{d_{k-1}}{d_{k-2}} , \qquad (3.20)$$

which explains (C). From (3.17) we find

$$\frac{\lambda'_{k,k-1}}{d'_{k-1}} = \frac{\lambda_{k,k-1}}{d_{k-1}} \cdot \frac{d_{k-1}}{d_{k-2}} \cdot \frac{d'_{k-2}}{d'_{k-1}} ,$$

hence $\lambda_{k,k-1}$ remains unchanged. From (3.18) we obtain

$$\frac{\lambda'_{i,k-1}}{d'_{k-1}} = \frac{\lambda_{k,k-1}}{d'_{k-1}} \cdot \frac{\lambda_{i,k-1}}{d_{k-1}} + \left[ 1 - \frac{\lambda_{k,k-1}}{d_{k-1}} \cdot \frac{\lambda_{k,k-1}}{d'_{k-1}} \right] \cdot \frac{\lambda_{i,k}}{d_k} ,$$

whence, by multiplying by $d_{k-1} \cdot d'_{k-1}$ and using (3.20),

$$d_{k-1} \cdot \lambda'_{i,k-1} = \lambda_{k,k-1} \cdot \lambda_{i,k-1} + ( d_{k-1} \cdot d'_{k-1} - \lambda^2_{k,k-1} ) \cdot \frac{\lambda_{i,k}}{d_k}$$

$$= \lambda_{k,k-1} \cdot \lambda_{i,k-1} + d_{k-2} \cdot \lambda_{i,k} \ .$$

Finally, from (3.19) we see

$$\frac{\lambda'_{i,k}}{d_k} = \frac{\lambda_{i,k-1}}{d_{k-1}} - \frac{\lambda_{k,k-1}}{d_{k-1}} \cdot \frac{\lambda_{i,k}}{d_k} \ ,$$

and (B) follows.

In our applications we often have a lattice $\Gamma$, of which a basis is given such that the associated matrix, $\mathcal{A}$ say, has the special form

$$\mathcal{A} = \begin{bmatrix} 1 & & & \\ & \ddots & & \emptyset \\ \emptyset & & 1 & \\ \theta_1 & \cdots & \theta_{n-1} & \theta_n \end{bmatrix} ,$$

where the $\theta_i$ are large integers, that may have several hundreds of decimal digits. We can compute a reduced basis of this lattice directly, using the matrix $\mathcal{A}$ itself as input for the $L^3$-algorithm. But it may save time and space to split up the computation into several steps with increasing accuracy, as follows.

Let $k$ be a natural number (the number of steps), and let $\ell$ be a natural number such that the $\theta_i$ have about $k \cdot \ell$ (decimal) digits. For $i = 1, \ldots, n$ and $j = 1, \ldots, k$ put

$$\theta_i^{(j)} = [\theta_i / 10^{\ell \cdot (k-j)}] \ ,$$

and define $\Psi_i^{(j)}$ by

$$\theta_i^{(j+1)} = 10^{\ell} \cdot \theta_i^{(j)} + \Psi_i^{(j)} \ .$$

Thus, the $\Psi_i^{(j)}$ are blocks of $\ell$ consecutive digits of $\theta_i$. Define for the relevant $j$ the $n \times n$ matrices

$$\mathcal{A}_j = \begin{bmatrix} 1 & & & \\ & \ddots & & \emptyset \\ \emptyset & & 1 & \\ \theta_1^{(j)} & \cdots & \theta_{n-1}^{(j)} & \theta_n^{(j)} \end{bmatrix} , \quad \mathcal{D}_j = \begin{bmatrix} & & \\ & \emptyset & \\ \Psi_1^{(j)} & \cdots & \Psi_n^{(j)} \end{bmatrix} ,$$

$$\mathcal{E} = \begin{bmatrix} 1 & & & \\ & \ddots & & \emptyset \\ & \emptyset & 1 & \\ & & & 10^{\ell} \end{bmatrix} .$$

Then it follows at once that

$$\mathcal{A}_{j+1} = \mathcal{E}\cdot\mathcal{A}_j + \mathcal{D}_j .$$

Notice that $\mathcal{A}_k = \mathcal{A}$, since $\theta_i^{(k)} = \theta_i$. Put $\mathcal{U}_0 = \mathcal{I}$, $\mathcal{B}_1 = \mathcal{A}_1$. For some $j \geq 1$ let $\mathcal{B}_j$ and $\mathcal{U}_{j-1}$ be known matrices. Then we apply the $L^3$-algorithm to $\mathcal{B} = \mathcal{B}_j$, $\mathcal{U} = \mathcal{U}_{j-1}$, and $\mathcal{U}^{-1}$. We thus find matrices $\mathcal{C}_j$, $\mathcal{U}_j$ and $\mathcal{U}_j^{-1}$ such that

$$\mathcal{C}_j = \mathcal{B}_j\cdot\mathcal{U}_{j-1}^{-1}\cdot\mathcal{U}_j .$$

Now put

$$\mathcal{B}_{j+1} = \mathcal{E}\cdot\mathcal{C}_j + \mathcal{D}_j\cdot\mathcal{U}_j .$$

By induction $\mathcal{B}_j$, $\mathcal{C}_j$ and $\mathcal{U}_j$ are defined for $j = 1, \ldots, k$. Note that

$$\mathcal{B}_{j+1}\cdot\mathcal{U}_j^{-1} = \mathcal{E}\cdot\mathcal{B}_j\cdot\mathcal{U}_{j-1}^{-1} + \mathcal{D}_j ,$$

so the $\mathcal{B}_j\cdot\mathcal{U}_{j-1}^{-1}$ satisfy the same recursive relation as the $\mathcal{A}_j$. Since $\mathcal{B}_1\cdot\mathcal{U}_0^{-1} = \mathcal{A}_1$, we have $\mathcal{B}_j\cdot\mathcal{U}_{j-1}^{-1} = \mathcal{A}_j$ for all $j$. Hence

$$\mathcal{C}_j = \mathcal{B}_j\cdot\mathcal{U}_{j-1}^{-1}\cdot\mathcal{U}_j = \mathcal{A}_j\cdot\mathcal{U}_j ,$$

and it follows that $\mathcal{C}_k$ and $\mathcal{A}_k$ are associated to bases of the same lattice, which is $\Gamma$. Moreover, since $\mathcal{C}_k$ is output of the $L^3$-algorithm, it is associated to a reduced basis of $\Gamma$.

Let us now analyse the computation time. For a matrix $\mathcal{M}$ we denote by $L(\mathcal{M})$ the maximal number of (decimal) digits of its entries. If the $L^3$-algorithm is applied to a matrix $\mathcal{B}$, with as output a matrix $\mathcal{C}$, then according to the experiences of Lenstra, Odlyzko (cf. Lenstra [1984], p. 7) and ourselves, the computation time is proportional to $L(\mathcal{B})^3$ in practice. Since $\mathcal{C}$ is associated to a reduced basis, we assume that

$$L(\mathcal{C}) \cong {}^{10}\log(\det\Gamma)/n .$$

In our situation, $L(\mathcal{A}_j) \cong \ell\cdot j$, $L(\mathcal{D}_j) \cong \ell$, and since $\det\mathcal{C}_j = \det\mathcal{A}_j =$

$\theta_n^{(j)}$, we have $L(\mathcal{C}_j) \cong \ell\cdot j/n$. Put $\mathcal{C}_j = (c_{i,h}^{(j)})$, $\mathcal{U}_j = (u_{i,h}^{(j)})$. then by $\mathcal{C}_j = \mathcal{A}_j\cdot\mathcal{U}_j$ and the special shape of $\mathcal{A}_j$ we have $c_{i,h}^{(j)} = u_{i,h}^{(j)}$ for $i = 1, \ldots, n-1$ and $h = 1, \ldots, n$, and

$$u_{n,h}^{(j)} = \left( - c_{1,h}^{(j)}\cdot\theta_1^{(j)} - \ldots - c_{n-1,h}^{(j)}\cdot\theta_{n-1}^{(j)} + c_{n,h}^{(j)} \right)/\theta_n^{(j)} .$$

It follows that $L(\mathcal{U}_j) \cong L(\mathcal{C}_j)$. So

$$L(\mathcal{B}_j) \cong \max \left( L(\mathcal{E}\cdot\mathcal{C}_{j-1}),\ L(\mathcal{D}_{j-1}\cdot\mathcal{U}_{j-1}) \right) \cong \ell + \ell\cdot(j-1)/n .$$

Instead of applying the $L^3$-algorithm once with $\mathcal{A}$ as input, we apply it $k$ times, with $\mathcal{B}_1, \ldots, \mathcal{B}_k$ as input. Thus we reduce the computation time by a factor

$$\frac{L(\mathcal{A})^3}{\sum_{j=1}^{k} L(\mathcal{B}_j)^3} \cong \frac{(\ell\cdot k)^3}{\sum_{j=1}^{k} \ell^3\cdot\left(1+\frac{j-1}{n}\right)^3} = \frac{k^3\cdot n^3}{\sum_{j=0}^{k-1} (n+j)^3} .$$

For $k$ between $2.5\cdot n$ and $3\cdot n$ this expression is maximal, about $0.4\cdot n^2$. So the reduction in computation time is considerable (a factor 10 already for $n = 5$ ). The storage space that is required is also reduced, since the largest numbers that appear in the input have $\ell\cdot\left(1+(k-1)/n\right)$ instead of $\ell\cdot k$ digits.

### 3.6. Finding all short lattice points: the Fincke and Pohst algorithm.

Sometimes it is not sufficient to have a lower bound for $\ell(\Gamma)$ or $\ell(\Gamma,\underline{y})$ only. It may be useful to know exactly all vectors $\underline{x} \in \Gamma$ such that $|\underline{x}| \leq C$ or $|\underline{x}-\underline{y}| \leq C$ for a given constant $C$. There exists an efficient algorithm for finding all solutions to these problems. This algorithm was devised by Fincke and Pohst [1985], cf. their (2.8) and (2.12). We give a description of this algorithm below.

The input of the algorithm is a matrix $\mathcal{B}$, whose column vectors span the lattice $\Gamma$, and a constant $C > 0$. The output is a list of all lattice points $\underline{x} \in \Gamma$ with $|\underline{x}| \leq C$, apart from $\underline{x} = \underline{0}$. We give the algorithm in Figure 2. We use the notation $\mathcal{X} = (x_{ij})$ for matrices $\mathcal{X} = \mathcal{A}, \mathcal{B}, \mathcal{R}, \mathcal{S}, \mathcal{U}$, and $\underline{x}_i$ for the column vectors of $\mathcal{X}$.

The algorithm can also be used for finding all vectors $\underline{x} \in \Gamma$ of which the

Figure 2. The Fincke and Pohst Algorithm.

```
        $\mathcal{A} := \mathcal{B}^T \cdot \mathcal{B}$ ;
        $q_{ij} := a_{ij}$  for  $1 \le i \le j \le n$ ;
        $q_{ji} := q_{ij}$ ,  $q_{ij} := q_{ij}/q_{ii}$  for  $1 \le i < j \le n$ ;
        $q_{k\ell} := q_{k\ell} - q_{ki} \cdot q_{i\ell}$  for  $i+1 \le k \le \ell \le n$  for  $1 \le i \le n$ ;
        $r_{ii} := \surd q_{ii}$  for  $1 \le i \le n$ ;
        $r_{ij} := r_{ii} \cdot q_{ij}$ ,  $r_{ji} := 0$  for  $1 \le j < i \le n$ ;
        compute  $\mathcal{R}^{-1}$ ;
        compute a row-reduced version  $\mathcal{S}^{-1}$  of  $\mathcal{R}^{-1}$ , and  $\mathcal{U}$, $\mathcal{U}^{-1}$  such
             that  $\mathcal{S}^{-1} = \mathcal{U}^{-1} \cdot \mathcal{R}^{-1}$ ;
        compute  $\mathcal{S} = \mathcal{R} \cdot \mathcal{U}$ ;
        determine a permutation  $\pi$  such that  $|\underline{s}_{\pi(1)}| \ge \ldots \ge |\underline{s}_{\pi(n)}|$ ,
             let  $\mathcal{S}'$  be the matrix with columns  $\underline{s}_{\pi^{-1}(i)}$  for $i = 1,\ldots,n$ ;
        $\mathcal{A} := \mathcal{S}'^T \cdot \mathcal{S}'$ ;
        $q_{ij} := a_{ij}$  for  $1 \le i \le j \le n$ ;
        $q_{ji} := q_{ij}$ ,  $q_{ij} := q_{ij}/q_{ii}$  for  $1 \le i < j \le n$ ;
        $q_{k\ell} := q_{k\ell} - q_{ki} \cdot q_{i\ell}$  for  $i+1 \le k \le \ell \le n$  for  $1 \le i \le n$ ;
        $i := n$ ;
        $T_i := C$ ;
        $U_i := 0$ ;
  (1)   $Z := \surd(T_i/q_{ii})$ ;
        $UB(x_i) := \lfloor Z-U_i \rfloor$ ;
        $x_i := \lceil -Z-U_i \rceil - 1$ ;
  (2)   $x_i := x_i + 1$ ;
        if  $x_i \le UB(x_i)$ , go to (4) ;
  (3)   $i := i + 1$ ;
        go to (2) ;
  (4)   if $i = 1$ , go to (5) ;
        $i := i - 1$ ;
        $U_i := \sum_{j=i+1}^{m} q_{ij} \cdot x_j$ ;
        $T_i := T_{i+1} - q_{i+1,i+1} \cdot (x_{i+1} + U_{i+1})^2$ ;
        go to (1) ;
  (5)   if  $x_i = 0$  for  $1 \le i \le n$ , terminate ;
        compute and print  $\underline{x} = \mathcal{U} \cdot (x_{\pi^{-1}(1)}, \ldots, x_{\pi^{-1}(n)})^T$ ;
        go to (2).
```

distance to a given non-lattice point $\underline{y}$ is at most a given constant $C$. Namely, let

$$\underline{y} = \sum_{i=1}^{n} s_i \cdot \underline{b}_i ,$$

and let $r_i$ be the integer nearest to $s_i$ for all $i$. Put

$$\underline{z} = \sum_{i=1}^{n} r_i \cdot \underline{b}_i .$$

Then $|\underline{y}-\underline{z}| < C'$ for some constant $C'$ ( $C' = \frac{n}{2}\cdot\sum|\underline{b}_i|$ will do). Since $\underline{z} \in \Gamma$ it suffices to search for all lattice points $\underline{u}$ with $|\underline{u}| \leq C + C'$, and compute for each such $\underline{u}$ also $\underline{x} = \underline{z} + \underline{u}$, since $|\underline{x}-\underline{y}| < C$ implies

$$|\underline{u}| \leq |\underline{x}-\underline{y}| + |\underline{y}-\underline{z}| \leq C + C' .$$

## 3.7. Homogeneous multi-dimensional approximation in the real case: real approximation lattices.

Let the linear form $\Lambda$ have the form

$$\Lambda = \sum_{i=1}^{n} x_i \cdot \vartheta_i .$$

We assume that $n \geq 2$. The case $n = 2$ has already been discussed in Section 3.2, but the method of this section works also for $n = 2$. In fact, it is in this case essentially the same method.

Let $C$ be a large enough integer, that is of the order of magnitude of $X_0^n$. Let $\gamma \in \mathbb{N}$ be a constant (we will explain its use later). We define the *approximation lattice* $\Gamma$ by giving the matrix

$$\mathcal{B} = \begin{pmatrix} \gamma & & & \\ & \ddots & & \varnothing \\ \varnothing & & \gamma & \\ [\gamma\cdot C\cdot\vartheta_1] & \ldots & [\gamma\cdot C\cdot\vartheta_{n-1}] & [\gamma\cdot C\cdot\vartheta_n] \end{pmatrix} ,$$

of which the column vectors $\underline{b}_1, \ldots, \underline{b}_n$ are a basis of the lattice. Then $\Gamma$ is a sublattice of $\mathbb{Z}^n$ of determinant $\gamma^{n-1}\cdot[\gamma\cdot C\cdot\vartheta_n]$, which is of size $C$. A lattice point $\underline{x}$ has the form

$$\underline{x} = \sum_{i=1}^{n} x_i \cdot \underline{b}_i = \left( \gamma \cdot x_1 , \ \ldots , \ \gamma \cdot x_{n-1} , \ \tilde{\Lambda} \right)^T ,$$

where the $x_i$ are integers, and

$$\tilde{\Lambda} = \sum_{i=1}^{n} x_i \cdot [\gamma \cdot C \cdot \vartheta_i] \ .$$

Clearly, $\tilde{\Lambda}$ is close to $\gamma \cdot C \cdot \Lambda$ . The length of the vector $\underline{x}$ now measures both $X_0$ and $|\Lambda|$ , which are exactly the two numbers we want to balance with each other. We express this in the following lemma.

<u>LEMMA 3.7.</u> *Let* $X_1$ *be a positive number such that*

$$\ell(\Gamma) \geq \sqrt{\left((n+1)^2 + (n-1) \cdot \gamma^2\right)} \cdot X_1 \ . \tag{3.21}$$

*Then* (3.1) *has no solutions with*

$$\frac{1}{\delta} \cdot \log(\gamma \cdot C \cdot c / X_1) \leq X \leq X_1 \ . \tag{3.22}$$

<u>Remark.</u> We apply this lemma for $X_1 = X_0$ . If condition (3.21) then fails, we must take a larger constant $C$ . If it holds for a constant $C$ of the size $X_0^n$ , then (3.22) yields a reduced lower bound for $X$ of size $\log X_0$ .

<u>Proof.</u> Let $x_1 , \ \ldots , \ x_n$ be a solution of (3.1) with $0 < X \leq X_1$ . Consider the lattice point

$$\underline{x} = \sum_{i=1}^{n} x_i \cdot \underline{b}_i = \left( \gamma \cdot x_1 , \ \ldots , \ \gamma \cdot x_{n-1} , \ \tilde{\Lambda} \right)^T ,$$

with $\tilde{\Lambda}$ as above. Then

$$|\underline{x}|^2 = \gamma^2 \cdot \sum_{i=1}^{n-1} x_i^2 + \tilde{\Lambda}^2 \leq (n-1) \cdot \gamma^2 \cdot X_1^2 + \tilde{\Lambda}^2 ,$$

and

$$|\tilde{\Lambda} - \gamma \cdot C \cdot \Lambda| \leq \sum_{i=1}^{n} |x_i| \cdot |[\gamma \cdot C \cdot \vartheta_i] - \gamma \cdot C \cdot \vartheta_i| \leq \sum_{i=1}^{n} |x_i| \ , \tag{3.23}$$

which is at most $n \cdot X_1$ . By (3.1), (3.21) and the definition of $\ell(\Gamma)$ we have

$$\gamma \cdot C \cdot c \cdot \exp(-\delta \cdot X) > |\gamma \cdot C \cdot \Lambda| \geq |\tilde{\Lambda}| - |\tilde{\Lambda} - \gamma \cdot C \cdot \Lambda|$$

$$\geq \sqrt{\left(\ell(\Gamma)^2-(n-1)\cdot\gamma^2\cdot X_1^2\right)} - n\cdot X_1 \geq X_1 ,$$

and (3.22) follows at once. □

Condition (3.21) can be checked by computing a reduced basis of the lattice $\Gamma$ by the $L^3$-algorithm, and applying Lemma 3.4. The parameter $\gamma$ is used to keep the "rounding-off error"

$$|[\gamma\cdot C\cdot\vartheta_i]-\gamma\cdot C\cdot\vartheta_i|$$

relatively small. This is of importance only if $C$ is not very large, usually only if one wants to make a further reduction step after the first step has already been made. For large $C$, simply take $\gamma = 1$.

It may be necessary, if $C$ is not very large, to use a more refined method of reducing the upper bound. To do so, we use the following lemma, which is a slight refinement of Lemma 3.7, together with the algorithm of Fincke and Pohst (cf. Section 3.6). It is particularly useful in the situation that one has different upper bounds for the $|x_i|$ for different $i$.

<u>LEMMA 3.8.</u> *Suppose that for a solution of* (3.1)

$$|\tilde{\Lambda}| > \sum_{i=1}^{n} |x_i| \tag{3.24}$$

*holds. Then*

$$X < \frac{1}{\delta}\cdot\log\left(\gamma\cdot C\cdot c/\left(|\tilde{\Lambda}|-\sum_{i=1}^{n}|x_i|\right)\right) . \tag{3.25}$$

<u>Proof.</u> Define the lattice point $\underline{x}$ as in the proof of Lemma 3.7. By (3.23) and (3.24)

$$|\Lambda| \geq \left(|\tilde{\Lambda}|-\sum_{i=1}^{n}|x_i|\right)/\gamma\cdot C > 0 .$$

The result follows at once by (3.1). □

We proceed as follows. Choose a constant $C_0$ such that if $|\tilde{\Lambda}| > C_0$ then the upper bounds for $|x_i|$ imply (3.24). In that case we have a new upper bound for $X$ from (3.25). In case $|\tilde{\Lambda}| \leq C_0$ we have an upper bound for the length of the vector $\underline{x}$. We compute all lattice points satisfying this bound by the algorithm of Fincke and Pohst, and check them for (3.1).

Summarizing, the reduction method presented above is based on the fact that a large solution of (3.1) corresponds to an extremely short vector in an appropriate approximation lattice. Since we can actually prove by computations that such short vectors do not exist, it follows that such large solutions do not exist. We will apply the above described techniques in Chapter 5.

## 3.8. Inhomogeneous multi-dimensional approximation in the real case: an alternative for the generalized Davenport lemma.

Let $\Lambda$ be the most general linear form that we study, viz.

$$\Lambda = \beta + \sum_{i=1}^{n} x_i \cdot \vartheta_i ,$$

where $n \geq 2$ (the case $n = 2$ has been dealt with in Section 3.3, but can be incorporated here also). To deal with this inhomogeneous case, two methods are available. The first method is a generalization of the method of Davenport that we discussed in Section 3.3. The second method is closer to the homogeneous case of the previous section.

First we explain briefly the generalized Davenport method. See Ellison [1971$^a$] (where only the case $n = 3$ is treated). Put

$$\vartheta'_i = \vartheta_i/\vartheta_n \quad \text{for} \quad i = 1, \ldots, n-1 , \quad \beta' = \beta/\vartheta_n ,$$

$$\Lambda' = \Lambda/\vartheta_n = \beta' + \sum_{i=1}^{n-1} x_i \cdot \vartheta'_i + x_n .$$

Let $(p_1, \ldots, p_{n-1}, q)$ be a simultaneous approximation to $\vartheta'_1, \ldots, \vartheta'_{n-1}$ with $q$ of the size of $X_0^{n-1}$, such that, for $i = 1, \ldots, n-1$,

$$|\vartheta'_i - p_i/q| < c'/q^{1+1/(n-1)}$$

for a small constant $c'$.

<u>LEMMA 3.9. (Davenport, Ellison).</u> *Suppose that*

$$\|q \cdot \beta'\| > 2 \cdot (n-1) \cdot X_0 \cdot c'/q^{1/(n-1)} .$$

*Then the solutions of* (3.1), (3.2) *satisfy*

$$X < \frac{1}{\delta}\cdot\log\left(q^{1+1/(n-1)}\cdot c/|\vartheta_n|\cdot c'\cdot(n-1)\cdot X_0\right) .$$

Proof. The result follows at once from

$$\|q\cdot\beta'\| \le |q\cdot\Lambda' + \sum_{i=1}^{n-1} x_i\cdot(p_i - q\cdot\vartheta_i')| \le$$

$$q\cdot|\vartheta_n|^{-1}\cdot c\cdot\exp(-\delta\cdot X) + (n-1)\cdot X_0\cdot c'/q^{1/(n-1)} . \qquad \square$$

To apply this generalized Davenport method in practice, it is necessary to compute the simultaneous approximations $(p_1,\dots,p_{n-1},q)$ . We indicated in Section 1.4 how this can be done with the $L^3$-algorithm. As lattice we take the one associated to the following matrix:

$$\begin{pmatrix} 1 & & & \\ [C\cdot\vartheta_1'] & -C & & \emptyset \\ \vdots & & \ddots & \\ [C\cdot\vartheta_{n-1}'] & \emptyset & & -C \end{pmatrix} ,$$

where $C$ is a constant of size $X_0^n$ . Then $\underline{c}_1$ , the first basis vector of a reduced basis, will have length of the size of $C^{(n-1)/n} \cong X_0^{n-1}$ . But $\underline{c}_1$ can be written as

$$\underline{c}_1 = \left( q,\ q\cdot[C\cdot\vartheta_1'] - C\cdot p_1,\ \dots,\ q\cdot[C\cdot\vartheta_{n-1}'] - C.p_{n-1} \right)^T$$

for some $p_1, \dots, p_{n-1}, q$ . It can be expected that $q$ is of size $X_0^{n-1}$ , and

$$q\cdot C\cdot|\vartheta_i' - p_i/q| \cong |q\cdot[C\cdot\vartheta_i'] - C\cdot p_i|$$

are of the size $X_0^{n-1}$ , so that $|\vartheta_i' - p_i/q|$ are of the size $X_0^{n-1}/C\cdot X_0^{n-1} = C^{-1} \cong X_0^{-n} \cong q^{-(1+1/(n-1))}$ , as desired.

The above method has been applied in practice to solve Thue and Thue-Mahler equations by Agrawal, Coates, Hunt and van der Poorten [1980] (using multi-dimensional continued fractions instead of the $L^3$-algorithm), Pethö and Schulenberg [1987], and Blass, Glass, Meronk and Steiner [1987$^a$], [1987$^b$]. So it has proved to be useful. However, we prefer another method, for several reasons. Firstly, it is close to the homogeneous case as described in the previous section, whereas the generalized Davenport method has no obvious counterpart for the homogeneous case. Secondly, it actually produces solutions for which the linear form $\Lambda$ is almost as near to zero as possible under the condition $X \leq X_0$. Thirdly, an analogous method for the p-adic case can be given (see Section 3.11). Finally, if a linear relation between the $\vartheta_i$ exists, but had not been noticed before (a situation that sometimes occurs when one solves e.g. Thue equations), the method detects these relations, by finding explicitly an extremely short lattice vector giving the coefficients of the relation. Concerning computation time we think that the two methods are about equally fast.

The method works as follows. We take the approximation lattice $\Gamma$ exactly as in the homogeneous case, cf. the previous section, with constants $\gamma$, $C$ chosen properly, i.e. $C$ is of the size $X_0^n$. Compute with the $L^3$-algorithm a reduced basis $\underline{c}_1, \ldots, \underline{c}_n$ of $\Gamma$. Let $\mathcal{C}$ be the matrix associated to this basis, and compute also the transformation matrix $\mathcal{U}$ with $\mathcal{C} = \mathcal{B}\cdot\mathcal{U}$, and its inverse $\mathcal{U}^{-1}$. Note that $\mathcal{B}^{-1}$, and hence also $\mathcal{C}^{-1}$, are easy to compute, namely by

$$\mathcal{B}^{-1} = \begin{pmatrix} 1/\gamma & & & \emptyset \\ & \ddots & & \\ \emptyset & & 1/\gamma & \\ -\dfrac{[\gamma\cdot C\cdot\vartheta_1]}{\gamma\cdot[\gamma\cdot C\cdot\vartheta_n]} & \cdots & -\dfrac{[\gamma\cdot C\cdot\vartheta_{n-1}]}{\gamma\cdot[\gamma\cdot C\cdot\vartheta_n]} & \dfrac{1}{[\gamma\cdot C\cdot\vartheta_n]} \end{pmatrix}$$

and the $L^3$-algorithm. Let $\underline{y} \in \mathbb{Z}^n$ be defined by

$$\underline{y} = \left(0, \ldots, 0, -[\gamma\cdot C\cdot\beta]\right)^T = \sum_{i=1}^{n} s_i\cdot\underline{c}_i ,$$

where the coefficients $s_i \in \mathbb{R}$ can be computed by

$$\left(s_1, \ldots, s_n\right)^T = \mathcal{C}^{-1}\cdot\underline{y} .$$

To be more precise, if $\mathcal{U}^{-1}$ has $\underline{u}$ as n th column, then $\mathcal{C}^{-1}$ has $\underline{u}/[\gamma\cdot C\cdot\vartheta_n]$ as n th column, so

$$\left(s_1, \ldots, s_n\right)^T = -\underline{u}\cdot[\gamma\cdot C\cdot\beta]/[\gamma\cdot C\cdot\vartheta_n] .$$

Now we apply Lemma 3.5 or 3.6, that provide a lower bound for $\ell(\Gamma,\underline{y})$. Then we can apply the following lemma.

<u>LEMMA 3.10.</u> *Let* $X_1$ *be a positive constant such that*

$$\ell(\Gamma,\underline{y}) \geq \sqrt{\left((n+2)^2+(n-1)\gamma^2\right)}\cdot X_1 \; . \tag{3.26}$$

*Then* (3.1) *has no solutions with*

$$\frac{1}{\delta}\cdot\log(\gamma\cdot C\cdot c/X_1) \leq X \leq X_1 \; . \tag{3.27}$$

<u>Remark.</u> We apply this lemma for $X_1 = X_0$ . If condition (3.26) then fails, we must take a larger constant $C$ . If it holds for a constant $C$ of the size $X_0^n$ , then (3.27) yields a reduced lower bound for $X$ of size $\log X_0$ .

<u>Proof.</u> Let $x_1, \ldots, x_n$ be a solution of (3.1) with $0 < X \leq X_1$ . Consider the lattice point

$$\underline{x} = \sum_{i=1}^{n} x_i\cdot\underline{b}_i = \left(\gamma\cdot x_1, \;\ldots,\; \gamma\cdot x_{n-1},\; \tilde{\Lambda}_0\right)^T ,$$

with

$$\tilde{\Lambda}_0 = \sum_{i=1}^{n} x_i\cdot[\gamma\cdot C\cdot\vartheta_i] \; .$$

Put $\tilde{\Lambda} = [\gamma\cdot C\cdot\beta] + \tilde{\Lambda}_0$ . Then

$$|\underline{x}-\underline{y}|^2 = \gamma^2\cdot\sum_{i=1}^{n-1} x_i^2 + \tilde{\Lambda}^2 \leq (n-1)\cdot\gamma^2\cdot X_1^2 + \tilde{\Lambda}^2 ,$$

and

$$\begin{aligned} |\tilde{\Lambda}-\gamma\cdot C\cdot\Lambda| &\leq |[\gamma\cdot C\cdot\beta]-\gamma\cdot C\cdot\beta| + \sum_{i=1}^{n} |x_i|\cdot|[\gamma\cdot C\cdot\vartheta_i]-\gamma\cdot C\cdot\vartheta_i| \\ &\leq 1 + \sum_{i=1}^{n} |x_i| \leq 1 + n\cdot X_1 \leq (n+1)\cdot X_1 \; . \end{aligned}$$

By (3.1), (3.26) and the definition of $\ell(\Gamma,\underline{y})$ the result follows, since

$$\begin{aligned} \gamma\cdot C\cdot c\cdot\exp(-\delta\cdot X) &> |\gamma\cdot C\cdot\Lambda| \geq |\tilde{\Lambda}| - |\tilde{\Lambda}-\gamma\cdot C\cdot\Lambda| \\ &\geq \sqrt{\left(\ell(\Gamma,\underline{y})^2-(n-1)\cdot\gamma^2\cdot X_1^2\right)} - (n+1)\cdot X_1 \geq X_1 \; . \end{aligned}$$

□

Again we may prove refinements of the above lemma, similar to Lemma 3.8 in the homogeneous case. We explained in Section 3.5. how to apply the Fincke and Pohst algorithm in the inhomogeneous case. We do not work that out here.

Summarizing, the method described above is based on the fact that a large solution of (3.1) in the inhomogeneous case leads to a lattice point extremely near to a fixed point in $\mathbb{Z}^n$. We can actually prove by some computations that such lattice points do not exist, so that such extreme solutions do not exist. The method outlined in this section is used in Chapter 8. Note that in the case $n = 2$ the method is essentially the same as the Davenport lemma.

### 3.9. Inhomogeneous zero-dimensional approximation in the p-adic case.

In the p-adic case we start with a very simple linear form $\Lambda$, to which also a very simple reduction method applies. Let $\Lambda$ be

$$\Lambda = \beta + x\cdot\vartheta ,$$

for $\beta, \vartheta \in \Omega_p$ such that $\beta/\vartheta \in \mathbb{Q}_p$, and $x \in \mathbb{Z}$, $x > 0$. It is obvious that in the real case with such a simple linear form $\Lambda$ inequality (3.1) has only finitely many solutions (we even don't need (3.2)), and that they are easy to compute. In the p-adic case however, inequality (3.3) may have infinitely many solutions, so we do need a bound like (3.4), and a reduction method.

Put $\vartheta' = -\beta/\vartheta$. Then $\vartheta' \in \mathbb{Q}_p$. Inequality (3.3) now becomes

$$\operatorname{ord}_p(\vartheta'-x) \geq c_1' + c_2\cdot x , \tag{3.28}$$

where $c_1'$, $c_2$ are constants with $c_2 > 0$. We assume that

$$x \geq -c_1'/c_2 .$$

Then (3.28) has no solutions if $\operatorname{ord}_p(\vartheta') < 0$. Hence we may assume that $\vartheta'$ is a p-adic integer. Let the p-adic expansion of $\vartheta'$ be

$$\vartheta' = \sum_{i=0}^{\infty} u_i\cdot p^i ,$$

where $u_i \in \{ 0, 1, \dots, p-1 \}$ for all $i \in \mathbb{N}_0$. Compute the p-adic digits $u_i$ far enough to be able to apply the following reduction lemma.

LEMMA 3.11. *Let* $X_1$ *be a positive constant. Let* $r$ *be the minimal index such that*

$$p^r > X_1 \ , \quad u_r \neq 0 \ . \tag{3.29}$$

*Then* (3.28) *has no solutions with*

$$(r-c_1')/c_2 < x \leq X_1 \ . \tag{3.30}$$

Remark. We apply the lemma with $X_1 = X_0$ . The assumption behind the lemma is that in the p-adic expansion of $\vartheta'$ no long sequences of zeroes appear. In fact, it seems that in our applications the numbers $u_i$ are distributed randomly over $\{ 0, 1, \ldots, p-1 \}$ . Then the minimal $r$ satisfying (3.29) will not be much larger than $\log X_0/\log p$ , and then (3.30) yields a reduced upper bound of size $\log X_0$ , as desired.

Proof. Let $x \leq X_1$ satisfy (3.28). Suppose that $\text{ord}_p(\vartheta'-x) \geq r+1$ . Then

$$x \equiv \sum_{i=0}^{r} u_i \cdot p^i \pmod{p^{r+1}} \ .$$

By $x \geq 0$ it follows from (3.29) that

$$x \geq \sum_{i=0}^{r} u_i \cdot p^i \geq u_r \cdot p^r \geq p^r > X_1 \ ,$$

which contradicts the assumption $x \leq X_1$ . Hence $\text{ord}_p(\vartheta'-x) \leq r$ , and (3.30) follows from (3.28). □

Remark. In the above proof it is essential that $x \geq 0$ . It is however not difficult to formulate a similar result that holds for all $x \in \mathbb{Z}$ , by looking, if $p \neq 2$ for p-adic digits $u_i$ that are not only $\neq 0$ but also $\neq p-1$ , and if $p = 2$ for p-adic digits $u_i$ with $u_i \neq u_{i+1}$ .

A method very similar to the one described above was used by Wagstaff [1979], [1981] for solving congruences such as $5^n \equiv 2 \pmod{3^n}$ . We apply the method in Chapter 4.

### 3.10. Homogeneous one-dimensional approximation in the p-adic case: p-adic continued fractions and approximation lattices of p-adic numbers.

Let $\Lambda$ have the form

$$\Lambda = x_1 \cdot \vartheta_1 + x_2 \cdot \vartheta_2 ,$$

where $\vartheta_1, \vartheta_2 \in \Omega_p$ such that $\vartheta = -\vartheta_1/\vartheta_2 \in \mathbb{Q}_p$, and $x_1, x_2 \in \mathbb{Z}$. We may assume that $\operatorname{ord}_p(\vartheta) \geq 0$. Now

$$\Lambda' = \Lambda/\vartheta_1 = - x_1 \cdot \vartheta + x_2 .$$

So (3.3) now means that the rational number $x_2/x_1$ is p-adically close to the p-adic number $\vartheta$.

In analogy of the real case it seems reasonable to study p-adic continued fraction algorithms. However, a p-adic continued fraction algorithm that provides all best approximations to a p-adic number seems not to exist. Therefore we introduce the concept of p-*adic approximation lattices*, as was done in de Weger [1986$^a$]. From this paper we adopt the best approximation algorithm, which is a generalization of the algorithm of Mahler [1961], Chapter IV. This algorithm goes back also on the euclidean algorithm, and thus is close to a continued fraction algorithm. But it is not a p-adic continued fraction algorithm in the sense that a p-adic number is expanded into a continued fraction, and that the approximations are then found by truncating the continued fraction.

Recall that for $\mu \in \mathbb{N}_0$ the rational integer $\vartheta^{(\mu)}$ is defined by $\operatorname{ord}_p(\vartheta-\vartheta^{(\mu)}) \geq \mu$ and $0 \leq \vartheta^{(\mu)} < p^\mu$. We define for any $\mu \in \mathbb{N}_0$ the p-adic approximation lattice $\Gamma_\mu$ by a matrix to which a basis of $\Gamma_\mu$ is associated, namely the matrix

$$\begin{bmatrix} 1 & 0 \\ \vartheta^{(\mu)} & p^\mu \end{bmatrix} .$$

Then it is easy to see that

$$\Gamma_\mu = \{ \left(x_1, x_2\right)^T \in \mathbb{Z}^2 \mid \operatorname{ord}_p(x_2 - x_1 \cdot \vartheta) \geq \mu \}$$

(cf. Lemma 3.13 in the next section, where we prove a more general result). The following algorithm computes the point of minimal length in $\Gamma_\mu$.

<u>Figure 3.</u> p-adic approximation algorithm.

```
    $\underline{x} := \left(1, \vartheta^{(\mu)}\right)^T$ ; $\underline{y} := \left(0, p^{\mu}\right)^T$ ;
    if $|\underline{x}| > |\underline{y}|$ , interchange $\underline{x}$ and $\underline{y}$ ;
(1) compute $K \in \mathbb{Z}$ such that $|\underline{y} - K \cdot \underline{x}|$ is minimal ;
    $\underline{y} := \underline{y} - K \cdot \underline{x}$ ;
    if $|\underline{x}| > |\underline{y}|$ , interchange $\underline{x}$ and $\underline{y}$ , and go to (1) ;
    print $\underline{x}$ .
```

With this algorithm it is possible to compute $\ell(\Gamma_{\mu})$ explicitly. Then we can apply the following lemma.

<u>LEMMA 3.12.</u> *Let* $X_1$ *be a constant such that*

$$\ell(\Gamma_{\mu}) > \sqrt{2} \cdot X_1 \ . \tag{3.31}$$

*Then* (3.3) *has no solutions with*

$$\left(\mu - 1 - c_1 + \mathrm{ord}_p(\vartheta_2)\right)/c_2 < x_j \leq X \leq X_1 \ . \tag{3.32}$$

<u>Remark.</u> We take $\mu$ such that $p^{\mu}$ is of the size of $X_0^2$ , and apply the lemma for $X_1 = X_0$ . Then we expect that $\ell(\Gamma_{\mu})$ is of the size of $X_0$ , so that (3.31) is a reasonable condition.

<u>Proof.</u> Apply the proof of Lemma 3.14 (in the next section) for $n = 2$ . □

The above method has been applied by Agrawal, Coates, Hunt and van der Poorten [1980]. We use it in Chapters 6 and 7.

### 3.11. Homogeneous multi-dimensional approximation in the p-adic case: p-adic approximation lattices.

We now study the case

$$\Lambda = \sum_{i=1}^{n} x_i \cdot \vartheta_i \ ,$$

where $\vartheta_i \in \Omega_p$ such that $\vartheta_i/\vartheta_j \in \mathbb{Q}_p$ , $x_i \in \mathbb{Z}$ for all $i, j$ , and with $n \geq 2$ . We may assume that $\mathrm{ord}_p(\vartheta_i)$ is minimal for $i = n$ . Put

$$\vartheta'_i = -\vartheta_i/\vartheta_n \quad \text{for} \quad i = 1, \ldots, n-1 .$$

Then $\vartheta'_i \in \mathbb{Z}_p$ for all $i$ . Put

$$\Lambda' = \Lambda/\vartheta_n = -\sum_{i=1}^{n-1} x_i \cdot \vartheta'_i + x_n .$$

The definition of the p-adic approximation lattices can be generalized directly from the one-dimensional case. Namely, for any $\mu \in \mathbb{N}_0$ we define $\Gamma_\mu$ as the lattice associated to the matrix

$$\mathscr{B}_\mu = \begin{pmatrix} 1 & & & \emptyset \\ & \ddots & & \\ \emptyset & & 1 & \\ \vartheta'^{(\mu)}_1 & \ldots & \vartheta'^{(\mu)}_{n-1} & p^\mu \end{pmatrix} .$$

Then we have the following result.

<u>LEMMA 3.13.</u> *The lattice* $\Gamma_\mu$ , *associated to the above defined matrix* $\mathscr{B}_\mu$ , *is equal to the set*

$$\Gamma_\mu = \{ (x_1, \ldots, x_n)^T \in \mathbb{Z}^n \mid \mathrm{ord}_p(\Lambda') \geq \mu \} .$$

<u>Proof.</u> For any $\underline{x} = (x_1, \ldots, x_n)^T \in \Gamma_\mu$ there exists a $\underline{z} = (z_1, \ldots, z_n)^T \in \mathbb{Z}^n$ such that $\underline{x} = \mathscr{B}_\mu \cdot \underline{z}$ . Then $x_i = z_i$ for $i = 1, \ldots, n-1$ , and

$$x_n = \sum_{i=1}^{n-1} z_i \cdot \vartheta'^{(\mu)}_i + z_n \cdot p^\mu \equiv \sum_{i=1}^{n-1} x_i \cdot \vartheta'_i \pmod{p^\mu} .$$

Hence $\mathrm{ord}_p(\Lambda') \geq \mu$ . Conversely, for any $\underline{x} = (x_1, \ldots, x_n)^T$ such that $\mathrm{ord}_p(\Lambda') \geq \mu$ there obviously exists a $\underline{z} \in \mathbb{Z}^n$ such that $\underline{x} = \mathscr{B}_\mu \cdot \underline{z}$ . □

Using the $L^3$-algorithm we can compute a lower bound for $\ell(\Gamma_\mu)$ . Then we can apply the following lemma, which is a direct generalization of Lemma 3.12.

<u>LEMMA 3.14.</u> *Let* $X_1$ *be a constant such that*

$$\ell(\Gamma_\mu) > \sqrt{n} \cdot X_1 . \tag{3.33}$$

*Then* (3.3) *has no solutions with*

$$(\mu - 1 - c_1 + \mathrm{ord}_p(\vartheta_n))/c_2 < x_j \leq X \leq X_1 . \tag{3.34}$$

Remark. We take $\mu$ such that $p^{\mu}$ is of the size of $X_0^n$ , and apply the lemma for $X_1 = X_0$ . Then we expect that $\ell(\Gamma_{\mu})$ is of the size of $X_0$ , so that (3.33) is a reasonable condition.

Proof. Let $x_1, \ldots, x_n$ be a solution of (3.3) with $X \le X_1$ . Then (3.33) prohibits the point $(x_1,\ldots,x_n)^T$ from being a lattice point in $\Gamma_{\mu}$ . Hence, by Lemma 3.13, $\mathrm{ord}_p(\Lambda') \le \mu-1$ , and (3.34) follows from (3.3). □

We will apply the results of this section in Chapters 6 and 7.

## 3.12. Inhomogeneous one- and multi-dimensional approximation in the p-adic case.

Finally we study an inhomogeneous p-adic form

$$\Lambda = \beta + \sum_{i=1}^{n} x_i \cdot \vartheta_i ,$$

where $\beta$, $\vartheta_i \in \Omega_p$ such that $\beta/\vartheta_j$, $\vartheta_i/\vartheta_j \in \mathbb{Q}_p$ and $x_i \in \mathbb{Z}$ for all i, j , and $n \ge 2$ . We assume that $\mathrm{ord}_p(\vartheta_i)$ is minimal for $i = n$ , and that $\mathrm{ord}_p(\beta) \ge \mathrm{ord}_p(\vartheta_n)$ . Put

$$\vartheta_i' = -\vartheta_i/\vartheta_n \quad \text{for} \quad i = 1, \ldots, n-1 , \quad \beta' = \beta/\vartheta_n ,$$

$$\Lambda' = \Lambda/\vartheta_n = \beta' - \sum_{i=1}^{n-1} x_i \cdot \vartheta_i' + x_n .$$

Then $\beta'$, $\vartheta_i' \in \mathbb{Z}_p$ for all i . As p-adic approximation lattices we take the lattices $\Gamma_{\mu}$ that were defined for the homogeneous case, i.e. for any $\mu \in \mathbb{N}_0$ the lattice $\Gamma_{\mu}$ that is associated to the matrix $\mathcal{B}_{\mu}$ (see Section 3.11). Put further

$$\underline{y} = \left( 0, \ldots, 0, \beta'^{(\mu)} \right)^T = \sum_{i=1}^{n} s_i \cdot \underline{c}_i \in \mathbb{Z}^n ,$$

where $\underline{c}_1, \ldots, \underline{c}_n$ is a reduced basis of $\Gamma_{\mu}$ , and $s_i \in \mathbb{R}$ . By Lemma 3.5 or 3.6 we can compute a lower bound for $\ell(\Gamma,\underline{y})$ . This is useful in view of the following lemma.

<u>LEMMA 3.15.</u> *The set* $\Gamma_\mu(\underline{y}) = \Gamma_\mu + \underline{y}$ *is equal to the set*

$$\Gamma_\mu(\underline{y}) = \{ (x_1,\ldots,x_n)^T \in \mathbb{Z}^n \mid \mathrm{ord}_p(\Lambda') \geq \mu \} .$$

<u>Proof.</u> Let $\underline{x} = (x_1,\ldots,x_n)^T$ satisfy $\underline{x} - \underline{y} \in \Gamma_\mu$ . Note that

$$\underline{x} - \underline{y} = ( x_1, \ldots, x_{n-1}, x_n - \beta'^{(\mu)} )^T .$$

By Lemma 3.13 we have

$$\mathrm{ord}_p\Big( \sum_{i=1}^{n-1} x_i \cdot \vartheta_i - (x_n - \beta'^{(\mu)})\Big) \geq p^\mu .$$

The left hand side is just $\mathrm{ord}_p(\Lambda')$ , which proves the lemma. □

Obviously, the length of the shortest vector in $\Gamma_\mu(\underline{y})$ (a translated lattice) is equal to $\ell(\Gamma_\mu,\underline{y})$ (unless in the case $\underline{y} \in \Gamma_\mu$ ). We have the following useful lemma.

<u>LEMMA 3.16.</u> *Let* $X_1$ *be a constant such that*

$$\ell(\Gamma_\mu,\underline{y}) > \sqrt{n}\cdot X_1 . \tag{3.35}$$

*Then* (3.3) *has no solutions with*

$$(\mu - 1 - c_1 + \mathrm{ord}_p(\vartheta_n))/c_2 < x_j \leq X \leq X_1 . \tag{3.36}$$

<u>Remark.</u> We take $\mu$ such that $p^\mu$ is of the size of $X_0^n$ , and apply the lemma for $X_0 = X_1$ . Then we expect that $\ell(\Gamma_\mu,\underline{y})$ is of the size of $X_0$ , so that (3.35) is a reasonable condition.

<u>Proof.</u> Let $x_1, \ldots, x_n$ be a solution of (3.3) with $X \leq X_1$ . Then (3.35) prohibits the point $(x_1,\ldots,x_n)^T$ from being in $\Gamma_\mu(\underline{y})$ . Hence, by Lemma 3.15, $\mathrm{ord}_p(\Lambda') \leq \mu-1$ , and (3.36) follows from (3.3). □

We shall not apply the above lemma in this thesis, so we have included it here only for the sake of completeness. However, when solving Thue-Mahler equations (see Section 8.6), it will be of use.

### 3.13. Useful sublattices of p-adic approximation lattices.

In our p-adic applications of solving diophantine equations via linear forms, we always have linear forms in logarithms of algebraic numbers, i.e. in

$$\Lambda = \beta + \sum_{i=1}^{n} x_i \cdot \vartheta_i$$

the $\beta$ and $\vartheta_i$'s are p-adic logarithms of algebraic numbers, say

$$\beta = \log_p(\alpha_0) \ , \quad \vartheta_i = \log_p(\alpha_i) \quad \text{for} \quad i = 1, \ldots, n \ .$$

In Section 2.3 we have seen that for a $\xi \in \mathbb{Q}_p$ if $\operatorname{ord}_p(1 \pm \xi) > 1/(p-1)$ then $\operatorname{ord}_p(\log_p(\xi)) = \operatorname{ord}_p(1 \pm \xi)$ . In our applications we apply this to

$$\xi = \alpha_0 \cdot \prod_{i=1}^{n} \alpha_i^{x_i} \ ,$$

for which $\operatorname{ord}_p(\xi - 1)$ is large. This implies that $\operatorname{ord}_p(\log_p(\xi))$ is large too, on which we based the definition of our approximation lattices. However, the converse is not necessarily true: $\operatorname{ord}_p(\log_p(\xi))$ being large does not imply that $\operatorname{ord}_p(\xi - 1)$ is large. This is due to the fact that the p-adic logarithm is a multi-branched function. To be more precise, for any root of unity $\zeta \in \mathbb{Q}_p$ we have $\log_p(\zeta) = 0$ (cf. Section 2.3). In $\mathbb{Q}_p$ there exist only the $(p-1)$ th roots of unity if $p$ is odd, and only $\pm 1$ as roots of unity if $p = 2$ . Let $\zeta$ be a primitive $(p-1)$ th root of unity if $p$ is odd, and $\zeta = -1$ if $p = 2$ . It follows that $\operatorname{ord}_p(\log_p(\xi))$ being large implies that for some $k \in \{ 0, 1, \ldots, p-2 \}$ (or $k \in \{ 0, 1 \}$ if $p = 2$ )

$$\operatorname{ord}_p(\log_p(\xi)) = \operatorname{ord}_p(\xi - \zeta^k) \ .$$

It turns out that the set of $x_1, \ldots, x_n$ such that $\operatorname{ord}_p(\xi - 1)$ (or $\operatorname{ord}_p(\xi \pm 1)$ if one wishes) is large, is a sublattice $\Gamma_\mu^*$ (or $\Gamma_\mu^\#$ ) of $\Gamma_\mu$ . In the following lemma we shall prove this fact, and indicate how a basis of this sublattice can be found. Then we can work with this sublattice instead of $\Gamma_\mu$ itself. Of course, in Lemmas 3.12, 3.14 and 3.16 we can replace $\Gamma_\mu$ by these sublattices $\Gamma_\mu^*$, $\Gamma_m^\#$ . For simplicity we assume that $\alpha_i \in \mathbb{Q}_p$ for all $i$ . We take $\alpha_0 = 1$ , and leave it to the reader to define appropriate translated lattices $\Gamma_\mu^*(\underline{y})$, $\Gamma_\mu^\#(\underline{y})$ for the case $\alpha_0 \neq 1$ .

<u>LEMMA 3.17.</u> *Let* $\alpha_1, \ldots, \alpha_n \in \mathbb{Q}_p$ *be given numbers with* $\operatorname{ord}_p(\alpha_i) = 0$ *for all* $i$ , *and* $\operatorname{ord}_p(\log_p(\alpha_i))$ *minimal for* $i = n$ . *Let* $x_1, \ldots, x_n \in \mathbb{Z}$ . *Put*

$$\xi = \prod_{i=1}^{n} \alpha_i^{x_i} , \quad \mu_0 = \mathrm{ord}_p(\log_p(\alpha_n)) .$$

*Put for any* $\mu \in \mathbb{N}_0$

$$\Gamma_\mu = \{ (x_1,\dots,x_n) \in \mathbb{Z}^n \mid \mathrm{ord}_p(\log_p(\xi)) \ge \mu + \mu_0 \} ,$$

$$\Gamma_\mu^* = \{ (x_1,\dots,x_n) \in \mathbb{Z}^n \mid \mathrm{ord}_p(\xi \pm 1) \ge \mu + \mu_0 \} ,$$

$$\Gamma_\mu^\# = \{ (x_1,\dots,x_n) \in \mathbb{Z}^n \mid \mathrm{ord}_p(\xi - 1) \ge \mu + \mu_0 \} .$$

*Then* $\Gamma_\mu^\# \subseteq \Gamma_\mu^* \subseteq \Gamma_\mu$ *are lattices. If* $p = 2$ *they are all equal. If* $p = 3$ *then* $\Gamma_\mu^* = \Gamma_\mu$ *. Let further* $p \ge 3$ *. Let* $\underline{b}_1, \dots, \underline{b}_n$ *be a basis of* $\Gamma_\mu$ *. Define* $k(\underline{x})$ *for any* $\underline{x} = (x_1,\dots,x_n)^T \in \Gamma_\mu$ *by*

$$\xi \equiv \zeta^{k(\underline{x})} \pmod{p^{\mu+\mu_0}} , \quad k(\underline{x}) \in \{ 0, 1, \dots, p-2 \} .$$

*Let* $\underline{b}_1', \dots, \underline{b}_n'$ *be a basis of* $\Gamma_\mu$ *such that*

$$k(\underline{b}_n') = \gcd\left(k(\underline{b}_1),\dots,k(\underline{b}_n)\right) .$$

*Put for* $i = 1, \dots, n-1$ *and* $p \ge 5$

$$\gamma_i^* \equiv k(\underline{b}_i')/k(\underline{b}_n') \pmod{(p-1)/2} , \quad |\gamma_i^*| \le (p-1)/4 ,$$

$$\underline{b}_i^* = \underline{b}_i' - \gamma_i^* \cdot \underline{b}_n' ,$$

*and for* $p \ge 3$ *also*

$$\gamma_i^\# \equiv k(\underline{b}_i')/k(\underline{b}_n') \pmod{(p-1)} , \quad |\gamma_i^\#| \le (p-1)/2 ,$$

$$\underline{b}_i^\# = \underline{b}_i' - \gamma_i^\# \cdot \underline{b}_n' .$$

*Further put for* $p \ge 5$

$$\gamma_n^* = \mathrm{lcm}\left(k(\underline{b}_n'),(p-1)/2\right)/k(\underline{b}_n') , \quad \underline{b}_n^* = \gamma_n^* \cdot \underline{b}_n' ,$$

*and for* $p \ge 3$ *also*

$$\gamma_n^\# = \mathrm{lcm}\left(k(\underline{b}_n'),p-1\right)/k(\underline{b}_n') , \quad \underline{b}_n^\# = \gamma_n^\# \cdot \underline{b}_n' .$$

*Then* $\underline{b}_1^*, \dots, \underline{b}_n^*$ *is a basis of* $\Gamma_\mu^*$ *, and* $\underline{b}_1^\#, \dots, \underline{b}_n^\#$ *is a basis of* $\Gamma_\mu^\#$ *.*

Proof. It is trivial that $\Gamma_\mu^\# \subseteq \Gamma_\mu^* \subseteq \Gamma_\mu$ , and that they are lattices. The equalities of the lattices for $p = 2, 3$ follow from the fact that ±1 are the only roots of unity in $\mathbb{Q}_p$ for $p = 2, 3$ . Note that $k(\underline{x})$ is (mod $(p-1)$) a linear function on $\Gamma_\mu$ . The points $\underline{x}$ of $\Gamma^*_\mu$ are characterized by $(p-1)/2 \mid k(\underline{x})$ , and the points $\underline{x}$ of $\Gamma^\#_m$ are characterized by $(p-1) \mid k(\underline{x})$ . It follows from the definitions in the lemma that for $i = 1, \dots, n-1$

$$k(\underline{b}^*_i) \equiv k(\underline{b}'_i) - \gamma^*_i \cdot k(\underline{b}'_n) \equiv 0 \pmod{(p-1)/2} ,$$

$$k(\underline{b}^\#_i) \equiv k(\underline{b}'_i) - \gamma^\#_i \cdot k(\underline{b}'_n) \equiv 0 \pmod{(p-1)} .$$

Note that $\underline{b}^*_1, \dots, \underline{b}^*_{n-1}, \underline{b}'_n$ and $\underline{b}^\#_1, \dots, \underline{b}^\#_{n-1}, \underline{b}'_n$ are both bases of $\Gamma_\mu$ . Write $\underline{x} \in \Gamma_\mu$ as

$$\underline{x} = \sum_{i=1}^{n-1} y^*_i \cdot \underline{b}^*_i + y^*_n \cdot \underline{b}'_n = \sum_{i=1}^{n-1} y^\#_i \cdot \underline{b}^\#_i + y^\#_n \cdot \underline{b}'_n$$

for integers $y^*_i, y^\#_i$ . Then it follows that

$$k(\underline{x}) \equiv y^*_n \cdot k(\underline{b}'_n) \pmod{(p-1)/2} ,$$

$$k(\underline{x}) \equiv y^\#_n \cdot k(\underline{b}'_n) \pmod{(p-1)} .$$

So $\underline{x} \in \Gamma^*_\mu$ if and only if $\gamma^*_n \mid y^*_n$ , and $\underline{x} \in \Gamma^\#_\mu$ if and only if $\gamma^\#_n \mid y^\#_n$ . This proves the result. □