Cowles Foundation Paper 818

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 17, No. 3, August 1992
Printed in U.S.A.

THE GENERALIZED BASIS REDUCTION ALGORITHM*

LÁSZLÓ LOVÁSZ AND HERBERT E. SCARF

Let $F(x)$ be a convex function defined in R^n, which is symmetric about the origin and homogeneous of degree 1, and let L be the lattice of integers Z^n. A definition of a reduced basis, $b^1, \dots, b^n$, of the lattice with respect to the distance function F is presented, and we describe an algorithm which yields a reduced basis in polynomial time, for fixed n. In the special case in which the bodies $\{x\colon F(x) \leqslant t\}$ are ellipsoids, the definition of a reduced basis is identical with that given by Lenstra, Lenstra and Lovász (1982) and the algorithm is the well-known basis reduction algorithm.

We show that the basis vector b^1, in a reduced basis, is an approximation to a shortest nonzero lattice point with respect to F and relate the basis vectors b^i to Minkowski's successive minima. The results lead to an algorithm for integer programming which executes in polynomial time for fixed n, but which avoids the ellipsoidal approximations required by Lenstra's algorithm. We also discuss the properties of a Korkine-Zolotarev basis for the lattice.

1. Introduction. Let C be a compact convex body in R^n, of positive volume and symmetric about the origin, and let L be the lattice of integer vectors in R^n. The body can be used to define a distance function $F(x) = \inf\{\lambda \geqslant 0 | x/\lambda \in C\}$, with the properties:

(1) $F(x)$ is convex,
(2) $F(-x) = F(x)$,
(3) $F(tx) = tF(x)$ for $t > 0$.

The dual body C^* is defined to be $\{y | y \cdot x \leqslant 1$ for all $x \in C\}$, and the dual distance function is $F^*(y) = \max_{x \in C} y \cdot x$. We shall assume that F is computable in polynomial time.

In order to determine a smallest nonzero lattice point according to the distance function F, we introduce the concept of a *reduced basis* with respect to F. Let $b^1, b^2, \dots, b^n$ be a basis for the integer lattice L. For each i we project C, along the vectors $b^1, \dots, b^{i-1}$, into the affine space $E_i = \langle b^i, \dots, b^n \rangle$ obtaining C_i. In other words, $x = x_i b^i + \cdots + x_n b^n \in C_i$ if and only if there are $\alpha_1, \dots, \alpha_{i-1}$ such that $x + \alpha_1 b^1 + \cdots + \alpha_{i-1} b^{i-1} \in C$. The lattice L_i, obtained by projecting L along $b^1, \dots, b^{i-1}$ into $\langle b^i, \dots, b^n \rangle$, is the set of integral linear combinations of the vectors $b^i, \dots, b^n$.

The distance function $F_i(x)$, associated with the projected body C_i, is defined for $x \in E_i$ by

$$F_i(x) = \min F\big(x + \alpha_1 b^1 + \cdots + \alpha_{i-1} b^{i-1}\big),$$

with the minimum taken over $\alpha_1, \dots, \alpha_{i-1}$. The function may, of course, be defined for all x in R^N by the same formula; if $x = \sum x_j b^j$, then $F_i(x)$ will be independent of

*Received November 28, 1990; revised April 19, 1991.
AMS 1980 subject classification. Primary 90C10, 52A43, 10E05.
IAOR 1973 subject classification. Main: Programming: integer. Cross Reference: Combinatorial analysis.
OR/MS Index 1978 subject classification. Primary: 627 Programming/integer.
Key words. Geometry of numbers, reduced basis, Lenstra's algorithm, integer programming.

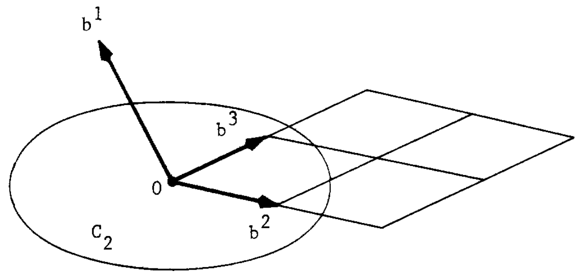

FIGURE 1

$x_1, \ldots, x_{i-1}$. It is elementary to show that

$$F_i(x) = \max\{x \cdot z | z \in C^*, b^1 \cdot z = 0, \ldots, b^{i-1} \cdot z = 0\}.$$

Fix $0 < \epsilon < \frac{1}{2}$. The basis is *reduced*, for this ϵ, if the following two conditions hold for $i = 1, \ldots, n-1$:

(1) $F_i(b^{i+1} + \mu b^i) \geqslant F_i(b^{i+1})$ for integral μ, and

(2) $F_i(b^{i+1}) \geqslant (1-\epsilon) F_i(b^i)$.

A basis $b^1, b^2, \ldots, b^n$, reduced or not, will be said to be *proper* if $F_j(b^i + \mu b^j) \geqslant F_j(b^i)$ for integral μ, and *all* $j < i$. In the arguments of this paper, we shall frequently find it useful to modify a basis $b^1, b^2, \ldots, b^n$ by selecting integers $\mu_{i,j}$, for $j < i$ such that the basis $c^1, \ldots, c^n$ defined by

$$c^i = b^i + \sum_{j=1}^{i-1} \mu_{i,j} b^j \tag{1}$$

is proper. This is done sequentially by taking $\mu_{i,j}$ so as to minimize

$$F_j\left(b^i + \mu_{i,i-1} b^{i-1} + \cdots + \mu_{i,j+1} b^{j+1} + \mu_{i,j} b^j\right),$$

given $\mu_{i,i-1}, \ldots, \mu_{i,j+1}$. The basis $\{c^i\}$, which will be reduced if $\{b^i\}$ is reduced, will be said to be a *proper basis associated with* $\{b^i\}$. If C is an ellipsoid—or, alternatively, if C is the unit ball and the lattice is a general lattice in R^n—a proper reduced basis is identical with the definition of a reduced basis in A. K. Lenstra, H. W. Lenstra, Jr. and L. Lovász (1982).

In §2, we discuss the properties of a reduced basis, demonstrating, in particular, that for such a basis b^1 is an approximation to the shortest nonzero lattice point. In addition, b^i is an approximation to a lattice point realizing the ith successive minimum, according to Minkowski. We also provide a polynomial algorithm for fixed n which finds the shortest nonzero lattice point rather than an approximation.

In §3, the basis reduction algorithm is described and shown to execute in polynomial time, for fixed n. In §4, we examine a special basis—the Korkine-Zolotarev basis—associated with a distance function F. Using the Korkine-Zolotarev basis, we provide an alternative demonstration of a theorem to be found in Kannan and Lovász (1988), that a lattice-free body K, in R^n, has associated with it a nonzero lattice point h, such that the width of the body in the direction h satisfies

$$\max_{x \in K}\{h \cdot x\} - \min_{x \in K}\{h \cdot x\} \leqslant c_0 n(n+1)/2,$$

with c_0, a universal constant.

Lenstra's polynomial algorithm (H. W. Lenstra, Jr. (1983)) for integer programming with a fixed number of variables makes use of the spherical basis reduction algorithm. He begins by a preliminary reduction to the problem of determining a lattice point in a convex polyhedron K, in R^n, defined by a system of linear inequalities $Ax \leqslant c$. To find such a lattice point, the polyhedron is approximated by an ellipsoid E, and a hyperplane with integer normals h is found so that the width of the ellipsoid in the direction h, $\max_{x \in E}\{h \cdot x\} - \min_{x \in E}\{h \cdot x\}$ is as small as possible, aside from a factor depending only on the number of variables, n. If this width is sufficiently large, the polyhedron is sure to contain a lattice point. In the alternative case, in which the width is not large, we consider the intersections of the polyhedron with the hyperplanes $hx = h_0$, with h_0 assuming all integral values between $\min_{x \in E} h \cdot x$ and $\max_{x \in E} h \cdot x$. The n-dimensional problem is thereby reduced to the problem of determining a lattice point in one of a small number of $n - 1$ dimensional polyhedra. Each of these polyhedra is then approximated by its own ellipsoid and the algorithm continues.

A nonzero lattice point h, which minimizes the width of the ellipsoid E, is a shortest nonzero lattice point for the body $(E - E)^*$, itself an ellipsoid. If this latter ellipsoid is transformed to a sphere by a linear transformation, an approximation to the shortest nonzero lattice point can be found using the spherical basis reduction algorithm for a general lattice.

The arguments of this note can be used to find a short nonzero lattice point for the body $(K - K)^*$ directly, thereby avoiding the series of ellipsoidal approximations. The basis reduction algorithm is applied to $C = (K - K)^*$, where $K = \{x \mid Ax \leqslant c\}$, with the distance functions

$$F_i(\xi) = \min_{\alpha} F_1(\xi + \alpha_1 b^1 + \cdots + \alpha_{i-1} b^{i-1})$$

$$= \max \xi \cdot (x - y), \quad \text{subject to}$$

$$Ax \leqslant c, \quad Ay \leqslant c, \quad b^1 \cdot (x - y) = 0, \ldots, b^{i-1} \cdot (x - y) = 0.$$

The analysis of §§3 and 4 can be used to provide an alternative to Lenstra's algorithm, free of ellipsoidal approximations, and which also executes in polynomial time for a fixed number of variables.

The general basis reduction algorithm requires the solution of many linear programs, and there are tradeoffs between using an ellipsoidal approximation to K, or working directly with the body, itself, to resolve the question of whether K contains a lattice point. A number of computational experiments are currently being attempted on integer programming problems of moderately large size to evaluate the merits of the two procedures.

2. Properties of a reduced basis.

THEOREM 1. *Let $b^1, \ldots, b^n$ be a reduced basis. Then*

$$F_{i+1}(b^{i+1}) \geqslant (\tfrac{1}{2} - \epsilon) F_i(b^i) \quad \text{for } i = 1, \ldots, n-1.$$

PROOF. We have the identity $\min F_i(x + \alpha b^i) = F_{i+1}(x)$ with the minimum taken over all real α. Since we can round α to the nearest integer μ, it follows that

$$\min F_i(x + \mu b^i) \leqslant F_{i+1}(x) + \tfrac{1}{2} F_i(b^i), \tag{2}$$

with the minimum taken over integer μ. If x is taken to be b^{i+1} then (2), in conjunction with the definition of a reduced basis, tells us that

$$(1-\epsilon)F_i(b^i) \leqslant F_i(b^{i+1}) = \min F_i(b^{i+1} + \mu b^i)$$
$$\leqslant F_{i+1}(b^{i+1}) + \tfrac{1}{2}F_i(b^i). \quad \square$$

THEOREM 2. *Let $b^1, \ldots, b^n$ be a reduced basis, and let $\lambda_1 = \min F(h)$, for all nonzero lattice points h. Then $\lambda_1 \geqslant F(b^1)\cdot(\frac{1}{2}-\epsilon)^{n-1}$.*

PROOF. Let $h = l_1 b^1 + \cdots + l_k b^k$, with $l_1, \ldots, l_k$ integral and l_k different from zero, be a shortest nonzero lattice point according to the distance function F. Then

$$\lambda_1 = F(h) \geqslant F_k(h) = |l_k| F_k(b^k) \geqslant F(b^1)\cdot\left(\tfrac{1}{2}-\epsilon\right)^{k-1}. \quad \square$$

Theorem 2 states that the first vector, b^1, in a reduced basis is an approximation to the shortest nonzero lattice point. In a similar fashion the other basis vectors approximate the *successive minima* of the lattice with respect to the distance function.

DEFINITION. $\lambda_1, \ldots, \lambda_n$ are the *successive minima* of the lattice with respect to F if there are lattice points $h^1, \ldots, h^n$, with $\lambda_i = F(h^i)$, such that for each $i = 1, \ldots, n$, h^i is the shortest lattice point which is linearly independent of $h^1, \ldots, h^{i-1}$.

An equivalent definition is $\lambda_i = \min\{\lambda | F(x) \leqslant \lambda$ contains i linearly independent lattice points$\}$. The successive minima λ_i are uniquely defined by the distance function F, but there may be more than one set of lattice points h^i, aside from $-h^i$, which realize these values. We have the following generalization of Theorem 2.

THEOREM 3. *Let $b^1, \ldots, b^n$ be a reduced basis. Then for $i = 1, \ldots, n$,*

$$F_i(b^i)\left(\tfrac{1}{2}-\epsilon\right)^{n-i} \leqslant \lambda_i \leqslant F_i(b^i)/\left(\tfrac{1}{2}-\epsilon\right)^{i-1}.$$

PROOF. There is no loss in generality in assuming that the basis $\{b^i\}$ is not only reduced but is also proper, in our sense that $F_j(b^i + \mu b^j) \geqslant F_j(b^i)$ for all $j < i$ and integral μ. This follows from the observation that if $\{b^i\}$ is replaced by a proper basis $\{c^i\}$ with $c^i = b^i + \sum_{j=1}^{i-1}\mu_{i,j}b^j$, the new basis is also reduced and, moreover, $F_i(c^i) = F_i(b^i)$.

We begin by showing that

$$F_1(b^i) \leqslant F_i(b^i)/\left(\tfrac{1}{2}-\epsilon\right)^{i-1}, \tag{3}$$

thereby demonstrating the right-hand side of the inequality in Theorem 3.

We first argue that

$$F_j(b^i) \leqslant F_i(b^i) + \tfrac{1}{2}\left[F_{i-1}(b^{i-1}) + \cdots + F_j(b^j)\right] \quad \text{for } j < i. \tag{4}$$

In the proof of Theorem 1, we have established the validity of (4) for $j = i-1$. Moreover, if (4) is correct for a given value of j, then the corresponding inequality holds for $j-1$, since

$$F_j(b^i) = \min_x F_{j-1}(b^i + xb^{j-1})$$
$$\geqslant \min_\mu F_{j-1}(b^i + \mu b^{j-1}) - \tfrac{1}{2}F_{j-1}(b^{j-1}) \quad \text{for integral } \mu$$
$$= F_{j-1}(b^i) - \tfrac{1}{2}F_{j-1}(b^{j-1}) \quad \text{since } \{b^i\} \text{ is proper.}$$

Estimating $\Sigma_1^{i-1} F_j(b^j)$ by means of Theorem 1, we see that

$$F_1(b^i) \leqslant F_i(b^i) \cdot \left\{1 + \tfrac{1}{2}\Sigma 1/\left(\tfrac{1}{2} - \epsilon\right)^{i-j}\right\}$$

$$\leqslant F_i(b^i)/\left(\tfrac{1}{2} - \epsilon\right)^{i-1}.$$

To demonstrate the inequalities on the left-hand side of Theorem 3, we write

$$h^1 = l_{11}b^1 + l_{12}b^2 + \cdots + l_{1n}b^n,$$

$$\vdots$$

$$h^i = l_{i1}b^1 + l_{i2}b^2 + \cdots + l_{in}b^n,$$

$$\vdots$$

$$h^n = l_{n1}b^1 + l_{n2}b^2 + \cdots + l_{nn}b^n,$$

with l_{ij} integral and with h^i linearly independent lattice points which realize the successive minima, i.e. $F_1(h^i) = \lambda_i$.

For each index i, there must be a pair of indices j and k with $j \leqslant i \leqslant k$ such that $l_{jk} \neq 0$, since otherwise

$$h^1 = l_{11}b^1 + \cdots + l_{1,i-1}b^{i-1},$$

$$h^i = l_{i1}b^1 + \cdots + l_{i,i-1}b^{i-1},$$

and the vectors $h^1, \ldots, h^i$ would be linearly dependent. For each i, therefore, let k be the largest index such that $l_{jk} \neq 0$ for some $j \leqslant i \leqslant k$. But then, since $|l_{jk}| \geqslant 1$,

$$\lambda_i \geqslant \lambda_j = F_1(h^j) \geqslant F_k(h^j) = |l_{jk}| \cdot F_k(b^k)$$

$$\geqslant F_i(b^i) \cdot \left(\tfrac{1}{2} - \epsilon\right)^{k-i} \geqslant F_i(b^i) \cdot \left(\tfrac{1}{2} - \epsilon\right)^{n-i}.$$

This demonstrates Theorem 3. □

If the basis is reduced and proper, inequalities relating λ_i and $F_1(b^i)$ can be obtained from (3) and the fact that $F_1(b^i) \geqslant F_i(b^i)$.

According to Minkowski, the successive minima satisfy the inequality $\lambda_1 \cdots \lambda_n \cdot \mathrm{vol}(C) \leqslant 2^n$. We can show that a proper reduced basis $b^1, \ldots, b^n$ approximates this result in the following sense:

THEOREM 4. *Let $b^1, \ldots, b^n$ be a proper reduced basis with respect to F. Then*

$$F_1(b^1) \cdot F_1(b^2) \cdots F_1(b^n) \cdot \mathrm{vol}(C) \leqslant 2^n/\left(\tfrac{1}{2} - \epsilon\right)^{n(n-1)/2}.$$

PROOF. The proof depends on the fact that for any basis

$$F_1(b^1) \cdot F_2(b^2) \cdots F_n(b^n) \cdot \mathrm{vol}(C) \leqslant 2^n. \tag{5}$$

To demonstrate (5), there is no loss in generality in taking $b^1, \ldots, b^n$ to be the standard basis of Z^n, since this can be achieved by a unimodular transformation

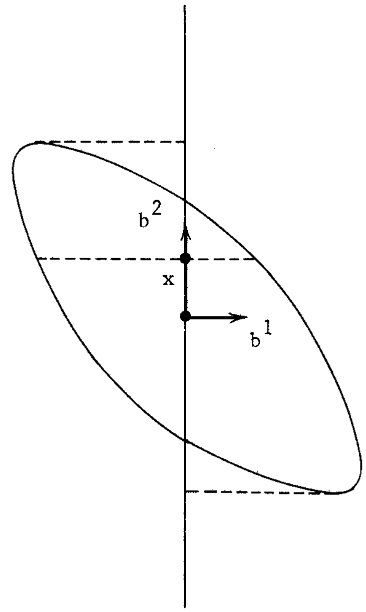

FIGURE 2

which does not change vol(C). Let us assume, by induction, that this inequality is satisfied for the $n-1$-dimensional body C_2 obtained by projecting b^1 into the affine space $\langle b^2, \dots, b^n \rangle$, so that

$$F_2(b^2) \cdots F_n(b^n) \cdot \mathrm{vol}(C_2) \leqslant 2^{n-1}.$$

But then $\mathrm{vol}(C) = \int_{C_2} l(x)\,dx$, and $l(x)$ the length of the intersection of the line $x + \alpha b^1$ with C. From the symmetry and convexity of C, $l(x) \leqslant l(0) = 2/F_1(b^1)$ so that $\mathrm{vol}(C) \leqslant 2\,\mathrm{vol}(C_2)/F_1(b^1)$, thereby demonstrating (5). Theorem 4 then follows from the previously established inequality (3):

$$F_1(b^i) \leqslant F_i(b^i)/\left(\tfrac{1}{2} - \epsilon\right)^{i-1}. \quad \square$$

If we are given a reduced basis, then a shortest vector h^1 can be calculated in polynomial time for fixed n. We do this by establishing bounds on the coordinates of lattice points satisfying $F_1(h) \leqslant F_1(b^1)$. Let $h = \Sigma l_j b^j$ be such a vector. Then

$$F_1(b^1) \geqslant F_1(h) \geqslant F_n(h) = |l_n| \cdot F_n(b^n)$$

so that

$$|l_n| \leqslant F_1(b^1)/F_n(b^n) \leqslant 1/\left(\tfrac{1}{2} - \epsilon\right)^{n-1}.$$

Now let us suppose that the coordinates $l_n, \dots, l_{i+1}$ have been selected. We find bounds for l_i as follows: Find the real α which minimizes $F_i(l_n b^n + \cdots + l_{i+1} b^{i+1} + \alpha b^i)$. If the minimum is greater than $F_1(b^1)$ then there is no h with these final $n-i$ coordinates satisfying $F_1(h) \leqslant F_1(b^1)$. If, on the other hand, the minimum is less than or equal to $F_1(b^1)$ then since

$$F_i\left(l_n b^n + \cdots + l_{i+1} b^{i+1} + l_i b^i\right) = F_i(h) \leqslant F_1(h) \leqslant F_1(b^1),$$

and F_i is symmetric and convex, we obtain $|l_i - \alpha| \cdot F_i(b^i) \leqslant 2F_1(b^1)$ and therefore $|l_i - \alpha| \leqslant 2/(\frac{1}{2} - \epsilon)^{i-1}$. This provides us with a tree of depth n and with a "small" number of branches at each node in which to search for the coordinates of the shortest vector. If the tree is used to calculate the shortest nonzero lattice point in actual numerical examples, the estimate $(\frac{1}{2} - \epsilon)^{i-1}$ should be replaced by $F_i(b^i)/F_1(b^1)$, which may be considerably smaller.

If we look for the ith successive minimum by considering those h with $F_1(h) \leqslant F_i(b^i)/(\frac{1}{2} - \epsilon)^{i-1}$ we obtain precisely the same set of inequalities for $l_n, \ldots, l_i$, but we do not have similar bounds for the first $i - 1$ coordinates of h. This yields a "small" number of hyperplanes of dimension $i - 1$, one of which contains a lattice point which realizes the ith successive minimum.

3. The basis reduction algorithm. An algorithm for finding a reduced basis ma easily be described. We begin with an initial basis $a^1, a^2, \ldots, a^n$ for the lattice, an move through a sequence of bases $b^1, b^2, \ldots, b^n$ according to the following rules: A each step of the algorithm, we consider the first index i for which one of th conditions

(1) $F_i(b^{i+1} + \mu b^i) \geqslant F_i(b^{i+1})$ for integral μ, and

(2) $F_i(b^{i+1}) \geqslant (1 - \epsilon)F_i(b^i)$

is not satisfied.

If the first condition is not satisfied, we replace b^{i+1} by $b^{i+1} + \mu b^i$, with μ the integer which minimizes $F_i(b^{i+1} + \mu b^i)$. If, after this replacement, the second condition obtains, we move to level $i + 1$. If the second condition is not satisfied, we interchange b^i and b^{i+1} and move to the preceding level $i - 1$, unless $i = 1$, in which case we remain at level 1.

In order to demonstrate convergence of the algorithm we consider the vector

$$F_1(b^1), \ldots, F_i(b^i), F_{i+1}(b^{i+1}), \ldots, F_n(b^n),$$

and remark that the maximum value of the components of the vector does not increase at any step of the basis reduction algorithm. If we replace b^{i+1} by $b^{i+1} + \mu b^i$, none of the terms change; if b^i and b^{i+1} are interchanged, $F_i(b^i)$ becomes $F_i(b^{i+1}) \leqslant (1 - \epsilon)F_i(b^i)$ and $F_{i+1}(b^{i+1})$ is replaced by

$$\begin{aligned} &\min F\left(b^i + \alpha_1 b^1 + \cdots + \alpha_{i-1} b^{i-1} + \alpha_{i+1} b^{i+1}\right) \\ &\qquad \leqslant \min F\left(b^i + \alpha_1 b^1 + \cdots + \alpha_{i-1} b^{i-1}\right) = F_i(b^i). \end{aligned}$$

It follows that at any step in the algorithm, $\max F_i(b^i) \leqslant \max F_i(a^i)$ equal to, say, U.

The basis reduction algorithm is known to converge in polynomial time, including the number of variables, n, for $F(x) = |x|$, and a general lattice given by an integer basis. The argument is based on two observations: first, that an interchange between b^i and b^{i+1} preserves the value of $F_j(b^j)$ for all indices other than i and $i + 1$, and secondly, that for $F(x) = |x|$, the product $F_i(b^i)F_{i+1}(b^{i+1})$ is constant when the vectors b^i and b^{i+1} are exchanged. This permits us to deduce that $D(b^1, \ldots, b^n) = \Pi(F_i(b^i))^{n-i}$ decreases by a factor of $(1 - \epsilon)$ at each interchange. It is easy to show that $\Pi(F_i(b^i))^{n-i} \geqslant 1$, for any basis, from which the polynomial convergence follows rèadily.

Constancy of $F_i(b^i)F_{i+1}(b^{i+1})$ is not valid for a general distance function, and the generalized basis reduction algorithm is not known to execute in polynomial time in the number of variables n. But the algorithm may be shown to be polynomial in the data of the problem for *fixed n*. We present two arguments for this conclusion, both of which depend on establishing lower bounds for the possible values assumed by $F_i(b^i)$ during the course of the algorithm.

To obtain such a lower bound, assume that $C \subset B(R)$, the ball of radius R. Then $F(x) \geqslant |x|/R$. Now let $b^1, \ldots, b^n$ be any basis for the lattice satisfying $F_i(b^i) \leqslant U$, and let $c^i = b^i + \sum_{j=1}^{i-1} \mu_{i,j} b^j$ be a proper basis associated with $\{b^i\}$, satisfying

$$F_i(c^i) = F_i(b^i) \quad \text{and} \quad F_1(c^i) \leqslant F_i(b^i) + \tfrac{1}{2} \sum_1^{i-1} F_j(b^j) \leqslant nU.$$

We have, therefore, $|c^i| \leqslant nUR$.

We estimate $F_i(b^i)$, from below, as follows:

$$\begin{aligned} F_i(b^i) &= \min F\left(b^i + \alpha_1 b^1 + \cdots + \alpha_{i-1} b^{i-1}\right) \\ &= \min F\left(b^i + \alpha_1 c^1 + \cdots + \alpha_{i-1} c^{i-1}\right) \\ &\geqslant \min\left|\left(b^i + \alpha_1 c^1 + \cdots + \alpha_{i-1} c^{i-1}\right)\right|/R. \end{aligned}$$

But $\min|(b^i + \alpha_1 c^1 + \cdots + \alpha_{i-1} c^{i-1})|$ is the distance between the vector b^i and the space $\langle c^1, \ldots, c^{i-1} \rangle$ and is therefore equal to

$$\left[G(c^1, \ldots, c^{i-1}, b^i)/G(c^1, \ldots, c^{i-1})\right]^{1/2},$$

where the Grammian

$$G(x^1, \ldots, x^i) = \det\left[(x^j, x^k)\right]_{j,k=1}^i.$$

Since $c^1, \ldots, c^{i-1}$ and b^i are integral, $G(c^1, \ldots, c^{i-1}, b^i) \geqslant 1$. Moreover,

$$G(c^1, \ldots, c^{i-1})^{1/2} \leqslant |c^1| \cdots |c^{i-1}| \leqslant (nUR)^{i-1}.$$

It follows that

$$F_i(b^i) \geqslant 1/\left[R(nRU)^{i-1}\right] \geqslant 1/\left[R(nRU)^{n-1}\right]$$

equal to, say, V.

We have already shown that each component of

$$F_1(b^1), \ldots, F_i(b^i), F_{i+1}(b^{i+1}), \ldots, F_n(b^n)$$

is bounded above by $U = \max F_i(a^i)$ throughout the course of the algorithm. Moreover, the first term in the sequence to change at any iteration decreases by a factor of $(1 - \epsilon)$. Our first argument for polynomial convergence is to observe that the maximal number of interchanges is therefore

$$\left[\log(U/V)/\log(1/(1-\epsilon))\right]^n.$$

(Simply record the times at which the first two basis vectors b^1 and b^2 are interchanged. Between any consecutive pair of such times we are faced with an identical problem with $n - 1$ variables.) Using our lower bound V we see that the number of interchanges of the basis reduction algorithm is bounded above by

$$[n \log(nUR)/\log(1/(1-\epsilon))]^n. \tag{6}$$

It may be useful to remark that if the basis reduction algorithm is executed using a *proper* basis $\{c^i\}$ associated with $\{b^i\}$, then the size of the vectors c^i will remain bounded from above by $|c^i| \leqslant nUR$. In practice, it may be sufficient to calculate this proper basis at periodic intervals.

The second argument for polynomiality, which achieves a different bound, depends on the observation that for a general distance function $F(x)$, the product $F_i(b^i)F_{i+1}(b^{i+1})$ increases by a factor less than or equal to 2 after an interchange of b^i and b^{i+1}. The argument makes use of the following theorem.

THEOREM 5. *Let S be a compact convex set in R^k, which is symmetric about the origin, and let x and y be two linearly independent vectors on the boundary of S. Define*

$$d_x = \max\{\alpha | \alpha x + \beta y \in S \text{ for some } \beta\} \quad \text{and}$$

$$d_y = \max\{\beta | \alpha x + \beta y \in S \text{ for some } \alpha\}.$$

Then $\frac{1}{2} \leqslant d_x/d_y \leqslant 2$.

PROOF. If $d_x x + \beta y \in S$, then either $\beta \geqslant d_x - 1$, or $\beta \leqslant 1 - d_x$. For if $0 \leqslant \beta < d_x - 1$, x is a strict convex combination of 0, $-y$, $d_x x + \beta y$, and is therefore interior to S; if $1 - d_x < \beta \leqslant 0$, x is a strict convex combination of 0, y, $d_x x + \beta y$ and is again interior to S. In the first case, $d_y \geqslant d_x - 1$. In the second case, since S is symmetric, the vector $-(d_x x + \beta y) \in S$ and again $d_y \geqslant d_x - 1$. It follows that

$$d_x/d_y \leqslant 1 + 1/d_y \leqslant 2,$$

since $d_y \geqslant 1$. The lower inequality follows from interchanging x and y. □

Theorem 5 may be used to show that the product $F_i(b^i)F_{i+1}(b^{i+1})$ increases by a factor not larger than 2 at any step of the basis reduction algorithm in which b^i and b^{i+1} are interchanged. Let $S = C_i \subset E_i = \langle b^i, \ldots, b^n \rangle$. Assume, without loss of generality, that $F_i(b^i) = 1$, and take $y = b^i$ and $x = b^{i+1}/F_i(b^{i+1})$, both of which are on the boundary of C_i. But then $F_{i+1}(x) = 1/d_x$, and $F^{\#}_{i+1}(b^i) = 1/d_y$, with $F^{\#}_{i+1}$ the distance function associated with the projection of C into $\langle b^i, b^{i+2}, \ldots, b^n \rangle$. It

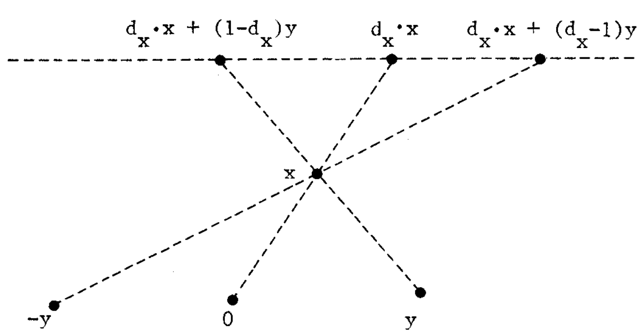

FIGURE 3

follows that

$$(7)\qquad F_i(b^{i+1})F^{\#}_{i+1}(b^i)/F_i(b^i)F_{i+1}(b^{i+1})$$
$$= \left[F_i(b^{i+1})/d_y\right]/\left[F_i(b^{i+1})/d_x\right] = d_x/d_y \leqslant 2.$$

Now consider

$$D(b^1,\dots,b^n) = \Pi\left(F_i(b^i)\right)^{\gamma^{n-i}},$$

with $\gamma = 2 + 1/\log(1/(1-\epsilon))$. It is a straightforward computation to show that our estimate (7) and the inequality $F_i(b^{i+1}) < (1-\epsilon)F_i(b^i)$ imply that $D(b^1,\dots,b^n)$ decreases by a factor of at least $(1-\epsilon)$ at each interchange required by the basis reduction algorithm. Since $V \leqslant F_i(b^i) \leqslant U$ at each step of the algorithm, we see that the number of interchanges is bounded above by

$$[(\gamma^n - 1)/(\gamma - 1)]\log(U/V)/\log(1/(1-\epsilon))$$
$$\leqslant [(\gamma^n - 1)/(\gamma - 1)]n\log(nUR)/\log(1/(1-\epsilon)),$$

an estimate which is much better than our previous estimate in terms of its dependence of UR. The preceding discussion has established the following theorem:

THEOREM 6. *The basis reduction algorithm terminates in a polynomial number of steps, for fixed n.*

Since the number of possible values of the vector

$$F_1(b^1),\dots,F_i(b^i),F_{i+1}(b^{i+1}),\dots,F_n(b^n)$$

is finite, the basis reduction algorithm executes in finite time even when $\epsilon = 0$. Bárány has recently demonstrated geometric convergence when $\epsilon = 0$ for the case of two variables. Consider two successive steps of the algorithm. Assume that the initial basis is (b^1, b^2), with b^2 the smallest lattice point on the line $b^2 + \alpha b^1$, and that $\delta_1 F(b^1) < F(b^2) < F(b^1)$. After the first interchange the basis is given by (b^2, b^1). Let μ^* minimize $F(b^1 + \mu b^2)$ for integral μ and assume that $\delta_2 F(b^2) < F(b^1 + \mu^* b^2) < F(b^2)$ so that another interchange is required leading to the basis $(b^1 + \mu^* b^2, b^2)$. Finally, let μ be the integer which minimizes $F(b^2 + \mu(b^1 + \mu^* b^2))$.

THEOREM 7 (BÁRÁNY). *If $\delta_1\delta_2 > \frac{1}{2}$ then $\mu = 0$ or 1. In either case the basis $(b^1 + \mu^* b^2, b^2 + \mu(b^1 + \mu^* b^2))$ is reduced.*

PROOF. We argue, first of all, that $|\mu^*| > 1$. If $\mu^* = 0$, there is a contradiction between $F(b^1 + \mu^* b^2) < F(b^2)$ and $F(b^2) < F(b^1)$. If $\mu^* = 1$, then b^2 is not the shortest integral vector on the line $b^2 + \alpha b^1$, and similarly for $\mu^* = -1$. To be specific, let us now assume that $\mu^* \leqslant -2$.

Consider the convex function $g(\alpha) = F(b^2 + \alpha(b^1 + \mu^* b^2))$. We have $g(0) = F(b^2)$ and from our assumptions $g(0) > \delta_1 F(b^1) > \delta_2\delta_1 F(b^1)$. Also,

$$g(1) = F(b^1 + (\mu^* + 1)b^2) \geqslant F(b^1 + \mu^* b^2) > \delta_2 F(b^2) > \delta_2\delta_1 F(b^1).$$

But

$$g(-1/\mu^*) = (1/|\mu^*|)F(b^1) \leqslant \tfrac{1}{2}F(b^1).$$

It follows, from the convexity of $g(\alpha)$, that if $\delta_1\delta_2 > \frac{1}{2}$, the integral minimum of $g(\alpha)$ is at $\alpha = 0$ or $\alpha = 1$. In the first case, the basis $(b^1 + \mu^* b^2, b^2)$ is reduced since $F(b^1 + \mu^* b^2) < F(b^2)$; in the second case, the basis $(b^1 + \mu^* b^2, b^1 + (\mu^* + 1)b^2)$ is reduced because $F(b^1 + \mu^* b^2) \leqslant F(b^1 + (\mu^* + 1)b^2)$. □

Theorem 7 implies that in $2p$ steps of the basis reduction algorithm, $F(b^1)$ will decrease by a factor of at least $(1/2)^p$. Since $F(h) \geqslant 1/R$ for any lattice point h, we have geometric convergence of the algorithm for $n = 2$ and $\epsilon = 0$. No argument is currently available for higher dimensions, and $\epsilon = 0$, unless we revise the order in which the steps of the algorithm are executed. For example, following a suggestion made by Bárány, let us assume that we always select the *largest* index i for which one of the conditions of a reduced basis is not satisfied. It follows that if we ever return to level 1, the basis $b^2, \dots, b^n$ is reduced with $\epsilon = 0$ for the $n - 1$ dimensional problem defined by C_2. If an interchange of b^1 and b^2 is then required, two possible cases arise:

(1) $F_1(b^2) \geqslant (1 - \delta)F_1(b^1)$ for some fixed $\frac{1}{2} < \delta < 1$. But then the basis $b^1, \dots, b^n$ will be δ-reduced for the original problem. Our previous analysis shows that there is a finite number, $N(n, \delta)$ of lattice points h, such that $F_1(h) < F_1(b^1)$, and, therefore, the algorithm requires an exchange of b^1 and b^2 not more than $N(n, \delta)$ times.

(2) At each return to level 1, we have $F_1(b^2) < (1 - \delta)F_1(b^1)$, and therefore the number of returns to level 1 is bounded above by

$$\log(U/V)/\log(1/(1 - \delta)).$$

We then use an inductive argument on n to achieve polynomial bounds on the running time of the algorithm for $\epsilon = 0$ and fixed n.

4. The Korkine-Zolotarev basis. A special basis for a lattice, the Korkine-Zolotarev basis, has been used very successfully by Lagarias, Lenstra and Schnorr (1986) to improve some classical estimates in the geometry of numbers relating the successive minima of a body C and its dual body $C^* = \{y \mid y \cdot x \leqslant 1 \text{ for all } x \in C\}$. In their analysis they approximate a general body by an ellipsoid, transform the ellipsoid to a sphere by a linear transformation and use specific properties of the spherical norm. We shall illustrate, by means of a few examples, that their arguments can be applied, virtually unchanged, to a general body without the prior step of an ellipsoidal approximation.

Let $b^1, b^2, \dots, b^n$ be defined recursively as follows: given $b^1, \dots, b^{i-1}, b^i$ minimizes $F_i(h)$ over all lattice points which are linearly independent of $b^1, \dots, b^{i-1}$. The vectors $b^1, b^2, \dots, b^n$ clearly form a basis, since otherwise there is an integer vector which can be written as a linear combination of the b^i with *some* fractional coefficients. But then by adding and subtracting suitable integral multiples of $\{b^j\}$, we obtain an integral vector $h = \alpha_1 b^1 + \cdots + \alpha_i b^i$, with α_i a proper fraction; h is independent of $b^1, \dots, b^{i-1}$ and gives a smaller value of $F_i(h)$ than does b^i.

A Korkine-Zolotarev basis is defined to be any *proper* basis $\{c^i\}$ associated with $\{b^i\}$. The basis satisfies the inequalities

$$F_1(c^i) \leqslant F_i(b^i) + \tfrac{1}{2}\sum_{1}^{i-1} F_j(b^j). \tag{8}$$

The Korkine-Zolotarev basis may not be unique; there may be several integral vectors independent of $b^1, \dots, b^{i-1}$ which minimize $F_i(h)$, and the integers $\mu_{i,j}$ defining a proper basis associated with $\{b^i\}$ need not be unique.

THEOREM 8. *Let $c^1, \ldots, c^n$ be a Korkine-Zolotarev basis. Then*

$$F_1(c^i)/((i+1)/2) \leqslant \lambda_i \leqslant ((i+1)/2) F_1(c^i).$$

PROOF. Let $h^1, \ldots, h^n$ realize the successive minima. For each i, at least one of the vectors $h^1, \ldots, h^i$ must be independent of $b^1, \ldots, b^{i-1}$, since otherwise the vectors would all lie in $\langle b^1, \ldots, b^{i-1} \rangle$ and be linearly dependent. It follows that $\max_{j \leqslant i} F_i(h^j) \geqslant F_i(b^i)$ and therefore

$$\lambda_i = F_1(h^i) = \max_{j \leqslant i} F_1(h^j) \geqslant F_i(b^i).$$

But then (8) implies that

$$F_1(c^i) \leqslant \lambda_i + \tfrac{1}{2}(\lambda_1 + \cdots + \lambda_{i-1}) \leqslant ((i+1)/2)\lambda_i.$$

This demonstrates the left-hand inequality of Theorem 8.

To obtain the right-hand side, notice that for

$$k \leqslant i,\ F_k(b^k) \leqslant F_k(c^i) \leqslant F_1(c^i), \text{ since } c^i \text{ is independent of } b^1, \ldots, b^{i-1}.$$

But

$$\lambda_i \leqslant \max_{j \leqslant i} F_1(c^j) \leqslant \max_{j \leqslant i} \left\{ F_j(b^j) + \tfrac{1}{2} \sum_{k \leqslant j-1} F_k(b^k) \right\}$$

$$\leqslant F_1(c^i) \max_{j \leqslant i} \left\{ 1 + \tfrac{1}{2} \sum_{k \leqslant j-1} 1 \right\} = ((i+1)/2) F_1(c^i). \quad \square$$

We remark that Theorem 8, in conjunction with Minkowski's inequality, implies that a Korkine-Zolotarev basis satisfies

$$F_1(c^1) \cdot F_1(c^2) \cdots F_1(c^n) \cdot \mathrm{vol}(C) \leq (n+1)!,$$

an improvement over the estimate of Theorem 4.

Let λ_1^* be the length of the shortest nonzero lattice point with respect to the dual body $C^* = \{y \mid y \cdot x \leqslant 1 \text{ for all } x \in C\}$. Minkowski's first theorem implies that $\lambda_1 \leqslant 2/(\mathrm{vol}(C))^{1/n}$ and $\lambda_1^* \leqslant 2/(\mathrm{vol}(C^*))^{1/n}$, so that an upper bound for $\lambda_1 \lambda_1^*$ may be obtained from a lower bound for the product of the volumes $\mathrm{vol}(C) \cdot \mathrm{vol}(C^*)$. A well-known ellipsoidal approximation to C is sufficient to produce the inequality $\lambda_1 \lambda_1^* \leqslant n^{3/2}$. A more sophisticated lower bound, given by Bourgain and Milman (1985), implies that there exists a universal constant c_0, such that $\lambda_1 \lambda_1^* \leqslant c_0 n$. This result is used to demonstrate the following property of a Korkine-Zolotarev basis.

THEOREM 9. *There is a universal constant c_0 such that for a Korkine-Zolotarev basis, $F_i(c^i)\lambda_1^* \leqslant c_0(n - i + 1)$.*

PROOF. We assume, without loss of generality, that the Korkine-Zolotarev basis consists of the n-units vectors $e^1, \ldots, e^n$, and let C_i be the projection of C into $e^i, \ldots, e^n$, with associated distance function F_i. The projection of the original lattice is the set of all $(x_i, \ldots, x_n)$ with integral coordinates. For this lattice and distance function, $\lambda_1 = F_i(c^i)$.

For the previous discussion, there is a nonzero lattice point $h^i = (h_i, \ldots, h_n)$ such that

$$F_i(c^i) \cdot \max\{h^i \cdot x | x \in C_i\} \leqslant c_0(n - i + 1).$$

But this linear function $h^i \cdot x$ may be extended to a linear function $h \cdot x$ in R^n by adding $i - 1$ zero coordinates to h^i, so that

$$F_i(c^i)\lambda_1^* \leqslant F_i(c^i) \cdot \max\{h \cdot x | x \in C\} \leqslant c_0(n - i + 1). \quad \square$$

Theorem 9 has an important application to the study of lattice-free bodies K which are not symmetric about the origin. As we shall see, any such body has associated with it a nonzero lattice point h such that

$$\max_{x \in K}\{h \cdot x\} - \min_{x \in K}\{h \cdot x\} \leqslant c_0 n(n + 1)/2.$$

The argument is based on Theorem 9, and a subsequent result in the paper by Kannan and Lovász which may be described as follows.

THEOREM 10. *Let $C = (K - K)$, with K a convex body, and let F be the distance function associated with C. For any basis $b^1, \ldots, b^n$, define $\rho = \Sigma F_i(b^i)$. Then the lattice translates of ρK cover R^n.*

PROOF. We show, by induction on n, that for any $x \in R^n$, there is a lattice point h with $x + h \in \rho K$. Notice that the hypotheses and conclusion of the theorem are unchanged if we replace K by any translate of itself; we may therefore assume that K has been translated so that both 0 and b^1 are contained in $F_1(b^1)K$. Let K' be the projection of K along the vector b^1 into $\langle b^2, \ldots, b^n \rangle$ and x' the corresponding projection of x.

By the induction assumption, there is a lattice point h' in $\langle b^2, \ldots, b^n \rangle$ such that $x' + h' \in \Sigma_2^n F_i(b^i)K'$ and therefore $x + \alpha b^i + h' \in \Sigma_2^n F_i(b^i)K$ for some α. It follows that

$$x + \lceil \alpha \rceil b^i + h' \in \sum_2^n F_i(b^i)K + (\lceil \alpha \rceil - \alpha)b^i \subseteq \Sigma F_i(b^i)K. \quad \square$$

If the body K is free of lattice points, then its lattice translates do not cover the origin, and therefore $\rho = \Sigma F_i(b^i) > 1$. We see from Theorem 9, that for such a body, $\lambda_1^* < c_0 \Sigma(n - i + 1) = c_0 n(n + 1)/2$. It should be remarked that the inductive argument provides an algorithm for calculating a lattice point in K, if $\rho \leqslant 1$. The reader may notice that this observation, combined with the fact that the shortest nonzero lattice point may be calculated in polynomial time for fixed n, leads to an algorithm for integer programming, which requires no ellipsoidal approximations, and which executes in polynomial time for a fixed number of variables.

Acknowledgements. The research of the first author was supported by the Hungarian National Foundation for Scientific Research, and of the second author by the Program on Discrete Mathematics at the Cowles Foundation and NSF grant SES-8807167.

We wish to thank David Applegate, Imre Bárány, Judith Bender, Bill Cook, Mark Hartman, Ravi Kannan, Jeffrey Lagarias, Tom Rutherford, David Shallcross and Mike Todd for a number of stimulating conversations on the subject of this paper. We are also grateful for the comments of the anonymous referee.

References

Bárány, I., private communication.

Bourgain, J. and Milman, V. D. (1985). Sections Euclidiennes et Volume des Corps Symétriques Convex dans R^n. *C. R. Acad. Sci. Paris*, **300** 435–437.

Kannan, R. and Lovász, L. (1988). Covering Minima and Lattice-Point-Free Convex Bodies. *Ann. of Math.* **128** 577–602.

Lagarias, J., Lenstra, H. W. Jr. and Schnorr, C. P. (1986). *Korkine-Zolotarev Bases and Successive Minima of a Lattice and its Reciprocal Lattice*, manuscript.

Lenstra, A. K., Lenstra, H. W. Jr. and Lovász, L. (1982). Factoring Polynomials with Rational Coefficients, *Math. Ann.* **261** 513–534.

Lenstra, H. W. Jr. (1983). *Integer Programming with a Fixed Number of Variables. Math. Oper. Res.* **8** 538–548.

Lovász, L. (1988). *Geometry of Numbers and Integer Programming*. Manuscript prepared for the 13th Annual Symposium on Mathematical Programming.

LOVÁSZ: DEPARTMENT OF COMPUTER SCIENCE, EÖTVÖS LORÁND UNIVERSITY, BUDAPEST, HUNGARY AND DEPARTMENT OF COMPUTER SCIENCE, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544

SCARF: THE COWLES FOUNDATION, YALE UNIVERSITY, NEW HAVEN, CONNECTICUT 06520