COMBINATORICA 6 (1) (1986) 1—13

ON LOVÁSZ' LATTICE REDUCTION AND THE NEAREST LATTICE POINT PROBLEM

L. BABAI

Received 11 June 1984
Revised 20 August 1985

Answering a question of Vera Sós, we show how Lovász' lattice reduction can be used to find a point of a given lattice, nearest within a factor of c^d ($c=$ const.) to a given point in $\mathbf{R}^d$. We prove that each of two straightforward fast heuristic procedures achieves this goal when applied to a lattice given by a Lovász-reduced basis. The verification of one of them requires proving a geometric feature of Lovász-reduced bases: a c_1^d lower bound on the angle between any member of the basis and the hyperplane generated by the other members, where $c_1=\sqrt{2}/3$.

As an application, we obtain a solution to the nonhomogeneous simultaneous diophantine approximation problem, optimal within a factor of C^d.

In another application, we improve the Grötschel—Lovász—Schrijver version of H. W. Lenstra's integer linear programming algorithm.

The algorithms, when applied to rational input vectors, run in polynomial time.

1. Introduction

A *lattice* in $\mathbf{R}^d$, defined by the basis $B=\{b_1, \ldots, b_d\}$ of $\mathbf{R}^d$, is the set $L=\sum_{i=1}^{d} \mathbf{Z}b_i$ of all integral linear combinations of B. Finding the shortest non-zero vector in L is a fundamental algorithmic problem, and lies at the heart of the solution of many diophantine problems in arithmetics, including integer programming (H. W. Lenstra, Jr. [12]), finding irreducible factors of polynomials (A. K. Lenstra [10]), minimal polynomials of algebraic numbers (Kannan, Lenstra, Lovász [8]) and simultaneous diophantine approximation in the first place (Lovász, see [11]).

Although the shortest vector problem may be NP-hard for integral input vectors (it is known to be NP-hard with respect to maximum norm, P. van Emde Boas [3]), a vector at most C^d times the shortest one suffices for most applications. These applications include those mentioned above as well as applications to the ellipsoid method in linear programming (Lovász [13]; cf. [5]), recent attacks on knapsack-based crypto-systems (Adleman [1], Shamir [15], Lagarias and Odlyzko [9]), and the disproof of Mertens' century-old conjecture in number theory (Odlyzko and te Riele,

AMS subject classification (1980): 68 C 25, 10 F 10, 10 F 15, 90 C 10

[14]). All these applications were made possible by *Lovász' lattice reduction* algorithm (see [11]), originally designed to give nearly optimal simultaneous diophantine approximation which, in turn, arose, as far as Lovász was concerned, from the need to eliminate the annoying full-dimensionality condition from the ellipsoid method in linear programming ([13], see Grötschel, Lovász, Schrijver [5]). Odlyzko reports that Lovász' algorithm performs substantially better in practice than predicted by the C^d theoretical worst-case bound. This observation was crucial for the number theoretic application [14].

Diophantine problems usually come in homogeneous and nonhomogeneous versions, and usually both have similar answers but the nonhomogeneous cases are more difficult to handle (cf. Cassels [2]).

In the case of the short lattice vector problem (a *homogeneous* approximation problem: we approximate zero), the corresponding *nonhomogeneous* problem is to find *the nearest lattice point to a given point in* $\mathbf{R}^d$. This problem is known to be NP-hard even in the Euclidean case (P. van Emde Boas [3]). However, as we shall see, a lattice point within C^d times the distance from the nearest one can be found efficiently (in polynomial time if the basis vectors have rational coordinates). Here, C is an absolute constant, and d is the dimension. We prove that each of two trivial heuristic procedures (Section 3) achieve this goal if we start from a Lovász-reduced basis.

The most important and immediate application of Lovász' lattice reduction algorithm was his (homogeneous) diophantine approximation algorithm, previously solved only in dimension one by the classical method of continued fractions. Vera Sós [17] gave a method, based on continued fractions, to solve the one-dimensional nonhomogeneous case optimally. We shall show how the approximate nearest lattice point procedure leads to nearly optimal nonhomogeneous simultaneous diophantine approximation (Section 7).

In Section 8, we improve the Grötschel—Lovász—Schrijver version of H. W. Lenstra's integer linear programming algorithm.

We note that R. Kannan [7] considered some of the problems discussed here (cf. Sect. 8 of this paper). He solved the nearest lattice point problem in d^{cd} arithmetic operations. He also showed that a shortest vector oracle can be used to find, in polynomial time, a lattice point nearest within a factor d to a given point in d-space.

Let me remark that I don't see any a priori reason why the nonhomogeneous approximation problem could not actually be *easier* than the homogeneous one.

Problem. Suppose we are given an oracle which solves the nearest lattice point problem within a constant factor, i.e., on an input (L, x) $(x \in \mathbf{R}^d$, L a lattice in $\mathbf{R}^d)$ the oracle outputs $w \in L$ such that $|w-x| \leqq C|u-x|$ for any $u \in L$. Can such an oracle be used to solve, in polinomial time, the shortest vector problem within a factor of $\exp(d^{1-\varepsilon})$ for some fixed $\varepsilon > 0$? (That is, on an input L, a lattice in $\mathbf{R}^d$, output a nonzero vector $w \in L$ such that $|w| \leqq |u| \exp(d^{1-\varepsilon})$ for any $u \in L$, $u \neq 0$.)

Acknowledgements. I am indebted to Vera Sós for drawing my attention to the importance of nonhomogeneous diophantine approximation and the approximately nearest lattice point problems. I would like to thank László Lovász for telling me about his lattice reduction algorithm and its numerous consequences, immediately after he had made the discovery at the end of 1981. Excellent seminar lectures by Éva

Tardos in Budapest helped me understand the subject, and I am indeed thankful to her.

Last but not least, I should like to thank Péter Frankl for suggesting an improvement of the error estimate in Theorem 3.1 and György Turán for pointing out an error in Section 7.

2. Lovász' lattice reduction

Let $b_1, \ldots, b_d \in \mathbf{R}^d$ be linearly independent vectors and $L = \sum_{i=1}^{d} \mathbf{Z} b_i$ the lattice generated by $B = \{b_1, \ldots, b_d\}$. Let $b_1^*, \ldots, b_d^*$ denote the orthogonalized sequence corresponding to $b_1, \ldots, b_d$. We thus have

$$b_i = \sum_{j=1}^{d} \mu_{ij} b_j^* = b_i^* + \sum_{1 \leqq j < i} \mu_{ij} b_j^* \tag{2.1}$$

where

$$\mu_{ij} = 0 \quad \text{for} \quad 1 \leqq i < j \leqq d, \tag{2.2}$$

and

$$\mu_{ii} = 1 \quad (i = 1, \ldots, d). \tag{2.3}$$

B is a *Lovász-reduced basis* of L if the following two conditions* hold:

$$|\mu_{ij}| \leqq 1/2 \quad \text{for} \quad 1 \leqq j < i \leqq d. \tag{2.4}$$

$$|b_i^*| \geqq \frac{|b_{i-1}^*|}{\sqrt{2}} \quad \text{for} \quad i = 2, \ldots, d. \tag{2.5}$$

($|x|$ denotes the Euclidean norm of $x \in \mathbf{R}^d$). For any lattice L, Lovász' algorithm [11] reaches a reduced basis in a finite number of steps. If the initial basis vectors have rational coordinates, then the reduced basis is reached in polynomial time.

Our algorithms will use a Lovász-reduced basis as input. In place of (2.4) and (2.5), it would be sufficient for our purposes to have any constant bound on the quantities $|\mu_{ij}|$ and $|b_{i-1}^*|/|b_i^*|$.

3. A nearby lattice point: two algorithms

Let L be a lattice in $\mathbf{R}^d$, given by a basis $B = \{b_1, \ldots, b_d\}$ and let $x \in \mathbf{R}^d$. Let u be the nearest neighbor of x in L.

NEARBY LATTICE POINT PROBLEM. *Given B and x, find $w \in L$ such that $|x-w| \leqq C_d |x-u|$, where C_d, the measure of approximation, is a function of d.*

* Actually, in place of (2.5) Lovász requires the insignificantly stronger inequality

$$|b_i^* + \mu_{i,i-1} b_{i-1}^*| \geqq \frac{\sqrt{3}}{2} |b_{i-1}^*|.$$

We give two straightforward heuristic procedures.

First procedure: ROUNDING OFF. Let $x=\sum_{i=1}^{d}\beta_i b_i$ and let α_i be the integer nearest to β_i. Set $w=\sum_{i=1}^{d}\alpha_i b_i$.

Second procedure: NEAREST PLANE. Let $U=\sum_{i=1}^{d-1}\mathbf{R}b_i$ be the linear subspace generated by $b_1, \ldots, b_{d-1}$ and let $L'=\sum_{i=1}^{d-1}\mathbf{Z}b_i$ be the corresponding sublattice of L.

Find $v\in L$ such that the distance between x and the affine subspace $U+v$ be minimal. Let x' denote the orthogonal projection of x on $U+v$. Recursively, find $y\in L'$ near $x'-v$. Let $w=y+v$.

Comment. In order to find v and x', we proceed as follows. Write x as a linear combination of the orthogonalized basis: $x=\sum_{i=1}^{d}\gamma_i b_i^*$. Let δ be the integer nearest to γ_d. Then $x'=\sum_{i=1}^{d-1}\gamma_i b_i^*+\delta b_d^*$, and $v=\delta b_d$.

Theorem 3.1. *If the basis B is Lovász-reduced, then the NEAREST PLANE procedure finds a lattice point w, nearest to x within a factor of $C_d=2^{d/2}$. Moreover, $|x-w|<2^{d/2-1}|b_d^*|$.*

For comparison, note that Lovász' algorithm [11] produces a nonzero lattice vector, shortest within a factor of $2^{(d-1)/2}$.

Perhaps more surprisingly, already the ROUNDING OFF procedure succeeds within a factor of C^d. This may be one more indication of the power of Lovász' lattice reduction. At the same time, this result answers Lovász' question [13], whether *an approximating lattice point can be found among the vertices of the parallelopiped cell containing x.* The answer is affirmative for approximations within a factor C_d.

Theorem 3.2. *If the basis B is Lovász-reduced, then the ROUNDING OFF procedure finds a lattice point w, nearest to x within a factor of $C_d'=1+2d(9/2)^{d/2}$.*

The proof of Theorem 3.1 is easy (Section 4). The proof of Theorem 3.2 (Section 6) requires a geometric result on the shape of Lovász-reduced parallelopipeds (Theorem 5.1).

On rational input vectors, both procedures clearly run in polynomial time.

We remark that C. P. Schnorr [16] has modified Lovász' algorithm to provide a better approximation for the shortest vector. His version finds, in polynomial time, a vector at most $(1+\varepsilon)^d$ times as long as the shortest one for any fixed ε. (The exponent in the running time depends on ε.) It is natural to ask if a similar improvement of the nearest lattice point algorithms is possible.

4. Nearest plane: error estimate

In this section, we prove Theorem 3.1.

For $d=1$ we find the nearest lattice point.

For $d\geqq 2$ observe that

$$|x-x'| \leqq |b_d^*|/2 \tag{4.1}$$

and

$$|x-x'| \leqq |x-u|, \tag{4.2}$$

because the translates $U+z\,(z\in L)$ are spaced at distance $|b_d^*|$ and $|x-x'|$ is the distance of x from the nearest such subplane.

From (4.1) we obtain by induction that

$$|x-w|^2 \leqq (|b_1^*|^2+\ldots+|b_d^*|^2)/4. \tag{4.3}$$

By (2.5), the right hand side is not greater than

$$|b_d^*|^2(2^{d-1}+2^{d-2}+\ldots+1)/4=|b_d^*|^2(2^d-1)/4,$$

hence

$$|x-w| < 2^{(d/2)-1}|b_d^*| \tag{4.4}$$

proving the second statement in Theorem 3.1.

We have to consider two cases.

Case (a). If $u\in U+v$ then clearly $u-v$ is the nearest lattice point to $x'-v$ in L' and therefore

$$|x'-w| = |x'-v-y| \leqq C_{d-1}|x'-u| \leqq C_{d-1}|x-u|. \tag{4.5}$$

Consequently, by (4.2) and (4.5),

$$|x-w| = (|x-x'|^2+|x'-w|^2)^{1/2} \leqq |x-u|(1+C_{d-1}^2)^{1/2} < C_d|x-u|.$$

Case (b). If $u\notin U+v$, then

$$|x-u| \geqq \frac{1}{2}|b_d^*|. \tag{4.6}$$

Comparing this inequality with (4.4) we obtain $|x-w|<2^{d/2}|x-u|$. ∎

5. On the shape of Lovász-reduced parallelopipeds

The following result, needed for the proof of Theorem 3.2, may be of independent interest.

Theorem 5.1. *Let $B=\{b_1, \ldots, b_d\}$ be a Lovász-reduced basis. Let θ_k denote the angle between b_k and the linear subspace $U_k=\sum_{j\neq k}\mathbf{R}b_j$. Then, for every k $(1\leqq k\leqq d)$,*

$$\sin\theta_k \geqq (\sqrt{2}/3)^d. \tag{5.1}$$

Note, that, while the condition of being Lovász-reduced depends on the order of $b_1, \ldots, b_d$, the conclusion is independent of this ordering.

Proof. We have to prove that, for any $m \in U_k$,

$$|b_k| \leqq c(d)|m - b_k|, \tag{5.2}$$

where

$$c(d) = (9/2)^{d/2}. \tag{5.3}$$

This is equivalent to (5.1), since

$$\sin \theta_k = \min_{m \in U_k} \frac{|m - b_k|}{|b_k|}.$$

Let $m = \sum_{i \neq k} \alpha_i b_i \in U_k$. Then m can be written as

$$m = \sum_{j=1}^{d} \beta_j b_j^*, \tag{5.4}$$

where, by (2.1),

$$\beta_j = \sum_{t \neq k} \alpha_t \mu_{tj}. \tag{5.5}$$

Now

$$m - b_k = \sum_{j=1}^{d} \gamma_j b_j^*, \tag{5.6}$$

where (again by (2.1))

$$\gamma_j = \beta_j - \mu_{kj}. \tag{5.7}$$

So far the symbol α_i has not been defined for $i = k$; let us set

$$\alpha_k = -1. \tag{5.8}$$

With this convention, from (5.5) we obtain

$$\gamma_j = \sum_{t=1}^{d} \alpha_t \mu_{tj} = \alpha_j + \sum_{t=j+1}^{d} \alpha_t \mu_{tj}. \tag{5.9}$$

Now our claim (5.2) reads as follows:

$$\sum_{j=1}^{k} \mu_{kj}^2 |b_j^*|^2 \leqq c(d)^2 \sum_{j=1}^{d} \gamma_j^2 |b_j^*|^2. \tag{5.10}$$

Claim. *Let $\alpha_1, \ldots, \alpha_d$ be arbitrary reals, $\alpha_k = -1$. Assume that the real numbers μ_{ij} satisfy* (2.2) *through* (2.4). *Let γ_j be defined by* (5.9). *Then*

$$\sum_{j=k}^{d} \gamma_j^2 \geqq (2/3)^{2(d-k)}. \tag{5.11}$$

Proof. Assuming the contrary, we deduce

$$|\gamma_j| < \varepsilon \quad \text{for} \quad k \leqq j \leqq d, \tag{5.12}$$

where $\varepsilon = (2/3)^{d-k}$.

By (5.9) and (2.2)—(2.4), we have

$$\begin{aligned}
\gamma_d &= \alpha_d \\
\gamma_{d-1} &= \alpha_{d-1} + \alpha_d \mu_{d,d-1} \\
\gamma_{d-2} &= \alpha_{d-2} + \alpha_{d-1}\mu_{d-1,d-2} + \alpha_d \mu_{d,d-2} \\
&\vdots \\
\gamma_k &= \alpha_k + \alpha_{k+1}\mu_{k+1,k} + \ldots + \alpha_d \mu_{d,k}
\end{aligned}$$

We claim that

$$|\alpha_j| < (3/2)^{d-j}\varepsilon \quad (k \leqq j \leqq d). \tag{5.13}$$

This is true for $j=d$ by (5.12) because $|\alpha_d| = |\gamma_d| < \varepsilon$. We proceed by reverse induction on j.

$$|\alpha_j| = \Big|\gamma_j - \sum_{t=j+1}^{d} \alpha_j \mu_{tj}\Big| \leqq |\gamma_j| + \sum_{t=j+1}^{d} |\alpha_t|/2.$$

Using the induction hypothesis, we obtain

$$|\alpha_j| < \varepsilon + \sum_{t=j+1}^{d} (3/2)^{d-t}\varepsilon/2 = (2/3)^{d-j}\varepsilon.$$

This proves (5.13).

Let us now set $j=k$ in (5.13). In view of (5.8) ($\alpha_k = -1$), we obtain $1 = |\alpha_k| < (3/2)^{d-k}\varepsilon$, a contradiction, proving (5.11). ∎

Now the proof of (5.10) is immediate. By (2.5), we have

$$|b_k^*|^2 \geqq 2^{j-k}|b_j^*|^2 \quad \text{for} \quad 1 \leqq j \leqq k \tag{5.14}$$

and

$$|b_k^*|^2 \leqq 2^{j-k}|b_j^*|^2 \quad \text{for} \quad k \leqq j \leqq d. \tag{5.15}$$

Thus the left hand side of (5.10) can be bounded as

$$\sum_{j=1}^{k} \mu_{kj}^2 |b_j^*|^2 \leqq \sum_{j=1}^{k} \mu_{kj}^2 2^{k-j}|b_k^*|^2 < 2^k |b_k^*|^2. \tag{5.16}$$

On the other hand, concerning the right hand side of (5.10) we obtain

(5.17)

$$\sum_{j=1}^{d} \gamma_j^2 |b_j^*|^2 \geqq \sum_{j=k}^{d} \gamma_j^2 |b_j^*|^2 \geqq \sum_{j=k}^{d} \gamma_j^2 2^{k-j}|b_k^*|^2 \geqq |b_k^*|^2 2^{k-d} \sum_{j=k}^{d} \gamma_j^2 \geqq |b_k^*|^2 (2/9)^{d-k}.$$

(We used (5.11) in the last step.)

A comparison of (5.17) and (5.16) yields (5.10), and thus completes the proof of the Theorem. ∎

6. Rounding off: error estimate

We apply Theorem 5.1 to prove Theorem 3.2.

Let $B=\{b_1, \ldots, b_d\}$ be a Lovász-reduced basis and $1 \leqq k \leqq d$. Let U_k denote the linear subspace generated by $B-\{b_k\}$. We shall use (5.2) which asserts that for any $m \in U_k$,

$$|b_k| \leqq c(d)|m-b_k|, \tag{6.1}$$

where

$$c(d) = (9/2)^{d/2}. \tag{6.2}$$

Let now x be the point to be approximated, and w the lattice point assigned to x by the ROUNDING OFF procedure. Then

$$w-x = \sum_{i=1}^{d} \delta_i b_i, \tag{6.3}$$

where

$$|\delta_i| \leqq 1/2 \quad (i = 1, \ldots, d). \tag{6.4}$$

Let u be the nearest lattice point to x, and let

$$u-w = \sum_{i=1}^{d} \varphi_i b_i \quad (\varphi_i \in \mathbf{Z}). \tag{6.5}$$

Claim. *The following inequality holds:*

$$|u-w| \leqq 2dc(d)|u-x|. \tag{6.6}$$

Proof. We may assume $u \neq w$. Let $|\varphi_k b_k| = \max_j |\varphi_j b_j| > 0$. Then

$$|u-w| \leqq \sum_{j=1}^{d} |\varphi_j b_j| \leqq d|\varphi_k b_k|. \tag{6.7}$$

On the other hand,

$$u-x = (u-w)+(w-x) = \sum_{i=1}^{d} (\varphi_i+\delta_i)b_i = (\varphi_k+\delta_k)(b_k-m),$$

where

$$m = -\frac{1}{\varphi_k+\delta_k} \sum_{j \neq k} (\varphi_j+\delta_j) b_j \in U_k.$$

Consequently, by (6.1) and (6.4),

$$|u-x| = |\varphi_k+\delta_k|\,|b_k-m| \geqq \frac{|\varphi_k|}{2c(d)}|b_k|. \tag{6.8}$$

Comparing (6.7) and (6.8) we obtain $|u-w| \leqq d|\varphi_k||b_k| \leqq 2dc(d)|u-x|$. ∎

The conclusion is now immediate. $|x-w| \leqq |x-u|+|u-w| \leqq |x-u|(1+2dc(d))$, proving Theorem 3.2. ∎

7. Nonhomogeneous simultaneous diophantine approximation

A nearly optimal nonhomogeneous diophantine approximation algorithm follows from any nearby lattice point procedure.

Let $\alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_d$ be real numbers. Given $\varepsilon>0$, let $q(\varepsilon)$ denote the smallest positive integer such that the following system of $d+1$ inequalities is solvable in integers $\bar{p}_1, \ldots, \bar{p}_d$ and $\bar{q}$.

$$|\bar{q}\alpha_i - \bar{p}_i - \beta_i| \leqq \varepsilon \quad \text{for} \quad i = 1, \ldots, d, \tag{7.1}$$

and

$$|\bar{q}| \leqq q(\varepsilon). \tag{7.2}$$

If no such $q(\varepsilon)$ exists we set $q(\varepsilon)=\infty$.

Theorem 7.1. *Let $\alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_d$ and $\varepsilon>0$ be given rational numbers. Then one can find in polynomial time*
(a) *either a proof that $q(\varepsilon)=\infty$,*
(b) *or integers $p_1, \ldots, p_d, q$ such that*

$$|q\alpha_i - p_i - \beta_i| \leqq C_d\varepsilon \tag{7.3}$$

and

$$|q| \leqq C_d q(\varepsilon) \tag{7.4}$$

where

$$C_d = 4\sqrt{d}\,2^{d/2}. \tag{7.5}$$

Here, the length of the input is measured by

$$I = \sum \text{length}\,(\alpha_i) + \sum \text{length}\,(\beta_i) + |\log \varepsilon|, \tag{7.6}$$

where the length of a fraction a/b is the numbers of bits of a and b together. (log denotes base 2 logarithm.)

In particular, the size of $q(\varepsilon)$, even if it is finite, is not part of the input length and no *a priori* information on $q(\varepsilon)$ is required.

First of all, we replace ε by δ such that $\delta \leqq \varepsilon > \delta/2$ and δ is a power of 2. This way, length $(\delta) < |\log \varepsilon| + 2$ and the possibly large quantity length(ε) will play no role in the algorithm. Clearly, $q(\delta) \leqq q(\varepsilon)$.

Next, we establish an upper bound on $q(\varepsilon)$. Let

$$Q = \prod_{i=1}^{d} \text{denom}\,(\alpha_i) \tag{7.7}$$

where denom(r) denotes the denominator of the fraction r.

Proposition 7.2. *For any $\varepsilon>0$, if $q(\varepsilon)$ is finite then $q(\varepsilon) \leqq Q$.*

Proof. In (7.1), $\bar{q}$ can be replaced by its residue mod Q. ∎

In the procedure, we shall have to guess the value of $q(\delta)$ within a factor of $\sqrt{2}$. Let s denote the guessed value. We shall perform the procedure for $s=1, 2, 4, \ldots, \tilde{Q}=2^{\lceil \log Q \rceil}$. Unless $q(\delta)=\infty$, one of the values s will satisfy

$$q(\delta)/\sqrt{2} < s < \sqrt{2}\, q(\delta). \tag{7.8}$$

Given s, the procedure runs as follows.

PROCEDURE APPR(s). Let $e_1, \ldots, e_{d+1}$ be the standard orthonormal basis of $\mathbf{R}^{d+1}$. Let

$$b_i = -e_i \quad (i = 1, \ldots, d)$$

and

$$b_{d+1} = \sum_{i=1}^{d} \alpha_i e_i + \frac{\delta}{s} e_{d+1}.$$

The set $B=\{b_1, \ldots, b_{d+1}\}$ is a basis of a lattice L. Let $x=\sum_{i=1}^{d} \beta_i e_i$.

Apply Lovász' lattice reduction algorithm to B and subsequently the NEAREST PLANE procedure to (x, L) to find a lattice point $w=\sum_{i=1}^{d} p_i b_i + q b_{d+1}$ near s.

The list $(p_1, \ldots, p_d, q)$ is the output of APPR(s).

PROCEDURE APPROX.

For $s=1, 2, 4, \ldots, \tilde{Q}$ *do* APPR(s).

Let S denote the set

$$S = \{s \mid \text{the output of APPR}(s) \text{ satisfies (7.3)}\}.$$

If $S=\emptyset$, print "$q(\varepsilon)=\infty$".

Else, let $s_0 \in S$ produce the smallest value of $|q|$ in the output of APPR(s_0). Print this output.

Claim 7.3. *The output of PROCEDURE APPROX satisfies Theorem 7.1.*

Proof. The polynomial running time is justified by the definition of $\tilde{Q}$.

Assume $q(\varepsilon)$ is finite. Then so is $q(\delta)$. Let s_1 be a power of 2 satisfying (7.8).

Let $u=\sum_{i=1}^{d} \bar{p}_i b_i + \bar{q} b_{d+1}$ where $(\bar{p}_1, \ldots, \bar{p}_d, \bar{q})$ satisfy (7.1) and (7.2) with δ in place of ε. Hence

$$|u-x|_{\max} = \max_i \left\{ |\bar{q}\alpha_i - \bar{p}_i - \beta_i|, \frac{|\bar{q}|\delta}{s_1} \right\}$$

where $|\ldots|_{\max}$ refers to maximum norm. Consequently, by (7.8),

$$|u-x|_{\max} \leqq \delta \max\left(1, \frac{|\bar{q}|}{s_1}\right) \leqq \delta\sqrt{2}.$$

Therefore $|u-x| \leqq \delta\sqrt{2d}$. By Theorem 3.1 we infer $|w-x| \leqq c_{d+1}|u-x| \leqq 2\sqrt{d}\,2^{d/2}\delta$. On the other hand,

$$|w-x| \geqq |w-x|_{\max} = \max_i \left\{ |q\alpha_i - p_i - \beta_i|, \frac{|q|\delta}{s_1} \right\}.$$

This implies

$$|q\alpha_i - p_i - \beta_i| \leqq 2\sqrt{d}\, 2^{d/2}\delta < 4\sqrt{d}\, 2^{d/2}\varepsilon$$

and

$$|q| \leqq 2\sqrt{d}\, 2^{d/2} s_1 < \sqrt{d}\, 2^{(d+3)/2} q(\delta) < \sqrt{d}\, 2^{(d+3)/2} q(\varepsilon).$$

Since s_0 is "at least as good" as s_1, the proof of Claim 7.3 and Theorem 7.1 is complete. ∎

8. Integer programming

The basic problem of integer programming is to find an integral solution to a system of linear inequalities if one exists. H. W. Lenstra [12] invented an algorithm to solve integer programming with a bounded number of variables in polynomial time. Grötschel, Lovász and Schrijver put Lenstra's result in the more general framework of finding integral points in convex bodies. By a convex body (K, d, r_1, r_2) we mean a closed convex set K in $\mathbf{R}^d$, contained in the ball of radius r_1 around the origin and containing some ball of radius $r_2 > 0$. In [5] (p. 175), Lenstra's result is rephrased as follows.

Theorem 8.1. *Given a convex body (K, d, r_1, r_2) by a separation oracle, we can decide if K contains an integral point, and find such a point if it exists. For fixed d the procedure is polynomial time.* ∎

Grötschel, Lovász and Schrijver [5] use a combination of the Ellipsoid Method and Lovász' lattice reduction to derive 8.1. They actually prove:

Theorem 8.2. *Given a convex body (K, d, r_1, r_2) by a separation oracle, one can achieve in polynomial time one of the following:*
(a) find an integral point in K;
(b) find an affine transformation A of $\mathbf{R}^d$ which maps $\mathbf{Z}^d$ onto itself such that the first coordinate of any point in AK is less than $f(d) = d^{7/2} 2^{d(d-1)/4}$ in absolute value. ∎

Note that this algorithm runs in polynomial time even for variable d and results in an integer programming algorithm (generalized in the above sense to convex bodies) where one problem in d variables is reduced to $\lceil 2f(d) \rceil$ problems in $(d-1)$ variables.

One can use the NEAREST PLANE procedure to improve the value of $f(d)$.

Theorem 8.3. *Theorem 8.2 remains valid if we replace $f(d)$ by $g(d) = d^{3/2} 2^{d/2}$.*

We remark that R. Kannan [7] found an algorithm that reduces the integer programming problem in d variables to a *polynomial* number of problems in $(d-1)$ variables. The cost of such a reduction step is proportional to d^{cd} (c a constant). While this time bound is not polynomial, the limited branching in Kannan's procedure assures a $d^{c'd}$ overall running time for the integer programming problem, substantially better than the $\exp(cd^2)$ bound that follows from 8.3. It would be interesting to know whether Kannan's bound remains valid for convex bodies given by separation oracles.

Proof. Let $B(x, R)$ denote the ball of radius R with center x in $\mathbf{R}^d$. The first phase of the algorithm described in [5] employs the "Shallow Cut Ellipsoid Method" to find a (rational) linear transformation T of $\mathbf{R}^d$ and a point x such that $B(x, d^{-3/2}R) \subseteqq TK \subseteqq B(x, R)$ for some R. We take this as our starting point.

The second phase is to find a Lovász-reduced basis $b_1, \ldots, b_d$ of the lattice $T\mathbf{Z}^d$.

In the third phase, we use the NEAREST PLANE procedure and Theorem 3.1 to find $w \in T\mathbf{Z}^d$ such that

$$|x-w| < 2^{d/2-1}|b_d^*|. \tag{8.1}$$

Clearly, $T^{-1}w$ is an integral point and the separation oracle will tell us whether $T^{-1}w$ belongs to K or not. If it does, we are done (the alternative (a) holds).

If it does not, then $w \notin TK$ and therefore $w \notin B(x, d^{3/2}R)$. In other words, $|x-w| > d^{-3/2}R$.

A comparison with (8.1) shows that

$$\frac{R}{|b_d^*|} < d^{3/2}2^{d/2-1} = \frac{g(d)}{2}. \tag{8.2}$$

The rest of the proof is routine. Let

$$z = \sum_{i=1}^{d} \alpha_i(z) b_i \tag{8.3}$$

be the representation of any $z \in \mathbf{R}^d$ as a linear combination of the b_i. Let π denote the orthogonal projection of $\mathbf{R}^d$ on the line $\mathbf{R}b_d^*$. Then

$$\pi(z) = \alpha_d(z) b_d^* \quad (z \in \mathbf{R}^d). \tag{8.4}$$

If z belongs to $TK \subseteqq B(x, R)$, we have $|x-z| \leqq R$, whence

$$|\pi(z)-\pi(x)| \leqq R. \tag{8.5}$$

Applying (8.4) to both z and x we deduce $|\alpha_d(z)-\alpha_d(x)||b_d^*| \leqq R$. Consequently, by (8.2) we obtain

$$|\alpha_d(z)-\alpha_d(x)| < g(d) \quad \text{for} \quad z \in TK. \tag{8.6}$$

Let now C be the linear transformation of $\mathbf{R}^d$ which maps $T^{-1}b_i$ to e_{d+1-i} where $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^t$ (the j^{th} entry is 1). $\mathbf{Z}^d$ is clearly invariant under C. Setting $z = Ty$ and applying CT^{-1} to both sides of (8.3) we obtain

$$Cy = \sum_{i=1}^{d} \alpha_i(Ty) e_{d+1-i}. \tag{8.7}$$

(8.6) then tells us that for any $y \in K$, the first coordinate of Cy is $\alpha_d(z)$ (where $z = Ty$) satisfying (8.6).

Let β be the integer nearest to $\alpha_d(x)$, and let us define the affine transformation A by

$$A(y) = Cy - \beta e_1 \quad (y \in \mathbf{R}^d). \tag{8.8}$$

Then by (8.6) and (8.7), for any $y \in K$, the first coordinate of Ay is less than $(g(d)+1)/2 < g(d)$ in absolute value. ∎

References

[1] L. ADLEMAN, On breaking the iterated Merkle—Hellman public key cryptosystem, *Proc.* **15***th ACM Symp. on Theory of Computing*, Boston (1983), 402—412.

[2] J. W. S. CASSELS, *An introduction to the geometry of numbers*, Springer, New York, (1971).

[3] P. VAN EMDE BOAS, Another NP-complete partition problem and the complexity of computing short vectors in a lattice, *Rep. MI/UVA 81—04*, Amsterdam (1981).

[4] M. GRÖTSCHEL, L. Lovász and A. SCHRIJVER, The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1** (1981), 186—197.

[5] M. GRÖTSCHEL, L. Lovász and A. SCHRIJVER, Geometric methods in combinatorial optimization, *in: Progress in Combinatorial Optimization* (W. R. Pulleyblank, ed.), *Proc. Silver Jubilee Conf. on Comb., Univ. Waterloo*, Vol. 1, 1982, Acad. Press, N. Y. (1984), 167—183.

[6] B. HELFRICH, An algorithm to construct Minkowski-reduced lattice-bases, *in: Proc. 2nd Ann. Symp. on Theoretical Aspects of Comp. Sci.* (STACS 85), *Springer Lect. Notes. in Comp. Sci.* **182** (1985), 173—179.

[7] R. KANNAN, Improved algorithms for integer programming and related lattice problems, *in: Proc. 15th ACM Symp. on Theory of Comp.*, (1983), 193—206.

[8] R. KANNAN, A. K. LENSTRA and L. Lovász, Polynomial factorization and nonrandomness of bits of algebraic and some transcendental numbers, *in: Proc. 16th Ann. ACM Symp. on Theory of Computing*, Washington, D. C. (1984), 191—200.

[9] J. LAGARIAS and A. M. ODLYZKO, Solving low density subset sum problems, *in: Proc. 24th IEEE Symp. on Foundations of Comp. Sci.*, (1983), 1—10.

[10] A. K. LENSTRA, Lattices and factorization of polynomials, *Report IW 190/81, Mathematisch Centrum*, Amsterdam (1981).

[11] A. K. LENSTRA, H. W. LENSTRA, JR. and L. Lovász, Factoring polynomials with rational coefficients, *Math. Ann.* **261** (1982), 515—534.

[12] H. W. LENSTRA, JR., Integer programming with a fixed number of variables. *Math. Oper. Res.* **8** (1983), 538—548.

[13] L. Lovász, *private communications*, 1981—1982.

[14] A. M. ODLYZKO and H. TE RIELE, Disproof of the Mertens conjecture, *J. reine angew. Math.* **357** (1985), 138—160.

[15] A. SHAMIR, A polynomial time algorithm for breaking the Merkle—Hellman cryptosystem, *Proc. 23rd IEEE Symp. on Foundations of Comp. Sci.*, Chicago, Illinois (1982), 145—152.

[16] C. P. SCHNORR, A hierarchy of polynomial time basis reduction algorithms, *in: Theory of Algorithms, Proc. Conf. Pécs (Hungary)* 1984, *Coll. Soc. J. Bolyai, to appear.*

[17] VERA T. Sós, On the theory of diophantine approximation II, *Acta Math. Acad. Sci. Hung.* (1958), 229—241.

[18] VERA T. Sós, Irregularities of partitions: Ramsey theory, uniform distribution, *in: Surveys in Combinatorics, Proc. 9th British Combinatorial Conference*, 1983 *(E. Keith Lloyd, ed.) London Math. Soc. Lect. Notes* **82,** Cambridge Univ. Press 1983.

László Babai

Department of Algebra
Eötvös University
Budapest, Hungary H-1088

and

Department of Computer Science
The University of Chicago
Chicago, Illinois 60 637