A simplified version of H.W. Lenstra's integer programing algorithm and some applications

by

A. Paz †

This research was supported by NSF Grant No. MCS-80-12716.
† Computer Sciences Department, Technion, Haifa, Israel.

Introduction

A very interesting algorithm has been recently suggested by H.W. Lenstra, Jr. [1] for solving integer programming problems. One part of that algorithm was further improved in [2]. The algorithm was shown to be polynomial in the length of the input, for a fixed number of variables. On the other hand the algorithm is impractical for a large number of variables and its implementation is not clear even for a small number of variables. We suggest here a few simplifications and improvements to that algorithm, making its implementation easy (though still impractical for a great number of variables). As a byproduct we show how to solve diophantine linear equations over the nonnegative integers. For a small number of variables (3 or 4) a practical and fast algorithm for solving such equations results.

1. Definitions

As this work relies heavily on the work of H.W. Lenstra [1] the reader is assumed to be familiar with that work and we shall use many of the notations used there. A convex set will be given here in the form

$$K = \{x \in R^n, (1,x)B \geq 0\} \tag{1}$$

where B is an $(n+1) \times m$ matrix of integers and $(1,x) = (1,x_n,x_{n-1}, \ldots, x_1)$. The number of variables involved (n) is the **dimension** of the set. K can also be given as the convex hull of its vertices and those vertices can be found by solving at most $\binom{m}{n} \leq m^n$ systems of n equations derived from B. Let $v_1,v_2 \cdots v_l$ be the vertices of K, $l > 1$. Let d be the dimension of the linear subspace generated by the vectors $v_j - v_0$, $1 \leq j \leq l$. Then d will be called the **rank** of K. If $d = n$ then K is of **full rank**.

An **n-dimensional linear diophantine equation** is an equation of the form

$$\sum_{i=n}^{1} a_i x_i = M \tag{2}$$

where the a_is and M are assumed to be nonnegative integers and a solution is sought over the nonnegative integers. An n-dimensional hyperplane is a set of points given in the form

$$P = \{x \in R^n: \ x\eta = M\} \tag{3}$$

where η is a column vector of integers (not necessarily nonnegative) and M is an integer.

A **translate** of a hyperplane is another hyperplane with the same vector η but a different M.

2. The intersection of a convex set and a hyperplane.

We describe first a few procedures to be used in the main algorithm.

Two convex sets K_1 of dimension n_1 and K_2 of dimension n_2 will be called I-equivalent if there is a 1-1 onto mapping τ from K_1 to K_2 such that $x \in K_1 \cap Z^{n_1}$ iff $\tau x \in K_2 \cap Z^{n_2}$. The first procedure considered is the following.

Procedure 'Cut'

Given a convex set K of dimension n of the form

$$K^{(0)} = \{x \in R^n: \ (1,x)B^{(0)} \geq 0\}$$

and a hyperplane

$$P = \{x \in R^n: \ x\eta = M\}$$

we want to find a new convex set $K^{(1)}$ of dimension $n-1$ and I-equivalent to $P \cap K^{(0)}$, so that $K^{(1)}$ will have the form

$$K^{(1)} = \{y \in R^{n-1}: \ (1,y)B^{(1)} \geq 0\}$$

where $B^{(1)}$ is $n \times m$ with all its entries integers. Let $\eta = (a_n, a_{n-1}, \ldots, a_1)^\tau$ and assume that $\gcd(a_1, \ldots, a_1) = 1$ (otherwise the gcd of the entries of η must divide M, or $|P| = \Phi$, and the equation $x\eta = M$ can be divided by $\gcd(a_n, \ldots, a_1)$).

The procecure consists of the following steps:

1. Let $f_n = a_n$ by definition and construct pairs of integers (t_i, s_i); $1 \le i \le n-1$ satisfying

$$t_i f_{i+1} - s_i a_i = \gcd(t_{i+1}, a_i) = f_i .$$

The evaluation of the pairs (t_i, s_i) can be done via the extended Euclidean algorithm and the time complexity (when counting the number of arithmetical operations) is linear in the length of the input (see [3]).

When the evaluation is completed we know that $a_n = f_n > f_{n-1} > \cdots > f_k = f_{k-1} = \cdots = f_1 = 1$. As $\gcd(a_n, a_{n-1} \cdots a_1) = 1$ we have that $f_1 = 1$ necessarily.

2. Construct the following matrix.

$$A = \begin{bmatrix}
a_n & s_{n-1} & a_n s_{n-2}/f_{n-1} & a_n s_{n-3}/f_{n-2} & a_n s_K/f_{K+1} & 0 & \cdots & 0 \\
a_{n-1} & t_{n-1} & a_{n-1} s_{n-2}/f_{n-1} & a_{n-1} s_{n-3}/f_{n-2} & a_{n-1} s_K/f_{K+1} & 0 & \cdots & 0 \\
a_{n-2} & 0 & t_{n-2} & a_{n-2} s_{n-3}/f_{n-2} & a_{n-2} s_K/f_{K+1} & 0 & \cdots & 0 \\
a_{n-3} & 0 & 0 & t_{n-3} & a_{n-3} s_K/f_{K+1} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \\
a_{K+1} & & & & a_{K+1} s_K/f_{K+1} & 0 & \cdots & 0 \\
a_K & & & & t_K & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & 1 & \ddots & 0 \\
a_1 & 0 & 0 & 0 & 0 & 0 & \cdots & 1
\end{bmatrix}$$

where the lower right corner of A is the unit matrix of dimension $(k-1) \times (k-1)$.

Comment: It follows from the definition of the f's that all the entries of A are integers. Let $A[j]$ be the $j \times j$ upper left corner of A and let $A[i, j+1]$ be the submatrix of A containig the first j rows, and column 2 to $j+1$ (inclusive). It is easily proven by induction that

$$\det(A[j]) = f_{n-j+1} \quad 2 \le j \le n$$

and

$$\det(A[1, j+1]) = s_{n-j} \quad 2 \le j \le n-1$$

and this implies that

$$\det(A) = \det[A(n)] = f_K = 1$$

Thus A is a unimodular matrix.

3. Construct A^{-1}. As the coefficients of A are integers and $\det(A) = 1$, A^{-1} must have the same properties. Let $\hat{A}$ be the matrix resulting from A^{-1} when its first row is multiplied by M. For any $(n-1)$-dimensional vector y we have that

$$(M,y)A^{-1} = (1,y)\hat{A} .$$

4. Set $B^{(1)} = \begin{bmatrix} 1 & \\ 0 & \\ \cdots & \hat{A} \\ 0 & \end{bmatrix} B^{(0)}$ and

$$K^{(1)} = \{y \in R^{n-1} \; (1,y)B^{(1)} \geq 0\}$$

We claim that $K^{(1)}$ is I-equivalent to $P \cap K^{(0)}$. By definition $B^{(1)}$ is $n \times m$.

Proof: For any $x \in P \cap K^{(0)} \cap Z^n$. Define y to be the $n-1$ last entries of the vector $xA = (M,y)$. Notice that, as $x \in L$ and the first column of A is y, by construction, the first entry of xA must be equal to M. The coefficients of x and A are integers implying that the coefficients of y are integers too.

Given $y \in K^{(1)} \cap Z^{n-1}$. Define x by

$$x = (1,y)\hat{A} = (M,y)A^{-1}$$

so that

$$x\eta = (M,y)A^{-1}\eta = (M,y)(1,0,0...,0)^T = M .$$

Again all the coefficients of x are integers, following from the fact that the coefficients of A^{-1} and of y are such. We also have the following:

$$(1,y)B^{(1)} = (1,y)\begin{bmatrix} 1 & \\ 0 & \\ \cdots & \hat{A} \\ 0 & \end{bmatrix} B^{(0)} =$$

$$= (1,1,y)\begin{bmatrix} 1 & 0 \; \ldots \; 0 \\ 0 & \\ \cdots & \hat{A} \\ 0 & \end{bmatrix} B^{(0)} = (1,(1,y)\hat{A})B^{(1)} =$$

$$= (1,(M,y)A^{-1})B^{(0)} = (1,x)B^{(0)}$$

Thus $(1,y)B^{(1)} \geq 0$ iff $(1,x)B^{(0)} \geq 0$ and $x\eta = M$. This completes the proof.

The following two remarks conclude this section.

Remark 1. Notice that the number of columns of $B^{(1)}$ is the same as the number of columns of $B^{(0)}$, i.e. the number of inequalities defining the consecutive convex sets is the same.

Remark 2. If, after the final reduction to dimension $n-1$, a new $(n-1)$-dimensional "cutting" hyperplane is provided allowing for another reduction to cimension $n-2$ and so on, it is very easy to find an integral vector in the original convex set corresponding to the integer in the final set (which is of dimension 1). This is done as follows.

Let the sequence of $\hat{A}$ matrices as constructed in stage 3 of the procedure be denoted by $\hat{A}^{(1)},\hat{A}^{(2)},\ldots,\hat{A}^{(n-1)}$ where $\hat{A}^{(1)}$ is $n \times n$ $\hat{A}^{(2)}$ is $(n-1) \times (n-1)$, etc. $\hat{A}^{(n-1)}$ is 2×2. Let $x^{(0)},x^{(1)},\ldots,x^{(n-1)}$ be the corresponding solution vectors, i.e. $x^{(0)}$ is n-dimensional, $x^{(1)}$ is $n-1$, etc. $x^{(n-1)}$ is one dimensional. Then

$$x^{(n-2)} = (1,x^{(n-1)})\hat{A}^{(n-1)},x^{(n-3)} = (1,x^{(n-2)})\hat{A}^{(n-2)},$$

to

$$x^{(0)} = (1\ x^{(1)}\hat{A}^{(1)}) .$$

Define the sequence of matrices $C^{(i)}$ as follows

$$C^{(1)} = \begin{bmatrix} 1 & \\ 0 & \\ \cdots & \hat{A}^{(1)} \\ 0 & \end{bmatrix}, \quad C^{(2)} = \begin{bmatrix} 1 & \\ 0 & \\ \cdots & \hat{A}^{(2)} \\ 0 & \end{bmatrix} C^{(1)},\ldots$$

$$C^{(n-1)} = \begin{bmatrix} 1 & \\ 0 & \hat{A}^{(n-1)} \end{bmatrix} C^{(n-1)}$$

Thus $(1\ x^{(0)}) = (1,x^{(n-1)})\ C^{(n-1)}$.

One must find therefore, at the i-th iteration step, the corresponding $C^{(i)}$ matrix which is computed out of $C^{(i-1)}$ and the corresponding new $\hat{A}^{(i)}$ matrix. After the last iteration the $x^{(0)}$ vector can be computed out of $x^{(n-1)}$ rightaway

as explained above.

Complexity analysis. If N is the length of the input and n is the dimension then: Step 1 is $0(N)$, Step 2 is $0(n^2)$ Step 3 is $0(n^3)$ and Step 4 is $0(n^3)$. The complexity of the procedure is therefore $0(n^3 + N)$ when we count the number of arithmetical operations.

3. Convex sets whose rank is not full.

The second procedure to be described reduces the dimension of a convex set whose rank is not full, using the previous procedure.

Let K be given as in (1) and assume that $\text{rank}(K) < n$.

Procedure "Reduce Dimension"

Let $v_1 \ldots v_l$ be the vertices of K, assumed to be of rank $d < n$. The procedure consists of the following steps.

1. Find a subset of the v's $w_0, w_1 \cdots w_d$ such that the vectors $w_i - w_0$, $1 \le i \le d$ span the space spanned by the vectors $v_i - v_0$, $1 \le i \le l$. Such a subset can be found easily, using methods from linear algebra.

2. Construct a determinant $D(v_j)$ as follows. The i-th row of $D(v_j)$ for $1 \le i \le d+1$ equals the vector w_{i+1} extended to an $n+1$ dimensional vector with last entry equal to 1. The i-th row of $D(v_j)$ for $d+2 \le i \le n$ equals the $(n+1)$-dimensional vector with all its entries equal to 0 except for the i-th entry which is equal to 1 (unit vectors). The $(n+1)$-st (and last) row of $d(v_j)$ is the vector v_j extended to an $n+1$-dimensional vector with its last entry equal to 1. As all the v_j's are linear combinations of the w_i's, by definition, we have that $D(v_j) = 0$, $0 \le j \le l$. Let x be the vector $x = (x_n \cdots x_1)$. It follows that all the vectors v_j, $0 \le j \le M$, are included in the hyperplane defined by the linear equation $D(x) = 0$ and therefore the whole convex set is included in that hyperplane. We have thus found a hyperplane such that

the given convex set can be defined in the form

$$K = P \cap K^{(0)}$$

where $K^0 = \text{conv}(v_0,\ldots,v_l)$ and $P = \{(x_n \cdots x_1)\colon D(x) = 0\}$. As all the constant entries in $D(x)$ are rational $D(x) = 0$ can be expressed as a hyperplane with all its coefficients integers.

3. Use procedure "cut" in order to find an I-equivalent convex set of dimension $n-1$.

Complexity analysis. It is easy to see that steps 1 and 2 are $O(n^3l)$. This follows from the fact that all the computations involved amount to the evaluation of $n \times n$ matrices, with at most one such evaluation for every vector v_j. Step 3 is $O(n^3+N)$ where N is the length of the input, as mentioned before. The whole procedure is therefore $O(n^3l+N) \leq O(n^3m+N)$.

4. Basis reduction

The next procedure we want to use is given explicitly in [2].

Procedure 'unimodular transformation'

Given a basis $u_1,\ldots,u_n$ for a lattice L such that all the coefficients of the vectors u_i, $1 \leq i \leq n$, are integers. Let U be a matrix whose rows are the vectors u_i. Find a unimodular transformation W such that the rows of the matrix WU are a **reduced** basis for L ('reduced' as in [2]). We want to get the unimodular matrix W only and are not interested in the reduced basis (the rows of WU).

It was shown in [2 prop. 1.26] that the complexity of the procedure is $O(n^4 \log N)$ where N is the square of the length of the maximal length vector among the vectors b_i, $1 \leq i \leq n$, when the coefficients of b_i are assumed to be integers.

One additional procedure is needed before constructing the main program which is in the next section.

5. Finding the cutting hyperplane.

Consider a convex set of the form

$$K = \{(x_n \ldots x_1) \in R^n: \ (1, x_n \cdots x_1)\, B \geq 0\}$$

as defined in (1). Assume that the vertices of K, denoted by $v_0, v_1 \cdots v_l$ are given and assume that K is of full rank. We want to find, in polynomial time (in the length of the input) a hyperplane of the form

$$P_M = \{(x_n \cdots x_1) \in R^{(n)}: \ x\eta = M\}$$

as defined in (2) such that the number of translates of P_M intersecting K is proportional to the number of integral points in the set $K \cap Z^n$.

The procedure is basically similar to a part of Lenstra's algorithm [1, sec. 1] with several simple but important chages to be described in the sequel. The procedure involves several steps. The reader is referred to Lenstra's paper for the implementation of the steps whose implementation is not described here.

Procedure "cutting planes"

1. Choose $n+1$ vertices out of the given vertices such that the volume of the n-simplex spanned by them is maximal. Let those vertices be $v_0, v_1, \ldots, v_n$. Construct a (nonsingular) $n \times n$ matrix whose rows are the vectors $v_i - v_0$, $1 \leq i \leq n$. Denote this matrix by U. Notice that all the entries of U are rational numbers.

2. Construct the matrix U^{-1}. The entries of U^{-1} are rational (as are the entries of U). Let U^{-1} be considered as an afine transformation $R^n \to R^n$. It follows from the definitions that $\mathrm{conv}(v_0, \ldots, v_n)$ is transformed by U^{-1} into a simplex which is a translate of the simplex whose vertices are the origin together with the n unit vectors, while K is transformed into a convex set which is included in a translate of the simplex whose vertices are the origin and the unit vectors doubled (i.e. (0,...,0), (2,0,...,0),...,(0,...,0,2)).

Moreover the natural lattice is transformed by U^{-1} into a lattice such that the rows of U^{-1} can be taken as a basis for it, to be denoted by $u_1 \cdots u_n$.

3. Use the "unimodular transformation" procedure to get the unimodular matrix W for the matrix U^{-1} above.

   Remark: Let u'_i, $1 \le i \le n$, be the rows of the matrix WU^{-1}. It was proved in [2, proposition (1.6)] that

$$\det(U^{-1}) \le \prod_{n}^{i=1} |u'_i| \le 2^{\frac{n(n-1)}{4}} \det(U^{-1})$$

4. Let u'_i be the rows of the matrix WU^{-1}, let $W = [w_{ij}]$ and let $|u'_j| = \max\{|u'_i|: 1 \le i \le n\}$. Let $W^{(j)}$ be the matrix resulting from W when its j-th row is replaced by the row of variables $(x_n..x_1)$. Return the hyperplanes

$$P_M = \{(x_n \cdots x_1): \det(W^{(j)}) = M\}$$

   where M ranges over all the integers such that $P_M \cap K \neq \Phi$.

We proceed now to show that our procedure is correct. We prove first the following.

Lemma 1. There are two hyperspheres S_1 and S_2 with corresponding radiuses r and R such that

$$S_1 \subset KU^{-1} \subset S_2$$

and

$$R/r = n(\sqrt{n}+1)$$

where KU^{-1} is the convex set resulting from K when U^{-1} is applied to its vertices and U^{-1} is as defined in step 2 of the above procedure.

Proof. As explained in steps 1 of the above procedure KU^{-1} contains a translate of the simplex with vertices (0,...,0), (1,0,...,0),...,(1,...,0,1) and is contained in a translate of the simplex with vertices (0,...,0), (2,0,...,0),...,(0,...,0,2) containing the first simplex in its interior. S_1 and S_2 can therefore be defined

as the hyperspheres, one inscribed in the first simplex and the other circumscribed over the second simplex correspondingly. We find r first. S_1 is tangent to the hyperlanes $x_i = 0$ implying that the center of S_1 is located at the point $(r,r,...,r)$. S_1 is tangent to the hyperplane $x_1 + x_2 \cdots + x_n = 1$ touching it at the point $(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$. Thus r must satisfy the equation $r^2 = n(\frac{1}{n} - r)$ (one can consider r to be the diagonal of a hypercube whose side has length $(\frac{1}{n} - r)$). Solving for r we get that

$$r = \frac{1-\sqrt{\frac{1}{n}}}{n-1}$$

(The other solution is not physical.)

To compute R we notice that the circumscribed hypersphere passes through the vertices of the simplex $(0,...,0),(2,0,...,0),(0,2,...,0),...,(0,0,...,2)$. Its diameter is therefore the diagonal of a hypercube whose side has length 2. The length of the diameter is therefore $2\sqrt{n}$. Thus $R = \sqrt{n}$. We have therefore that $2\sqrt{n}$.

$$\frac{R}{r} = \frac{\sqrt{n}(n-1)}{1-\sqrt{\frac{1}{n}}} = \frac{n(n-1)}{\sqrt{n}-1} = \frac{(\sqrt{n}+1)n(n-1)}{n-1} = n(\sqrt{n}+1)$$

Q.E.D.

Let $c_1 = 2^{\frac{n(n-1)}{4}}$, let $c_2 = n(\sqrt{n} + 1)$ and let u'_j be the vector defined in step 4 of the given procedure. The hyperplanes parallel to the set of vectors $\{u'_i: i \neq j\}$ (the u'_i's are defined in step 4 of the given procedure) will be called, for the purpose of the next theorem, lattice planes. Using the same arguments as the ones used by Lenstra [1] one can now prove the following theorem (set in a more general form).

Theorem 2. For a given convex set K defined as in (1). If $r \geq \frac{t}{2}|u'_j|\sqrt{n}$ (r as defined in the previous lemma) then there are at least t^n points in $K \cap Z^n$. If

$r < \frac{t}{2}|u'_j|\sqrt{n}$ then at most $t \cdot c_1 \cdot c_2 \cdot \sqrt{n}$ lattice planes have nonempty intersection with KU^{-1}.

Corollary: If more than $t \cdot c_1 \cdot c_2$ lattice planes have nonempty intersection with KU^{-1} then there are at at least t^n points in $K \cap Z^n$.

The proof is the same as the proof given by Lenstra but with the constant c_2 as defined and justified here. The additional constant t used here stems from the fact that if $r \geq \frac{t}{2}|u'_j|\sqrt{n}$ then the simplex in which r is inscribed can be split into t^n similar simplices every one of them containing an inscribed hypersphere with radius $\geq \frac{1}{2}|u'_j|\sqrt{n}$.

If we apply now the transformation U (inverse to U^{-1}) to KU^{-1} we get back K. Applying U to the "reduced" basis of the transformed lattice: $u'_1, \ldots, u'_n$, we get a basis for the natural lattice consisting of the rows of the matrix W (W is unimodular) $w_1,\ldots,w_n$ (whose entries are integers) and u'_j is transformed into w_j. But U is a linear afine transformation preserving parallelism, so the resulting hyperplanes (step 4) are parallel to all the vectors w_i, $i \neq j$. It follows that the hyperplanes as defined in step 4 have the following property: If the number of translates of P_M whose intersection with K is nonempty is bigger than $t \cdot c_1 \cdot c_2$ then at least t^n points are included in $K \cap Z^n$. To find the specific translates P_M having nonempty intersection with K, substitute the vertices v_i of K for the variables row in $w^{(j)}$ resulting in

$$M_i = \det \begin{bmatrix} d_1 \\ \vdots \\ d_{j-1} \\ v_i \\ d_{j+1} \\ \vdots \\ d_n \end{bmatrix}$$

Then P_M will cut K for all M with

$$M_0 = \lceil \min M_i \rceil \leq M \leq \lfloor \max M_i \rfloor = M_1$$

We have thus shown that the procedure is correct.

We will show now that the complexity of step 3 of our procedure is polynomial in the length of the input. One can multiply first U^{-1} by the gcd of the denominators of its entries resulting in $\bar{U}$. All the entries of $\bar{U}$ are now integers without changing the complexity of step 3. As mentioned before (see description of procedure 'unimodular transformation'), for this case (vectors with integer coefficients) the complexity of that part of the algorithm is $0(n^4 \log N)$ where N is the square of the length of the maximal length row vector in the matrix $\bar{U}$. Let Q be the maximal value among the coefficients of the matrix B defining the convex set K. Then the magnitude of the row vectors of $\bar{U}$ are at most exponential in Q with the exponent a polynomial in n. Therefore log N is linear in the length of the input.

The complexity analysis of the procedure now follows:

Step 1 is $0(n^3\binom{m}{n})$

Step 2 is $0(n^3)$

Step 3 is $0(n^4N)$ where N is the length of the input.

Step 4 is $0(n^3)$

The procedure is therefore polynomial (even linear) in the length of the input for fixed n.

6. The main program

We can describe now our version of the integer programming algorithm as follows.

IPA(n, K, C)

Input: an integer n; an n-dimensional convex set $K = \{x \in R^n: (1,x)B \geq 0\}$ where B is an $(n+1) \times m$ matrix of integers, and K is assumed to be bounded so

that $m \geq n+1$; C an $(n+1) \times m_1$ matrix of integers where $n+1 \leq m_1 \leq m+1$.

1. If $n = 1$, then for every integer $x \in K$ (x is now a vector of dimension 1) return the m_1-dimensional vector $(1,x)C$ and stop.

2. Find the vertices and the dimension d of K.

3. If $d < n$ then apply procedure 'reduce dimension' to K with output (of procedure) $K^{(1)}$ and reducing matrix $\hat{A}^{(1)}$ (see definition of procedure 'cut'). Then set $K \leftarrow K^{(1)}$, $C \leftarrow \begin{bmatrix} 1 & \\ 0 & \\ \cdots & \hat{A}^{(1)} \\ 0 & \end{bmatrix} C$, $n \leftarrow n-1$ and apply $IPA(n,K,C)$.

4. Here $d = n$. Apply procedure 'cutting planes' to K.

5. For every cutting hyperplane P_M, $M_0 \leq M \leq M_1$, returned in 4 and K apply procedure 'cut.' Let $K_M^{(1)}$ and $\hat{A}_M^{(1)}$ be the resulting $(n-1)$-dimensional convex sets and corresponding reducing matrices. Set $K_M \leftarrow K_M^{(1)}$, $n \leftarrow n-1$, $C_M = \begin{bmatrix} 1 & \\ 0 & \\ \cdots & \hat{A}_M^{(1)} \\ 0 & \end{bmatrix} C$. For every $M_0 \leq M \leq M_1$, apply $IPA(n-1,K_M,C_M)$.

6. For given n and K apply **IPA**(n,K,B).

Complexity analysis. All the steps of the above algorithm are polynomial (even linear) in the length of the input. The number of times recursion is applied is at most n and whenever recursion is applied the coefficients grow (in their length) by a factor which depends on n-only. It follows that the algorithm is polynomial in the length of the input.

7. Comparison with Lenstra's algorithm

The advantages of our algorithm when compared with Lenstra's are listed below.

1. The procedure 'reduce dimension' is given explicitly and in an easy programmable form. It replaces the first part of section 2 in [1].

2. The most significant difference is included in steps 4 and 5. Our transformation on U^{-1}, replacing Lenstra's transformation τ, is given here explicitly, has **rational** coefficients and is used only as an **auxiliary** step needed for the finding of the cutting hyperplanes - which have integral coefficients. After the cutting plane is found, using rational arithmetic, the computation returns to the **natural** lattice where integer arithmetic only is used. Our algorithm can, therefore, be easily transcribed into a computer program. It is also worth noticing that proposition 1.26 in [2] analysing the complexity of the procedure involved in the finding of a reduced basis for a given lattice (procedure 'unimodular transformation' here) assumes that the coefficients of the given base are **integers**. This condition can be met here, multiplying U^{-1} (whose rows are the given basis) by the gcd of its coefficients. As mentioned above, the only part of our algorithm using rational arithmetic is the auxiliary part where a set of cutting hyperplanes with certain properties is sought. It can be show that the above part of our algorithm is reducible to a diophantine approximation problem but we will not pursue this subject here, as we do not know how to solve the corresponding diophantine approximation problem, in polynomial time, for the general case. We shall consider however the case $n = 2$ in the appendix where a simplification of the general algorithm (for $n = 2$) is shown based on the above observation.

3. Some further simplifications might be possible based on:

   a. The large amount of parallelism present in the various steps of the algorithm.

   b. The possibility of saving some of the computations in symbolic form e.g. the matrix A^{-1} in step 3 of the 'procedure cut' can be computed right away from step 1 without going through step 2, as that matrix

(A^{-1}) can be shown to have a very simple form.

c. The possibility of preconditioning (see example at the end of this paper).

d. The vertices of the convex sets obtained after every reduction can be computed directly from the intersection of the cutting planes with the previous convex set (before the reduction): The reduction transforms those intersections points into the vertices of the new convex sets.

Those possible simplifications and others will be investigated in a subsequent work.

8. Linear diophantine equations

Given a linear diophantine equation

$$\sum_{i=n}^{1} a_i x_i = M$$

where the a_i's and M are nonnegative. A solution is sought over the nonnegative integers.

This problem is easily reduced to an integer programming problem of dimension $n-1$ as follows:

Let P_M be the hyperplane

$$P_M = \{x \in R^n: \sum_{i=n}^{1} a_i x_i = M\}$$

Let $K^{(1)}$ be the positive orthant:

$$K^{(0)} = \{x: (1,x)I \geq 0\}$$

where I is the unit matrix of dimension $n+1$. Use the procedure 'cut' on P_M and $K^{(0)}$ resulting in the equivalent convex set: $K^{(1)} = \{y \in R^{n-1}: (1,y)B^{(1)} \geq 0\}$ where

$B^{(1)} = \begin{bmatrix} 1 & \\ 0 & \\ \cdots & \hat{A} \\ 0 & \end{bmatrix}$ and $\hat{A}$ is defined as in step 3 of procedure 'cut.'

Apply now $IPA(n-1,K^{(1)},B^{(1)})$. This algorithm for solving linear diophantine equations reduces to a very simple algorithm for small n (e.g. $n=3$ or $n=4$) due to the simple form of the first iterations of the *IPA* for this case, and due to an additional possible simplification explained in the appendix. The following example, with $n=4$, provides some insight and shows the additional simplifications possible for small n.

9. An example

The following equation is mentioned in [4] as an example for another problem.

$$271\,x_4 + 281\,x_3 + 283\,x_2 + 277\,x_1 = M$$

It was shown in [4] that the biggest M for which the above equation has no solution over the nonnegative integers is $M = 13022$.

1. Set $K^{(0)} = \{(x_4x_3x_2x_1);\ (1x_4x_3x_2x_1)\,I \geq 0\}$ and apply procedure 'cut':

   1.1. Find (t_i,s_i): $28\cdot 271 - 27\cdot 281 = 1 = f_3 = f_2 = f_1 = 1$.

   1.2. Construct the A matrix

$$A^{(1)} = \begin{bmatrix} 271 & 27 & 0 & 0 \\ 281 & 28 & 0 & 0 \\ 283 & 0 & 1 & 0 \\ 277 & 0 & 0 & 1 \end{bmatrix}$$

   and then A^{-1} (as A is almost triangular this is an easy task) and then $\hat{A}$ we find that

$$\hat{A}^{(1)} = \begin{bmatrix} 28\cdot M & -27\cdot M & 0 & 0 \\ -281 & 271 & 0 & 0 \\ -28\cdot 283 & 27\cdot 283 & 1 & 0 \\ -28\cdot 277 & 27\cdot 277 & 0 & 1 \end{bmatrix}$$

   so

$$K^{(1)} = \{(y_3,y_2,y_1):\ (1,y_3y_2y_1)B^{(1)} \geq 0\}$$

   where

$$B^{(1)} = \begin{bmatrix} 1 & \\ 0 & \\ 0 & \hat{A}^{(1)} \\ 0 & \end{bmatrix}.$$

2. Apply now the main algorithm to $K^{(1)}$.

2.1. Find the vertices of $K^{(1)}$. As the i-th row of A is perpendicular to the j-th column of A^{-1}, for $j \neq i$, the vertex corresponding to those j columns can easily be found out of the i-th row of A. The vertices are:

$$(\frac{27 \cdot M}{271},0,0), (\frac{28 \cdot M}{281},0,0), (0, \frac{M}{283},0), (0,0, \frac{M}{277}).$$

2.2. $d = 3$. (This will always hold for this particular case.)

2.3. Apply procedure 'cutting planes.' As $K^{(1)}$ is already a simplex we can construct right away the matrix U which is (notice that $\frac{28 \cdot M}{281} - \frac{27 \cdot M}{271} = \frac{M}{281 \cdot 271}$ by 1.1).

$$U = \begin{bmatrix} \frac{M}{281 \cdot 271} & 0 & 0 \\ \frac{-27 \cdot M}{271} & \frac{M}{283} & 0 \\ \frac{-27 \cdot M}{271} & 0 & \frac{M}{277} \end{bmatrix}$$

Out of this we get U^{-1} easily:

$$U^{-1} = \begin{bmatrix} \frac{281 \cdot 271}{M} & 0 & 0 \\ \frac{27 \cdot 283 \cdot 281}{M} & \frac{283}{M} & 0 \\ \frac{27 \cdot 277 \cdot 281}{M} & 0 & \frac{277}{M} \end{bmatrix}$$

The rows of U^{-1} are the basis which must be reduced. Applying the procedure 'unimodular transformation' (which turns out to be quite simple here as U^{-1} is triangular) and then get the cutting planes

$$5y_3 + 141y_2 + 138y_1 = N \qquad (*)$$

We have left the value of M undefined to stress the fact that some of the calculations are independent on M. The values of N above depend on the value of M. Let e.g. $M = 13022$ then the only vaue of N for which

the above plane cuts $K^{(1)}$ is found to be $N = 6487$.

We can proceed now to reduce our problem to a 2-dimensional problem using again the procedure 1. We find first that

$$5 \cdot 113 - 141 \cdot 4 =$$

which implies the transformation

$$A^{(2)} = \begin{bmatrix} 5 & 4 & 0 \\ 141 & 113 & 0 \\ 138 & 0 & 1 \end{bmatrix}$$

$$\hat{A}^{(2)} = \begin{bmatrix} 113 \cdot 6487 & -4 \cdot 6487 & 0 \\ -141 & 5 & 0 \\ -113 \cdot 138 & 4 \cdot 138 & 1 \end{bmatrix}$$

$$C^{(2)} = \begin{bmatrix} 1 & & \\ 0 & \hat{A}^{(2)} & \\ 0 & & \end{bmatrix} \cdot \hat{A}^{(1)} = \begin{bmatrix} 1 & -5143 & 31139 & -25948 & 0 \\ 0 & 1 & -6 & 5 & 0 \\ 0 & 110 & -663 & 552 & 1 \end{bmatrix}$$

The resulting 2-dimensional equivalent convex set is now

$$K^{(2)} = \{(z_2 z_1): \ (1 \ z_2 z_1) \ C^{(2)} \geq 0\}$$

Its vertices are found to be

$$(4932, \frac{7}{3}), \ (\frac{25948}{5}, 0), \ (\frac{31139}{6}, 0)$$

Applying procedure 'cutting planes' to this new convex set we get the cutting lines

$$z_2 + 110 \, z_1 = N$$

The only translate of the above family of lines cutting $K^{(2)}$ is the one with $N = 5189$. We apply now again procedure 'cut'. As

$$111 \cdot 1 - 1 \cdot 110 = 1$$

we get

$$A^{(3)} = \begin{bmatrix} 1 & 1 \\ 110 & 111 \end{bmatrix}$$

$$\hat{A}^{(3)} = \begin{bmatrix} 111 \cdot 5189 & 5189 \\ -110 & 1 \end{bmatrix}$$

$$C^{(3)} = \begin{bmatrix} 1 & \\ 0 & \hat{A}^{(3)} \end{bmatrix} C^{(2)} =$$

$$= \begin{bmatrix} 1 & -46 & 15572 & -10381 & -5189 \\ 0 & 0 & -3 & 2 & 1 \end{bmatrix}$$

So $K^{(3)} = [w\colon\ (1\ w)C^{(3)} \geq 0]$ The one-dimensional inequalities involved in $K^{(3)}$ are:

$$\begin{cases} 46 \geq 0 \\ 3w \leq 15572 \\ 2w \geq 10381 \\ w \geq 5189 \end{cases} \text{ or } \begin{cases} 46 \geq 0 \\ w \leq 5190.66 \\ w \geq 5190.5 \\ w \geq 5189 \end{cases}$$

(The first and last are superfluous.) There is no integral solution to the above system and therefore no solution to the original problem. We modify now the problem a little bit by increasing the right hand side of the original equation by 1, to 13023. Most of the computation done previously can be used for this new M. Thus we can evaluate $\hat{A}^{(1)}$ immediately substituting 13023 for M so that $K^{(1)}$ is now defined. The same is true for the following computations, resulting in the same family of translates cutting $K^{(1)}$: The planes

$$5y_3 + 141y_2 + 138y = N\ ,$$

but now the only translate cutting $K^{(1)}$ is the one with $N = 6488$. The resulting $A^{(2)}$ matrix is the same as before but $\hat{A}^{(2)}$ is now

$$\hat{A}^{(2)} = \begin{bmatrix} 113 \cdot 6488 & -4 \cdot 6488 & 0 \\ -141 & 5 & 0 \\ -113 \cdot 138 & 4 \cdot 138 & 1 \end{bmatrix}$$

resulting in

$$C^{(2)} = \begin{bmatrix} 1 & & \\ 0 & \hat{A}^{(2)} & \\ 0 & & \end{bmatrix} C^{(1)} =$$

$$= \begin{bmatrix} 1 & -5172 & 31171 & -25952 & 0 \\ 0 & 1 & -6 & 5 & 0 \\ 0 & 110 & -663 & 552 & 1 \end{bmatrix}$$

and $K^{(2)} = \{(1\ z_2\ z_1)\colon\ (1\ z_2\ z_1)C^{(2)} \geq 0\}$. The vertices of $K^{(2)}$ are now found:

$$(\frac{31171}{6},0),\ \frac{(25952}{5},0),\ (\frac{226}{3},\frac{139}{3}),\ (112,46).$$

As in the previous case the family of translates cutting $K^{(2)}$ is found to be

$$2\ z_2 + 221\ z_1 = N$$

and all translates with $N = 10381 + t$ where $0 \leq t \leq 9$ have a nonempty intersection with $K^{(2)}$. From

$$2 \cdot 111 - 221 \cdot 1 = 1$$

we get that

$$A^{(3)} = \begin{bmatrix} 2 & 1 \\ 221 & 111 \end{bmatrix}$$

and $\bar{A}^{(3)}$ can assume any of the following forms, with $0 \leq t \leq 9$.

$$\bar{A}^{(3)} = \begin{bmatrix} (10381+t)111 & -(10381+t) \\ -221 & 2 \end{bmatrix}$$

We compute now $C^{(3)}$

$$C^{(3)} = \begin{bmatrix} 1 & \\ 0 & \bar{A}^{(3)} \end{bmatrix} C^{(2)} =$$

$$= \begin{bmatrix} 1 & 5209+t & 28-3t & 5191+3t & -10381-t \\ 0 & -1 & 0 & -1 & 2 \end{bmatrix}$$

and

$$K^{(3)} = \{w\colon\ (1,w)\ C^{(3)} \geq 0\}$$

The resulting inequalities for w are

$$w \leq 5209 + t$$

$$28 - 3t \geq 0$$

$$w \leq 5191 + 3t$$

$$2w \geq 10381 + t \quad \text{or} \quad w \geq 5190.5 + \frac{t}{2}$$

(the first two are superfluous) where $0 \leq t \leq 1$.

Every integral solution of the above inequalities for any $0 \leq t \leq 9$ generates a solution to the original equation via the transformation

$$(1,x_4,x_3,x_2,x_1) = (1,w)\ C^{(3)}$$

Taking e.g. $t = 3$ reuslts in $5192 \le w \le 5200$ and with $w = 5197$ we get $(x_4,x_3,x_2,x_1) = (15,19,3,10)$. Notice that the number of different solutions found for this case ($M = 13023$) is 125 while in the previous case ($M = 13022$) no solution existed.

Let us sum up this example in "quasi" parametric form.

(1) Starting from the equation

$$271x_4 + 281x_3 + 283x_2 + 277x_1 = M;\ x_i \ge 0.$$

(2) The first reduction resulted in:

$$281y_3 + 7925y_2 + 7756y_1 \le M \cdot 28$$
$$271y_3 + 7641y_2 + 7479y_1 \ge M \cdot 27\ ;\ y_i \ge 0.$$

(3) The cutting planes are

$$5y_3 + 141y_2 + 138y_1 = N$$

with

$$\left\lceil \frac{135}{271} M \right\rceil \le N \le \left\lfloor \frac{141}{283} M \right\rfloor.$$

(4) The second reduction resulted in

$$z_2 + 110z_1 \ge 57N - 28M$$
$$6z_2 + 663z_1 \le 59N - 27M$$
$$5z_2 + 552z_1 \ge 4N\ ;\ z_i \ge 0\ .$$

(5) The cutting lines are

$$2z_2 + 221z_1 = N_1$$

where N_1 ranges between the ceiling of the minimal and the floor of the maximal value among the values

$$\{\frac{59}{3}N - 9M,\ \frac{8}{5}N,\ 14M - \frac{53}{2}N\}$$

(6) The final reduction resulted in

$$w \le 28M - 57N + N_1$$
$$w \le 3N_1 - 4N$$
$$w \ge N_1\ .$$

(7) For every w in th above range we get a solution as below:

$$\begin{bmatrix} x_4 \\ x_3 \\ x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 28 & -57 & 1 & -1 \\ -27 & 59 & -3 & 0 \\ 0 & -4 & 3 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} M \\ N \\ N_1 \\ w \end{bmatrix}$$

After the above "preconditioning" one can solve now the equation (1) for any given M as follows:

a. Find the range of N as in (3).

b. For every such N (if any) find the range of N_1 as in (5).

c. For every N and N_1 found, find the range of w as in (6).

d. Plug in every found triple N, N_1, w and the given value of M into (7) to get a solution of (1).

We would like to make now the following remarks:

1. If a linear diophantine equation is to be solved for several right hand side values M, then part of the computations done for the first M can be used for the consequent M's.

2. When solving a 3-dimensional diophantine equation the equivalent convex set $K^{(1)}$ is of dimension 2 and for this case (dimension 2) a further simplification is possible in the procedure 'cutting planes' which is explained in the appendix. Actually that simplification can always be used for the general *IPA* at the stage when the dimension has been reduced to 2.

3. It might be possible to calculate all the solutions to a given linear diophantine equation concurrently and in a parametric form as was done in the above example.

4. It follows from the appendix that we need to apply the procedure 'cutting planes' only once for linear equations of dimension 4. When reduc-

ing the dimension of the I-equivalent convex set from 3 to 2. An mentioned in section 4 the complexity of that step is $O(n^4 \log N)$ where $n = 3$ and N is the square of the length of the maximal length row vector in MU^{-1} (step 3 in our example). Let $Q = \max(a_i)$ where the a_i's are the coefficients of the given equation. Then, as shown in the example (and true in general for 4-dimensional equations) $N \leq \sqrt{3} \cdot Q^3$.

References

[1] H.W. Lenstra, Jr. "Integer programming with a fixed number of variables" Report 81-03, Mathematish Institut, Amsterdam 1981.

[2] A.K. Lenstra, H.W. Lenstra, Jr. and L. Lovasz. "Factoring polynomials with rational coefficients" Report 82-05, Mathematish Institut, Amsterdam 1982.

[3] G.H. Bradley. "Algorithm and bound for the greatest common divisor of n integers" **Communications of the ACM**, Vol. 13, No. 7, pp. 433-436.

[4] H.S. Wilf. "A circle of lights algorithm for the "money-changing problem"" **American Math Monthly** 85 (1978), pp. 562-565.

Appendix

Consider the following problem:

Given K points in n-space $x^{(1)}, \ldots, x^{(K)}$, $K > n$. Let $a_1x_1 + a_2x_2 + \cdots + a_nx_n = M_i$, $i = 1,2$ be two parallel hyperplanes with integral coefficients and M_i such that

$$a_1x_1^{(j)} + a_2x_2^{(j)} + \cdots + a_nx_n^{(j)} \geq M_1$$

and

$$a_1x_1^{(j)} + a_2x_2^{(j)} + \cdots + a_nx_n^{(j)} \leq M_2$$

for all j, $1 \leq j \leq K$, and such that M_2 is the smallest integer and M_1 is the biggest integer satisfying the above inequalities. For any two such planes

define

$$\Delta(a_1 \cdots a_n) = M_2 - M_1$$

The problem is to find $a_1 \cdots a_n$ which minimize $\Delta(a_1 \cdots a_n)$. A fast algorithm for this problem could simplify the procedure 'cutting planes' in the *IPA*. Unfortunately we do no have such an algorithm for the general case but we conjecture that it exists.

We do have an algorithm for the above problem for $n = 2$ whose complexity is linear in the length of the input. That algorithm is described below.

Let the points be $(x_1^{(i)}, x_2^{(i)})$ $1 \le i \le K$ and $K > 2$. Choose any two points say $(x_1^{(1)}, x_2^{(1)})$ and $(x_1^{(2)}, x_2^{(2)})$, and construct a line $a_1x_1 + a_2x_2 = a_3$, passing through them. It follows that

$$a_1(x_1^{(i_0)} - x_1^{(j_0)}) + a_2(x_2^{(i_0)} - x_2^{(j_0)}) = 0 \tag{1}$$

for any (i_0, j_0). Let $b_1x_1 + b_2x_2 = b_{3i}$ be the two lines we want to find. Then

$$b_{32} \ge \max(b_1x_1^{(i)} + b_2x_2^{(i)}:\ 1 \le i \le K) \tag{2}$$

and

$$b_{31} \le \min(b_1x_1^{(i)} + b_2x_2^{(i)}:\ 1 \le i \le K)$$

To minimize $b_{32} - b_{31}$ we must minimize

$$\max(|\, b_1(x_1^{(i)} - x_1^{(j)}) + b_2(x_2^{(i)} - x_2^{(j)})|:\ K \ge i > j \ge 1) \tag{3}$$

Now by (1) $(x_2^{(i_0)} - x_2^{(j_0)}) = -\frac{a_1}{a_2}(x_1^{(i_0)} - x_1^{(j_0)})$ so that the entry in (3) with $(i,j) = (i_0, j_0)$ for some (i_0, j_0) equals to

$$|\, b_1(x_1^{(i_0)} - x_1^{(j_0)}) - b_2 \frac{a_1}{a_2}(x_1^{(i_0)} - x_1^{(j_0)})| = \tag{4}$$

$$= |(b_1 - b_2 \frac{a_1}{a_2})(x_1^{(i_0)} - x_1^{(j_0)})|$$

The value $|\, b_1 - b_2 \frac{a_1}{a_2}|$ can be decreased by increasing b_2, using diophan-

tine approximation. Let $(b_1 - b_2 \frac{a_1}{a_2}) = \varepsilon(b_2)$ then $b_1 = b_2 \frac{a_1}{a_2} + \varepsilon(b_2)$. We can therefore represent the other entries in (3) in the form

$$|b_2 \frac{a_1}{a_2}(x_1^{(i)} - x_1^{(j)}) + \varepsilon(b_2)(x_1^{(i)} - x_1^{(j)}) + b_2(x_2^{(i)} - x_2^{(j)})| =$$

$$= |b_2(\frac{a_1}{a_2}(x_1^{(i)} - x_1^{(j)}) + (x_2^{(i)} - x_2^{(j)}) + \varepsilon(b_2)(x_1^{(i)} - x_1^{(j)})| \qquad (5)$$

It is clear that the values (5) increase when b_2 increases, from some value of b_2 and on.

The procedure for minimizing (3) can therefore be defined as follows.

1. Initialize. Choose i_0,j_0 which maximize (3) with $b_1 = 1$ and $b_2 = 0$. For this i_0,j_0, b_1 and b_0 (4) is bigger than (5) for all $(i,j) \neq (i_0,j_0)$.

2. Using the diophantine approximation method find successive best approximations (b_1,b_2) to $\frac{a_1}{a_2}$, increasing b_1 until (4) is smaller than (5) for some $(i,j) \neq (i_0,j_0)$. Let $\bar{b}_1,\bar{b}_2$ be the values of b_1,b_2 when this happens and let $\bar{\bar{b}}_1,\bar{\bar{b}}_2$ be the values of b_1,b_2 in the previous step. Choose among the pairs $(\bar{b}_1,\bar{b}_2)$ and $(\bar{\bar{b}}_1,\bar{\bar{b}}_2)$ the pair which minimizes (3) or choose $(\bar{b}_1,\bar{b}_2)$ if both pairs induce the same value for (3).

The reader will easily convince himself that the resulting line is the line which solves our problem. It follows from the theory of continued fractions that the number of arithmetical operations involved is $O(\log a_2)$ which is linear in the length of the input.