Discrete Mathematics 58 (1986) 215–220
North-Holland

SMALL SOLUTIONS OF LINEAR DIOPHANTINE EQUATIONS

I. BOROSH, M. FLAHIVE and B. TREYBIG
Texas A & M University, College Station, TX 77843, U.S.A.

Received 21 October 1983
Revised 17 June 1985

Let $Ax = B$ be a system of $m \times n$ linear equations with integer coefficients. Assume the rows of A are linearly independent and denote by X (respectively Y) the maximum of the absolute values of the $m \times m$ minors of the matrix A (the augmented matrix (A, B)). If the system has a solution in nonnegative integers, it is proved that the system has a solution $X = (x_i)$ in nonnegative integers with $x_i \leqslant X$ for $n - m$ variables and $x_i \leqslant (n - m + 1)Y$ for m variables. This improves previous results of the authors and others.

1. Introduction

Given a system of linear equations with integral coefficients which is assumed to have a non-trivial integral solution, can one guarantee the existence of a "small" solution? Answers to this question have had many applications in various branches of mathematics and theoretical computer science.

In the case of a homogeneous system Siegel's Lemma [5] uses the pigeonhole principle to give such a bound in terms of the coefficients. This bound has been used repeatedly in work in Diophantine approximation and transcendence theory such as the proof of Roth's theorem [5] and Baker's estimates for linear forms of logarithms [1].

In [11] the problem of guaranteeing nontrivial, small non-negative integral solutions to a system of linear Diophantine equations arose in connection with a topological question. In [3] a bound depending on the minors of highest order was obtained. This bound was improved in [4] to nY, where n is the number of variables and Y is the maximum minor of order equal to the rank. In [7] the authors considered the more general problem of integer solutions to a system of linear equations and inequalities. They obtained a representation of all rational solutions and derived from that a bound similar to the one obtained in [4].

The bounds in [3], [4], as well as rougher estimates established by Cook in an unpublished manuscript, were used [6] to prove that the problem of obtaining nonnegative solutions to a system of linear Diophantine equations is NP. Recently, several related problems were found to be solvable in polynomial time. In each of these problems the analysis of the running time of the algorithm depends on bounds on the solution of a linear Diophantine equation. As an

0012-365X/86/$3.50

example, Kachian [9] (see [8] also) showed that the linear programming problem can be solved in polynomial time. Lenstra [10] then showed that the integer linear programming problem with a bounded number of variables is polynomial.

In this paper, we continue the study of the size of small solutions to a system of linear Diophantine equations. The new results are given in Theorems 1 and 2. We precede the statements of these results with the establishment of some notation and a discussion of their usefulness in obtaining small solutions.

2. Notation and results

We let $Ax = B$ be a matrix equation of the form

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1,n+1} \\ \vdots \\ a_{m,n+1} \end{bmatrix}, \tag{1}$$

where each a_{ij} is an integer. Given distinct integers $1 \leq j_1, \ldots, j_m \leq n+1$, let

$$d_{j_1,\ldots,j_m} = \det \begin{bmatrix} a_{1j_1} & \cdots & a_{1j_m} \\ \vdots & \ddots & \vdots \\ a_{mj_1} & \cdots & a_{mj_m} \end{bmatrix}, \tag{2}$$

$$X = \sup\{|d_{j_1,\ldots,j_m}| : j_i \leq n\}, \tag{3}$$

$$Y = \sup\{|d_{j_1,\ldots,j_m}|\}. \tag{4}$$

Given distinct integers $1 \leq j_1 < \cdots < j_m \leq n+1$ let $d_{\{j_1,\ldots,j_m\}} = |d_{j_1,\ldots,j_m}|$. Since [2] shows the theorems of this paper hold if $B = 0$, we assume $B \neq 0$.

In [2] it was conjectured that if the rows of A are linearly independent, then if there is a positive integral solution $x = (x_i)$ to $Ax = B$, there is such a solution x where $\sup x_i \leq Y$. Thus, if the bound Y holds, then one way to find such a solution $x = (x_i)$ is to find an $m \times m$ non zero minor of A, $d = d_{j_1,\ldots,j_m}$, and then find x by guessing values for x_i, $i \notin \{j_1, \ldots, j_m\}$, and solving for the x_i, $i \in \{j_1, \ldots, j_m\}$. However, this same technique would work, and involve the same amount of computation, if all we knew was that $d \neq 0$, and the x_i, $i \notin \{j_1, \ldots, j_m\}$ could be chosen small ($\leq Y$) but we did not know the size of the x_i, $i \in \{j_1, \ldots, j_m\}$. It is in this direction we have worked in finding the following Theorems 1 and 2. Indeed, Theorem 1 shows that for some $d = d_{j_1,\ldots,j_m} \neq 0$ the x_i, $i \notin \{j_1, \ldots, j_m\}$, can be chosen $\leq X$ and the remaining $x_i \leq (n-m)X + Y$. Theorem 2 refines the results of Theorem 1 in the case that all the $m \times m$ minors of A are nonzero.

Theorem 1. *If the rows of A are linearly independent and $x = (x_i)$ is a nonzero nonnegative integral solution to* (1), *then there exists a nonzero integral solution $y = (y_i)$ to* (1) *and integers $1 \leq j_1 \leq \cdots \leq j_m \leq n$ such that $d_{j_1,\ldots,j_m} \neq 0$, $0 \leq y_p \leq X$ for $p \notin \{j_1, \ldots, j_m\}$ and $0 \leq y_p \leq (n-m)X + Y$ otherwise.*

Theorem 2. *If all the $m \times m$ minors of A are nonzero and there exists a nonnegative integral solution $x = (x_i)$ to (1) such that $x_i > Y$ for some i, then there exist distinct integers $j_1, \ldots, j_{m+1}$ such that*

(a) *for each k, $(-1)^k d_{j_1,\ldots,j_{k-1},j_{k+1},\ldots,j_{m+1}}$ has the same sign as $(-1)^m d_{j_1,\ldots,j_m}$, and*

(b) *there exists a nontrivial nonnegative integral solution $y = (y_i)$ to (1) such that*

(i) *$y_i \leq (n-m)X + Y$ for all i, and*

(ii) *$y_i < |d_{j_1,\ldots,j_m}|$ for $i \notin \{j_1, \ldots, j_{m+1}\}$, and*
$y_{i_0} < |d_{j_1,\ldots,j_{i_0-1},j_{i_0+1},\ldots,j_{m+1}}|$ for some $i_0 \in \{j_1, \ldots, j_{m+1}\}$.

Proofs of theorems Since the rows of A are linearly independent, then $m \leq n$. If $m = n$, then $d = d_{1,\ldots,m} \neq 0$ and both results follow from Cramer's Rule, the second theorem vacuously. Thus, assume $m < n$.

Proof of Theorem 1. Given integers $1 \leq j_1, \ldots, j_m \leq n$, equation (1) implies

$$\begin{aligned} d_{j_1,\ldots,j_m} x_{j_1} &= -\sum_i d_{i,j_2,\ldots,j_m} x_i + d_{n+1,j_2,\ldots,j_m}, \\ &\vdots \\ d_{j_1,\ldots,j_m} x_{j_m} &= -\sum_i d_{j_1,\ldots,j_{m-1},i} x_i + d_{j_1,\ldots,j_{m-1},n+1}, \end{aligned} \tag{5}$$

where i ranges over $\{1, \ldots, n\} - \{j_1, \ldots, j_m\}$ in each summand.

Let $x = (x_i)$ be a nontrivial nonnegative integral solution to (1) with a minimal number of coordinates larger than X, say w. Reorder the variables if necessary to obtain $x_i > X$ if $1 \leq i \leq w$. If $w = 0$ the theorem holds, so suppose $w \geq 1$.

Lemma 1. *Let $1 \leq h \leq w$ be a positive integer such that $d_{j_1,\ldots,j_m} = 0$ for any choice of $1 \leq j_1 < \cdots < j_h \leq w < j_{h+1} < \cdots < j_m \leq n$. Then, $d_{j_1,\ldots,j_m} = 0$ for any choice of $1 \leq j_1 < \cdots < j_{h-1} \leq w < j_h < \cdots < j_m \leq n$.*

Proof. Suppose that $d = d_{j_1,\ldots,j_m} \neq 0$ for some choice of $1 \leq j_1 < \cdots < j_{h-1} \leq w < j_h < \cdots < j_m \leq n$. We may assume without loss of generality that $d > 0$. For any $j_{m+1} \in \{1, \ldots, w\} - \{j_1, \ldots, j_{h-1}\}$ the hypothesis implies that if $k \geq h$, the coefficient of $x_{j_{m+1}}$ in row k of (5) is zero. Therefore, for any integer q, $y = (y_i)$ defined by

$$\begin{aligned} y_{j_p} &= x_{j_p} + q d_{j_1,\ldots,j_{p-1},j_{m+1},j_{p+1},\ldots,j_m}, \quad \text{for } 1 \leq p \leq h-1, \\ y_{j_{m+1}} &= x_{j_{m+1}} - qd, \\ y_i &= x_i, \quad \text{otherwise}, \end{aligned} \tag{6}$$

is an integral solution to (1). Since $x_i > X$ for all $i = j_1, \ldots, j_{h-1}$, there is a positive integer q for which y is nonnegative and at most $w-1$ coordinates are greater than X. This contradicts the minimality of w and completes the proof of Lemma 1. $\square$

Remark. The independence of the rows of A combined with successive applications of Lemma 1, yields that neither $h = w$ nor $h = m$ satisfies the hypothesis of Lemma 1.

We continue now with the proof of Theorem 1. By way of contradiction suppose that $w > m$. By the above Remark, for some $1 \leq j_1 < \cdots < j_m \leq w$, $d = d_{j_1,\ldots,j_m}$ is not zero, so without loss of generality let $d > 0$. Taking $j_{m+1} \in \{1, \ldots, w\} - \{j_1, \ldots, j_m\}$ and $h = m + 1$, y defined as in (6) is an integral solution to (5) for any integer q. Since $d > 0$, there is a positive integer q for which y is a nontrivial nonnegative integral solution to (1) with at most $w - 1$ coordinates larger than X. This contradiction implies that $w \leq m$.

To complete the proof of Theorem 1 now suppose that $x = (x_i)$ is a nontrivial nonnegative solution to (1) such that:

(i) there exists $J = \{j_1, \ldots, j_m\}$, where $d_J \neq 0$;

(ii) if $i \notin J$, then $x_i \leq X$, and among all such solutions satisfying (i), (ii) we have

(iii) $\sum_{i \notin J} x_i$ is a minimum.

Without loss of generality suppose that we have reordered the x_i's so that $x_1 \geq \cdots \geq x_m$, $x_{m+1} \geq \cdots \geq x_n$, and $J = \{1, \ldots, m\}$.

If $x_i \geq d_{\{1,\ldots,m+1\}-\{i\}}$ for all $i = 1, \ldots, m + 1$, then using $q = \operatorname{sign}(d)$ and $j_i = i$, for $i = 1, \ldots, n$, the solution y defined by (6) satisfies $\sum_{i \notin J} y_i < \sum_{i \notin J} x_i$, a contradiction. Hence, there exists $1 \leq i \leq m + 1$ such that $x_i < d_{\{1,\ldots,m+1\}-\{i\}}$.

Suppose then that i is the largest integer $j \leq m + 1$ such that $x_j < d_{\{1,\ldots,m+1\}-\{j\}}$. If $x_i < x_{m+1}$, let $J' = \{1, \ldots, m + 1\} - \{i\}$ and note that $\sum_{p \notin J'} x_p < \sum_{p \notin J} x_p$, a contradiction. If $x_i \geq x_{m+1}$, we find that if (5) is applied where $j_1 = 1, \ldots, j_{i-1} = i - 1$, $j_i = m + 1$, $j_{i+1} = i + 1, \ldots, j_m = m$, then the first equation of (5) is of the form $Dx_1 = A_i x_i + A_{m+2} x_{m+2} + \cdots + A_n x_n + A_{n+1}$, $|D| > x_i \geq x_{m+2} \geq \cdots \geq x_n$, and where $|A_i| \leq X$ for $i \leq n$ and $|A_{n+1}| \leq Y$. Since $|D| > 1$, we thus have that $x_m \leq \cdots \leq x_1 \leq (n - m)X + Y$, proving Theorem 1. $\square$

Before we proceed with the proof of Theorem 2 we shall prove

Lemma 2. *If each $m \times m$ minor of A is nonzero, and there is a nontrivial nonnegative solution to $Ax = 0$, then there exist distinct $j_1, \ldots, j_{m+1}$ such that, for each k, $(-1)^{k-1} d_{j_1,\ldots,j_{k-1},j_{k+1},\ldots,j_{m+1}}$ has the same sign as $(-1)^m d_{j_1,\ldots,j_m}$.*

Proof. Let *h be a nontrivial nonnegative integral solution to $Ax = 0$* with a minimum number of nonzero coordinates. Reorder the variables, if necessary, to obtain $h = (h_1, \ldots, h_t, 0, 0, \ldots, 0)$ with each $h_j > 0$. Since $t \geq 1$, the system (5) with $\{j_1, \ldots, j_m\} = \{1, \ldots, m\}$ implies that $t \geq m + 1$.

Suppose that $t > m + 1$ and let A' be the matrix consisting of the first t columns of A. Then, $h' = (h_1, \ldots, h_t)$ is a nontrivial solution to $A'x = 0$. Since $t - m \geq 2$, there is an integral solution h'' to $A'x = 0$ such that h' and h'' are linearly independent. For $l = \min\{h''_i/h_i,\ i = 1, \ldots, t\}$, $h'' - lh'$ is a nonnegative rational solution to $A'x = 0$ with at least one zero coordinate. The linear independence of h', h'' implies $h'' - lh'$ is nontrivial, so by multiplying by an appropriate integer we contradict the minimality of t. Hence $t = m + 1$.

Let A' consist of the first $m+1$ columns of A. Then all solutions of $A'x = 0$ are multiples of

$$(d_{2,\ldots,m+1}, -d_{1,3,\ldots,m+1}, \ldots, (-1)^m d_{1,\ldots,m}).$$

Since there is a positive solution all coordinates of this solution must be of the same sign. □

Proof of Theorem 2. If there is no nontrivial nonnegative integral solution to $Ax = 0$, then by Theorem 4 of [3] $x_i \leqslant Y$, for all i. Hence there is such a solution to $Ax = 0$, so let h and $j_1, \ldots, j_{m+1}$ satisfy the conclusion of Lemma 2. Reorder the variables, if necessary, to obtain $\{j_1, \ldots, j_{m+1}\} = \{1, \ldots, m+1\}$ and

$$0 < d_{1,\ldots,m} = \min\{d_{\{1,\ldots,m+1\}-\{i\}}\} = D.$$

For $i > m+1$, the ith coordinate of x will be reduced by utilizing multiples of the solutions $r^{(i)}$ to $Ax = 0$, defined by

$$r_j^{(i)} = \begin{cases} -d_{1,\ldots,i-1,j,i+1,\ldots,m}, & j \leqslant m \\ D, & j = i \\ 0, & \text{otherwise} \end{cases}$$

balanced with a suitable multiple of h, where h is defined as in Lemma 2. Namely, for $s = \max\{x_i\}$, y defined by

$$y = x + (n-m)Xsh - \sum_{i \geqslant m+2} \left[\frac{x_i}{D}\right] r^{(i)}$$

is an integral solution to (1) with $0 \leqslant y_j = x_j - [x_j/D]D < D$, for all $j > m+1$. Also for all $j \leqslant m+1$,

$$\begin{aligned} y_j &= x_j + (n-m)Xsh_j - \sum_{i \geqslant m+2} \left[\frac{x_i}{D}\right] r_j^{(i)} \\ &\geqslant (n-m)Xs - \sum_{i \geqslant m+2} s\,|r_j^{(i)}| \\ &\geqslant Xs > 0. \end{aligned}$$

Now let $l = \min_{i \leqslant m+1}\{[y_i/h_i]\} = [y_{i_0}/h_{i_0}]$, and define $z = y - lh$. Then z is a nonnegative integral solution to (1) in which $0 \leqslant z_i = y_i \leqslant D$ for all $i \geqslant m+2$. Taking $D^* = d_{\{i,\ldots,m+1\}\setminus\{i_0\}}$ it is observed that $z_{i_0} \leqslant D^*$. If z is trivial, then y is a solution of $Ax = B$ and $Ax = 0$, implying $B = 0$, a contradiction.

Also, for all $j \leqslant m+1$, $j \neq i_0$, we have

$$D^* z_j = a_j^{(i_0)} z_{i_0} + \sum_{i \geqslant m+2} a_j^{(i)} z_i + a_j^{(n+1)},$$

where, for $i \leqslant n$, each $a_j^{(i)}$ is an $m \times m$ minor of A, and each $a_j^{(n+1)}$ is an $m \times m$

minor of the augmented matrix. Hence, for all $j \leqslant m+1$,

$$z_j \leqslant |a_j^{(i_0)}| \frac{z_{i_0}}{D^*} + \sum_{i \geqslant m+2} |a_j^{(i)}| \frac{z_i}{D^*} + \frac{|a_j^{(n+1)}|}{D^*}$$
$$\leqslant (n-m)X + Y,$$

since for all $i \geqslant m+2$, $z_i \leqslant D \leqslant D^*$.

References

[1] A. Baker, Transcendental Number Theory (Cambridge Univ. Press, 1975) 13.
[2] I. Borosh, A sharp bound for positive solutions of homogeneous linear Diophantine equations, Proc. AMS 60 (1976) 19–21.
[3] I. Borosh and L.B. Treybig, Bounds on positive integer solutions of linear Diophantine equations, Proc. AMS 55 (1976) 299–304.
[4] I. Borosh and L. B. Treybig, Bounds on positive integral solutions of linear Diophantine equations, II, Canad. Math. Bull. 22(3) (1979) 357–361.
[5] J.W.S. Cassels, An Introduction to Diophantine Approximation, Cambridge Tracts in Math., Phys. No. 45 (Cambridge Univ. Press, NY, 1957).
[6] M.R. Garey and D.S. Johnson, Computers and Intractability, a Guide to the Theory of NP-Completeness (Freeman, San Francisco, 1979).
[7] J. von zur Gathen and M. Sieveking, A bound on solutions of linear integer equalities and inequalities, Proc. AMS. 72 (1978) 155–58.
[8] M. Groschel, L. Lovasz and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, Combinatorica 1 (1981) 169–197.
[9] L.G. Kachian, A polynomial algorithm in linear programming, Dokl. Akad. Nauk SSSR 244 (1979) 1093–1096.
[10] H.W. Lenstra, Integer Programming with a fixed number of variables, Univ. of Amsterdam, Dept. of Math. Tech., Report 81–03 (1981).
[11] L.B. Treybig, Bounds in piecewise linear topology, Trans. AMS 201 (1975) 383–405.