16

Short vectors in lattices

In this chapter, we present a polynomial-time algorithm for factoring univariate polynomials with integer coefficients. We will also indicate how the algorithm can be modified so as to also work for bivariate polynomials over a field where we have univariate factorization, such as $\mathbb{Q}$ or a finite field. The main technical ingredient, short vectors in lattices, will be the central topic of this chapter.

16.1. Lattices

The methods we discuss in this chapter deal with computational aspects of the *geometry of numbers*, a mathematical theory initiated by Hermann Minkowski in the 1890s. This theory produces many results about Diophantine approximation, convex bodies, embeddings of algebraic number fields in $\mathbb{C}$, and the ellipsoid method for rational linear programming.

Let $f = (f_1, \ldots, f_n) \in \mathbb{R}^n$. In this chapter, we use the **norm** (or 2-norm, or Euclidean norm) of f, given by

$$\|f\| = \|f\|_2 = \Big(\sum_{1 \le i \le n} f_i^2 \Big)^{1/2} = (f \star f)^{1/2} \in \mathbb{R},$$

where $f \star g = \sum_{1 \le i \le n} f_i g_i \in \mathbb{R}$ is the usual **inner product** of two vectors f and $g = (g_1, \ldots, g_n)$ in $\mathbb{R}^n$ (often written as (f, g), or $\langle f, g \rangle$, or $f \cdot g^T$ in the literature). The vectors f and g are **orthogonal** if $f \star g = 0$.

DEFINITION 16.1. *Let $n \in \mathbb{N}$ and $f_1, \ldots, f_n \in \mathbb{R}^n$ with $f_i = (f_{i1}, \ldots, f_{in})$. Then*

$$L = \sum_{1 \le i \le n} \mathbb{Z} f_i = \{ \sum_{1 \le i \le n} r_i f_i : r_1, \ldots, r_n \in \mathbb{Z} \}$$

*is the **lattice** or $\mathbb{Z}$-module generated by $f_1, \ldots, f_n$. If these vectors are linearly independent, they are a **basis** of L. The **norm of L** is $|L| = |\det(f_{ij})_{1 \le i,j \le n}| \in \mathbb{R}$. Lemma 16.2 below implies that it is well defined, in other words, that the norm is independent of the choice of the generators of L.*

LEMMA 16.2. *Let* $N \subseteq M \subseteq \mathbb{R}^n$ *be lattices, generated by* $g_1, \ldots, g_n$ *and* $f_1, \ldots, f_n$, *respectively, where* $f_i = (f_{i1}, \ldots, f_{in})$ *and* $g_i = (g_{i1}, \ldots, g_{in})$. *Then* $\det(f_{ij})_{1 \le i,j \le n}$ *divides* $\det(g_{ij})_{1 \le i,j \le n}$.

PROOF. For $1 \le i, j \le n$ there exist $a_{ij} \in \mathbb{Z}$ such that $g_i = \sum_{1 \le j \le n} a_{ij} f_j$. Hence $|\det(g_{ij})| = |\det(a_{ij})| \cdot |\det(f_{ij})|$, and the claim follows. □

If we let $N = M$ in the above lemma, so that $f_1, \ldots, f_n$ and $g_1, \ldots, g_n$ both generate the same lattice, we see that $|\det(f_{ij})| = |\det(g_{ij})|$. Hence the norm is indeed independent of the choice of basis of L. Geometrically, $|L|$ is the volume of the parallelepiped spanned by $f_1, \ldots, f_n$, and Hadamard's inequality (Theorem 16.6) says that $|L| \le \|f_1\| \cdots \|f_n\|$ holds.

EXAMPLE 16.3. We let $n = 2$, $f_1 = (12, 2)$, $f_2 = (13, 4)$ and $L = \mathbb{Z}f_1 + \mathbb{Z}f_2$. Figure 16.1 shows some lattice points of L near the origin of the plane $\mathbb{R}^2$. The norm of L is

$$|L| = \left| \det \begin{pmatrix} 12 & 2 \\ 13 & 4 \end{pmatrix} \right| = 22$$

and equals the area of the blue parallelogram in Figure 16.1. Another basis of L is $g_1 = (1, 2)$ and $g_2 = (11, 0)$, and g_1 is a shortest vector in L with respect to the Euclidean norm $\|\cdot\|$. ◇

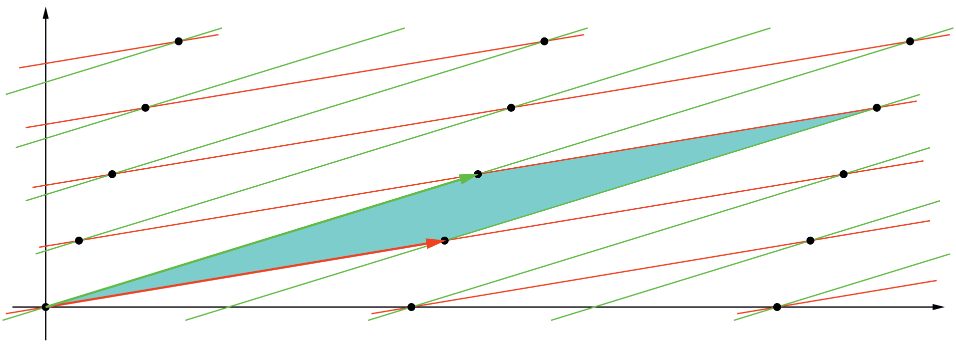

FIGURE 16.1: The lattice in $\mathbb{R}^2$ generated by $(12, 2)$ (red) and $(13, 4)$ (green).

A natural question is to compute a shortest vector in a given lattice. This problem is "$\mathcal{NP}$-hard", and there is no hope for efficient algorithms. But for our current application, the factorization of polynomials with integer coefficients, it will be sufficient to compute a "relatively short" vector, a problem for which Lenstra, Lenstra & Lovász (1982) first gave a polynomial time algorithm. Their "short vector" is guaranteed to be off by not more than a specified factor, which depends on the dimension but not the lattice itself.

16.2. Lenstra, Lenstra and Lovász' basis reduction algorithm

We briefly review the Gram-Schmidt orthogonalization procedure from linear algebra. Given an arbitrary basis $(f_1,\ldots,f_n)$ of $\mathbb{R}^n$, it computes an orthogonal basis $(f_1^*,\ldots,f_n^*)$ of $\mathbb{R}^n$ by essentially performing Gaussian elimination on the **Gramian matrix** $(f_i \star f_j)_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$ (Section 25.5). The f_i^* are defined inductively as follows.

$$f_i^* = f_i - \sum_{1\le j<i} \mu_{ij} f_j^*, \text{ where } \mu_{ij} = \frac{f_i \star f_j^*}{f_j^* \star f_j^*} = \frac{f_i \star f_j^*}{\|f_j^*\|^2} \text{ for } 1 \le j < i. \tag{1}$$

In particular, $f_1^* = f_1$. We will call $(f_1^*,\ldots,f_n^*)$ the **Gram-Schmidt orthogonal basis** of $(f_1,\ldots,f_n)$, and the f_i^* together with the μ_{ij} form the **Gram-Schmidt orthogonalization** (or GSO for short) of $f_1,\ldots,f_n$. The GSO has rational coefficients if $f_1,\ldots,f_n$ have, and then the cost for computing the GSO is $O(n^3)$ arithmetic operations in $\mathbb{Q}$.

We consider the f_i and f_i^* to be row vectors in $\mathbb{R}^n$, and define three $n\times n$ matrices F, F^*, and M in $\mathbb{R}^{n\times n}$:

$$F = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \; F^* = \begin{pmatrix} f_1^* \\ \vdots \\ f_n^* \end{pmatrix}, \; M = (\mu_{ij})_{1\le i,j\le n},$$

where $\mu_{ii} = 1$ for $i \le n$, and $\mu_{ij} = 0$ for $1 \le i < j \le n$. Then M is lower triangular with ones on the diagonal, and (1) reads:

$$F = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ \vdots & \ddots & \\ \mu_{n1} & \cdots & 1 \end{pmatrix} \begin{pmatrix} f_1^* \\ \vdots \\ f_n^* \end{pmatrix} = M \cdot F^*. \tag{2}$$

EXAMPLE 16.4. We let $n = 3$, $f_1 = (1,1,0)$, $f_2 = (1,0,1)$, $f_3 = (0,1,1)$, and calculate $f_1^* = f_1 = (1,1,0)$,

$$\mu_{21} = \frac{f_2 \star f_1^*}{f_1^* \star f_1^*} = \frac{1}{2}, \quad f_2^* = f_2 - \mu_{21} f_1^* = \left(\frac{1}{2}, -\frac{1}{2}, 1\right),$$

$$\mu_{31} = \frac{f_3 \star f_1^*}{f_1^* \star f_1^*} = \frac{1}{2}, \quad \mu_{32} = \frac{f_3 \star f_2^*}{f_2^* \star f_2^*} = \frac{1}{3}, \quad f_3^* = f_3 - \mu_{31} f_1^* - \mu_{32} f_2^* = \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right),$$

$$F = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 1 \\ -\frac{2}{3} & \frac{2}{3} & \frac{2}{3} \end{pmatrix} = M \cdot F^*.$$

We have $\|f_1\|^2 = \|f_2\|^2 = \|f_3\|^2 = 2$ and $\|f_1^*\|^2 = 2$, $\|f_2^*\|^2 = 3/2$, $\|f_3^*\|^2 = 4/3$. $\diamond$

The following theorem collects the properties of the Gram-Schmidt orthogonalization that we will need. The proof is left as Exercise 16.2.

THEOREM 16.5. *Let $f_1, \ldots, f_n \in \mathbb{R}^n$ be linearly independent, and $f_i^*, \ldots, f_n^*$ their Gram-Schmidt orthogonal basis. Let $0 \leq k \leq n$, and let $U_k = \sum_{1 \leq i \leq k} \mathbb{R} f_i \subseteq \mathbb{R}^n$ be the $\mathbb{R}$-subspace spanned by $f_1, \ldots, f_k$.*

(i) $\sum_{1 \leq i \leq k} \mathbb{R} f_i^* = U_k$.

(ii) f_k^* *is the projection of* f_k *onto the orthogonal complement*

$$U_{k-1}^{\perp} = \{f \in \mathbb{R}^n \colon f \star u = 0 \text{ for all } u \in U_{k-1}\}$$

of U_{k-1}*, and hence in particular* $\|f_k^*\| \leq \|f_k\|$.

(iii) $f_1^*, \ldots, f_n^*$ *are pairwise orthogonal, that is,* $f_i^* \star f_j^* = 0$ *if* $i \neq j$.

(iv) $\det \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \det \begin{pmatrix} f_1^* \\ \vdots \\ f_n^* \end{pmatrix}$.

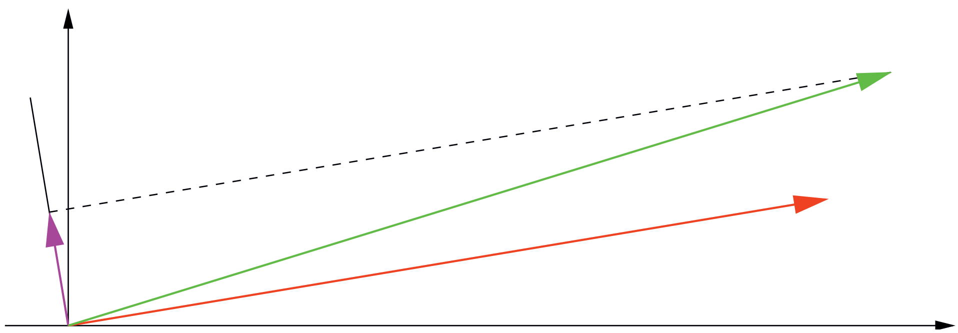

FIGURE 16.2: The Gram-Schmidt orthogonal basis of $(12,2)$ and $(13,4)$.

EXAMPLE 16.3 (continued). We have $f_1^* = f_1 = (12,2)$,

$$\mu_{21} = \frac{f_2 \star f_1^*}{f_1^* \star f_1^*} = \frac{41}{37}, \quad f_2^* = f_2 - \mu_{21} f_1^* = \Big(-\frac{11}{37}, \frac{66}{37}\Big).$$

This is illustrated in Figure 16.2: the vector f_2^* (pink) is the projection of f_2 (green) onto the orthogonal complement of f_1 (red). $\diamond$

An immediate consequence of Theorem 16.5 is the following famous inequality.

THEOREM 16.6 *Hadamard's inequality.*
Let $A \in \mathbb{R}^{n\times n}$, with row vectors $f_1, \ldots, f_n \in \mathbb{R}^{1\times n}$, and $B \in \mathbb{R}$ such that all entries of A are at most B in absolute value. Then

$$|\det A| \le \|f_1\| \cdots \|f_n\| \le n^{n/2} B^n.$$

PROOF. We may assume that A is nonsingular and the f_i are linearly independent. If $(f_1^*, \ldots, f_n^*)$ is their Gram-Schmidt orthogonal basis, then Theorem 16.5 implies that

$$\left|\det \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}\right| = \left|\det \begin{pmatrix} f_1^* \\ \vdots \\ f_n^* \end{pmatrix}\right| = \|f_1^*\| \cdots \|f_n^*\| \le \|f_1\| \cdots \|f_n\|.$$

The second inequality follows from noting that $\|f_i\| \le n^{1/2} B$ for all i. □

Of course, the theorem also holds for the column vectors of A.

The next lemma exhibits the connection between the Gram-Schmidt orthogonal basis and short vectors.

LEMMA 16.7. *Let $L \subseteq \mathbb{R}^n$ be a lattice with basis $(f_1, \ldots, f_n)$, and let $(f_1^*, \ldots, f_n^*)$ be its Gram-Schmidt orthogonal basis. Then for any $f \in L \setminus \{0\}$ we have*

$$\|f\| \ge \min\{\|f_1^*\|, \ldots, \|f_n^*\|\}.$$

PROOF. Let $f = \sum_{1\le i\le n} \lambda_i f_i \in L \setminus \{0\}$ be arbitrary, with all $\lambda_i \in \mathbb{Z}$, and let k be the highest index such that $\lambda_k \ne 0$. Substituting $\sum_{1\le j\le i} \mu_{ij} f_j^*$ for f_i yields

$$f = \sum_{1\le i\le k} \lambda_i \sum_{1\le j\le i} \mu_{ij} f_j^* = \lambda_k f_k^* + \sum_{1\le i<k} \nu_i f_i^*$$

for some appropriate $\nu_i \in \mathbb{R}$. Then

$$\begin{aligned} \|f\|^2 = f \star f &= \left(\lambda_k f_k^* + \sum_{1\le i<k} \nu_i f_i^*\right) \star \left(\lambda_k f_k^* + \sum_{1\le i<k} \nu_i f_i^*\right) \\ &= \lambda_k^2 (f_k^* \star f_k^*) + \sum_{1\le i<k} \nu_i^2 (f_i^* \star f_i^*) \ge \lambda_k^2 \cdot \|f_k^*\|^2 \\ &\ge \|f_k^*\|^2 \ge \min\{\|f_1^*\|^2, \ldots, \|f_n^*\|^2\}, \end{aligned}$$

where we used the pairwise orthogonality of the f_i^* and that $\lambda_k \in \mathbb{Z} \setminus \{0\}$. □

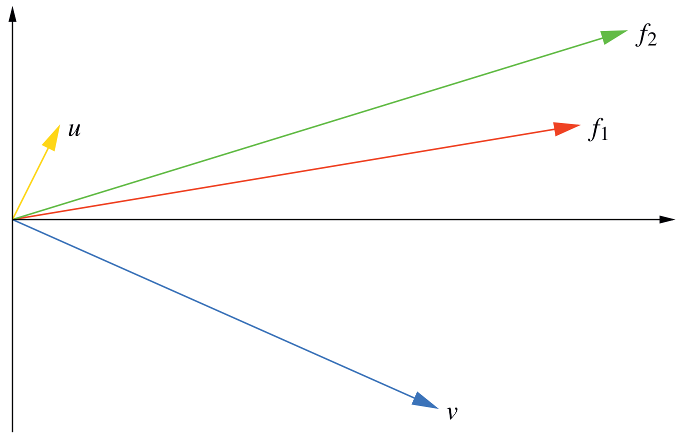

FIGURE 16.3: The vectors computed by the basis reduction algorithm 16.10 for the lattice of Example 16.3.

Our goal is to compute a short vector in L. If the Gram-Schmidt orthogonal basis of $(f_1, \ldots, f_n)$ is a basis for the lattice L generated by $f_1, \ldots, f_n$, then the lemma says that one of the f_i^* is a shortest vector. But usually the f_i^* are not even in L, as in Example 16.4. Lemma 16.7 provides a lower bound on the length of nonzero vectors in L, in terms of the GSO. With the following definition, we get a similar, though somewhat weaker, bound in terms of the original basis.

DEFINITION 16.8. *Let $f_1, \ldots, f_n \in \mathbb{R}^n$ be linearly independent and $(f_1^*, \ldots, f_n^*)$ the corresponding Gram-Schmidt orthogonal basis. Then $(f_1, \ldots, f_n)$ is* ***reduced*** *if $\|f_i^*\|^2 \leq 2\|f_{i+1}^*\|^2$ for $1 \leq i < n$.*

THEOREM 16.9.
Let $(f_1, \ldots, f_n)$ be a reduced basis of the lattice $L \subseteq \mathbb{R}^n$ and $f \in L \setminus \{0\}$. Then $\|f_1\| \leq 2^{(n-1)/2} \cdot \|f\|$.

PROOF. We have $\|f_1\|^2 = \|f_1^*\|^2 \leq 2\|f_2^*\|^2 \leq 2^2\|f_3^*\|^2 \leq \cdots \leq 2^{n-1}\|f_n^*\|^2$. Thus $\|f\| \geq \min\{\|f_1^*\|, \ldots, \|f_n^*\|\} \geq 2^{-(n-1)/2}\|f_1\|$, using Lemma 16.7. □

We now present an algorithm that computes a reduced basis of a lattice $L \subseteq \mathbb{Z}^n$ from an arbitrary basis. One can use this to find a reduced basis of a lattice in $\mathbb{Q}^n$, by multiplying with a common denominator of the given basis vectors. For $\mu \in \mathbb{R}$, we write $\lceil \mu \rfloor = \lfloor \mu + 1/2 \rfloor$ for the integer nearest to μ.

ALGORITHM 16.10 Basis reduction.
Input: Linearly independent row vectors $f_1, \ldots, f_n \in \mathbb{Z}^n$.
Output: A reduced basis $(g_1, \ldots, g_n)$ of the lattice $L = \sum_{1 \leq i \leq n} \mathbb{Z} f_i \subseteq \mathbb{Z}^n$.

1. **for** $i = 1, \ldots, n$ **do** $g_i \longleftarrow f_i$
   compute the GSO $G^*, M \in \mathbb{Q}^{n \times n}$, as in (1) and (2), $\quad i \longleftarrow 2$
2. **while** $i \leq n$ **do**
3. **for** $j = i-1, i-2, \ldots, 1$ **do**
4. $g_i \longleftarrow g_i - \lceil \mu_{ij} \rfloor g_j$, update the GSO { replacement step }
5. **if** $i > 1$ and $\|g_{i-1}^*\|^2 > 2\|g_i^*\|^2$
   then exchange g_{i-1} and g_i and update the GSO, $\quad i \longleftarrow i-1$
   else $i \longleftarrow i+1$
6. **return** $g_1, \ldots, g_n$

step	$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$	M	$\begin{pmatrix} g_1^* \\ g_2^* \end{pmatrix}$	action
4	$\begin{pmatrix} 12 & 2 \\ 13 & 4 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ \frac{41}{37} & 1 \end{pmatrix}$	$\begin{pmatrix} 12 & 2 \\ -\frac{11}{37} & \frac{66}{37} \end{pmatrix}$	row 2 $\longleftarrow$ row 2 $-$ row 1
5	$\begin{pmatrix} 12 & 2 \\ 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ \frac{4}{37} & 1 \end{pmatrix}$	$\begin{pmatrix} 12 & 2 \\ -\frac{11}{37} & \frac{66}{37} \end{pmatrix}$	exchange rows 1 and 2
4	$\begin{pmatrix} 1 & 2 \\ 12 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ \frac{16}{5} & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 2 \\ \frac{44}{5} & -\frac{22}{5} \end{pmatrix}$	row 2 $\longleftarrow$ row 2 $- 3 \cdot$ row 1
6	$\begin{pmatrix} 1 & 2 \\ 9 & -4 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ \frac{1}{5} & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 2 \\ \frac{44}{5} & -\frac{22}{5} \end{pmatrix}$	

TABLE 16.4: Trace of the basis reduction algorithm 16.10 on the lattice of Example 16.3.

In fact, Algorithm 16.10 does more than required: Lemma 16.12 (iii) below implies that $|\mu_{ij}| \leq 1/2$ holds for the GSO of the reduced basis $(g_1, \ldots, g_n)$. A reduced basis with this additional property is "almost orthogonal".

EXAMPLE 16.3 (continued). Table 16.4 traces the algorithm on the lattice of Example 16.3, and Figure 16.3 depicts the vectors g_i in the computation. We start with $g_1 = f_1 = (12, 2)$ (red) and $g_2 = f_2 = (13, 4)$ (green). In the second row of Table 16.4, g_2 is replaced by $u = g_2 - \lceil 41/37 \rfloor g_1 = (1, 2)$ (yellow). Then $g_1 = f_1$ and $g_2 = u$ are exchanged in the third row. In the last row, $v = g_2 - \lceil 16/5 \rfloor g_1 = f_1 - 3u = (9, -4)$ (blue) is computed, and the algorithm returns the reduced basis $u = (1, 2)$ and $v = (9, -4)$. We can see clearly in Figure 16.3 that the final $g_1 = u$ (the yellow vector) is much shorter than the two input vectors f_1, f_2, and that the computed basis u, v (the yellow and the blue vectors) is nearly orthogonal. $\diamond$

In the example above, the final g_1 is actually a *shortest* vector. This seems to happen quite often, but Theorem 16.9 only guarantees that the norm of the first vector in the computed basis is bigger by a factor of at most $2^{(n-1)/2}$ than the norm of a shortest vector, where n is the dimension of the lattice.

16.3. Cost estimate for basis reduction

THEOREM 16.11.
Algorithm 16.10 correctly computes a reduced basis of L and runs in polynomial time. It uses $O(n^4 \log A)$ arithmetic operations on integers whose length is $O(n \log A)$, where $A = \max_{1\leq i\leq n} \|f_i\|$.

The idea of the estimate on the number of arithmetic operations is as follows. Each execution of steps 4 or 5 has polynomial cost, and it is sufficient to bound the number of passes through step 5 with an exchange. In fact, at first glance it is not obvious that the algorithm terminates at all, since the decrease and increase of i in step 5 might continue forever. The crucial point then is to exhibit a value D with the following properties: It is always a positive integer, reasonably small in the beginning, and does not change in the algorithm except that at each exchange step it decreases (at least) by a factor of $3/4$. Therefore only few exchange steps can happen.

To structure the somewhat lengthy proof, we first investigate in the following two lemmas how the GSO of $(g_1, \ldots, g_n)$ changes in steps 4 and 5.

LEMMA 16.12. *(i) We consider one execution of step 4, and let $\lambda = \lceil \mu_{ij} \rfloor$ for short. Let G, G^*, M and H, H^*, N in $\mathbb{Q}^{n\times n}$ be the matrices of the g_k, g_k^*, μ_{kl} before and after the replacement, respectively, and $E = (e_{kl}) \in \mathbb{Z}^{n\times n}$ the matrix which has $e_{kk} = 1$ for all k, $e_{ij} = -\lambda$, and $e_{kl} = 0$ otherwise. Then*

$$H = EG, \quad N = EM, \text{ and } H^* = G^*.$$

(ii) The following invariant holds before each execution of step 4:

$$|\mu_{il}| \leq \frac{1}{2} \text{ for } j < l < i.$$

(iii) The Gram-Schmidt orthogonal basis $g_1^*, \ldots, g_n^*$ does not change in step 4, and after the loop in step 3 we have $|\mu_{il}| \leq 1/2$ for $1 \leq l < i$.

PROOF. (i) The equality $H = EG$ is just another way of saying that g_i is replaced by $g_i - \lambda g_j$ and all other g_k remain unchanged. Since $j < i$, for any $k \leq n$ the space spanned by $g_1, \ldots, g_k$ remains the same, and hence the orthogonal vectors $g_1^*, \ldots, g_n^*$