JOURNAL OF COMPLEXITY 3, 372–387 (1987)

A Simplex Variant Solving an $m \times d$ Linear Program in $O(\min(m^2, d^2))$ Expected Number of Pivot Steps*

ILAN ADLER

Department of Industrial Engineering and Operations Research, University of California, Berkeley, California 94720

RICHARD M. KARP

Computer Science Division, University of California, Berkeley, California 94720

AND

RON SHAMIR

Department of Computer Science, School of Mathematics, Tel Aviv University, Israel

We present a variant of the Simplex method which requires on the average at most 2 $(\min(m, d) + 1)^2$ pivots to solve the linear program min c^Tx, $Ax \geq b$, $x \geq 0$ with $A \in R^{m\times d}$. The underlying probabilistic distribution is assumed to be invariant under inverting the sense of any subset of the inequalities. In particular, this implies that under Smale's spherically symmetric model this variant requires an average of no more than $2(d + 1)^2$ pivots, independent of m, where $d \leq m$.

1. INTRODUCTION

The Simplex Method for Linear Programming, originated by Dantzig in 1947, is one of the most frequently used algorithms in industry and gov-

* This paper originally appeared in December 1983 as Report No. UCB/CSD 83/158 of the Computer Science Division, University of California at Berkeley. The results were reported at the TIMS/ORSA meeting, Orlando, Florida, in November 1983.

0885-064X/87 $3.00

ernment. The ordinary measure of complexity of this method is the number of pivot steps it requires to solve a linear program, expressed as a function of the dimensions of the problem. Vast practical experience indicates that this function is linear, or at most polynomial (Dantzig, 1963; Kuhn and Quandt, 1963). However, examples have been constructed for several variants of the Simplex method, showing that in the worst case the number of pivots may grow exponentially with the dimensions (Klee and Minty, 1972; Jeroslow, 1973; and others). Karmarkar's (1984) algorithm has a worst-case running time which is polynomial in the length of the problem data, and appears to be competitive with, or even superior to, the Simplex method on many classes of problems.

Recently, several works have tried to explain the efficiency of the Simplex method by approaching the complexity issue probabilistically: Assuming some distribution of the problem data, this approach tries to show that the *average* number of pivots grows slowly with the problem's dimensions. To quote these results denote the number of variables in the problem by d and the number of inequalities by n, and assume $d \leq n$. We use c to denote a constant and $c(d)$ to denote a function of d only. Borgwardt (1982a,b) showed that a parametric simplex variant requires an average of at most $c \cdot n \cdot d^2 \cdot (d + 1)^2$ pivots for a probabilistic model which generates only feasible linear programs. Smale (1983a) showed that the Parametric Self Dual Simplex requires an average of at most $c(d)$ $(\log(n - d))^{d(d+1)}$ pivots when the problem data are drawn from a spherically symmetric distribution. Megiddo (1986) improved that bound to $\bar{c}(d)$. This implies that the number of pivots tends to a finite limit when d is fixed and $m \to \infty$. However, this limit is superexponential in d. Adler (1983) and Haimovich (1983) demonstrated that some Parametric Simplex variants require an average of at most d steps once a vertex of the feasible region is given, but these results do not have immediate consequences for the full (Phase I–II) Simplex method.

We defined a family of Simplex variants which we called Constraint-By-Constraint (CBC) algorithms (Adler *et al.*, 1986). We showed that under probabilistic assumptions which are weaker than Smale's, these algorithms require an average of no more than $c(d)$ pivots where $c(d)$ is between $d \cdot 1.5^d$ and 2^{2d}, depending on the algorithm and the probabilistic model. In this paper we show that one of these variants, the Parametric-CBC algorithm, with proper initialization, requires at most $2(d + 1)^2$ pivots on the average, independent of m. Our probabilistic model requires that the problem data be nondegenerate and be generated by a distribution which is invariant under changing the sense of any subset of the inequalities defining the problem. Since Smale's probabilistic model satisfies these assumptions, this implies that the Parametric-CBC algorithm requires an average of at most $2(d + 1)^2$ pivots for Smale's model.

This result is one of three studies which obtain a quadratic bound on the expected number of pivots of a Simplex variant. The other two are by Todd (1983) and Adler and Megiddo (1983), who obtained an $O(d^2)$ bound on the expected number of pivots for the Self-Dual Simplex algorithm. The three studies make identical probabilistic assumptions and all three employ lexicographic pivot rules.

Megiddo (1985) has observed that, although the Parametric-CBC algorithm and the Self-Dual algorithm are in general quite different, their lexicographic versions execute exactly the same sequence of pivots. Thus all three investigations are concerned with the same Simplex variant. However, viewing this variant as a special case of the Parametric-CBC algorithm enabled us to apply the results of Adler (1983) and Haimovich (1983), and thereby to obtain a simple and direct quadratic bound.

2. Preliminaries

For a matrix $A \in R^{m\times d}$, we denote by A_i or $A_{i\cdot}$ the ith row of A, and by $A_{\cdot i}$ the ith column of A. If S is a sequence of indices of rows (columns), we denote by A_S $(A_{\cdot S})$ the submatrix obtained by taking only the rows (columns) in S.

The *Linear Programming Problem* is

$$(P)\quad \begin{aligned} &\min\ c^T x \\ &\text{s.t. } Ax \geq b \\ &x \geq 0, \end{aligned}$$

where $c, x \in R^d$, $b \in R^m$, $A \in R^{m\times d}$. The constraints of the form $A_{i\cdot}x \geq b_i$ are called *matrix constraints*, to be distinguished from the $x_i \geq 0$ *sign constraints*. Define also

$$M := \begin{bmatrix} I \\ A \end{bmatrix}, \qquad v := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b \end{bmatrix} \in R^{d+m}, \qquad n := d + m.$$

So an equivalent presentation of (P) is min c^Tx, $Mx \geq v$. Let $D \in R^{k+1}$, $j := \max(k, l)$, If every $j \times j$ submatrix of D is nonsingular we say that D is *strongly nondegenerate*.

The *Parametric Objective Problem* is

$$\min c^T x + \lambda \bar{c}^T x, \qquad c, \bar{c} \in R^d, \lambda \in R$$
$$Mx \geq v,$$

where we wish to find the optimal solutions for all values of the parameter λ. Here c is called the *objective* and $\bar{c}$ the *co-objective*.

This problem can be solved by the *Parametric Objective Algorithm* (Gass and Saaty, 1955; Dantzig, 1963), which is a variant of the Simplex method. The algorithm starts at a vertex optimal with respect to $c^T x$ in $F := \{x \mid Mx \geq v\}$, and (assuming nondegeneracy) when λ increases follows a connected path of edges and vertices of F. This path is called the *efficient path*. The union of the efficient paths for co-objectives $\bar{c}$ and $-\bar{c}$ is called the *co-optimal path*. We call a vertex or an edge of F $(c, \bar{c})$*-co-optimal* if it is on the co-optimal path generated when c is the objective and $\bar{c}$ the co-objective.

Every inequality of the form $M_i x \geq v_i$ can be thought of as a half-space in R^d determined by the hyperplane $M_i x = v_i$ and a sign choice with respect to that hyperplane. The opposite sign choice would yield the inequality $M_i x \leq v_i$. Given k hyperplanes in R^d, $k \geq d$, every one of the 2^k possible sign combinations determines a constraint set or an *instance*. Assuming nondegeneracy, every instance is either infeasible or d-dimensional, in which case it is called a *cell*.

Let $S = \{s_1, \ldots, s_d\} \subset \{1, \ldots, k\}$ be a set of d distinct indices of hyperplanes. S is called a *basic sequence* of M. Denote

$$M_s := \begin{pmatrix} M_{s_1} \\ \vdots \\ M_{s_d} \end{pmatrix}, \qquad v_s := \begin{pmatrix} v_{s_1} \\ \vdots \\ v_{s_d} \end{pmatrix}.$$

If $\det(M_s) \neq 0$, then these hyperplanes intersect in *vertex* $\bar{x} = (M_s)^{-1} v_s$. In that case we also say that S is the *basis* corresponding to $\bar{x}$. Under nondegeneracy this is a one-to-one correspondence.

Adler (1983) and Haimovich (1983) showed that (assuming nondegeneracy) every vertex of the arrangement of hyperplanes is co-optimal in exactly $(d + 1)$ of the 2^d cells incident on it.

A convenient way to present the sign choices is a sign matrix. A $k \times k$ matrix J is called a *k-sign matrix* (denoted $J \in \mathrm{SM}(k)$) if

$$J_{ij} = \begin{cases} \pm 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Thus for every matrix $A \in R^{k \times d}$

$$(JA)_{i\cdot} = \begin{cases} A_{i\cdot} & \text{if } J_{ii} = +1 \\ -A_{i\cdot} & \text{if } J_{ii} = -1. \end{cases}$$

So all the instances determined by the generating hyperplanes $M_i x = v_i$, $i = 1, \ldots, m$, can be represented by

$$JM \geq Jv, \qquad J \in \mathrm{SM}(m)$$

and clearly $|\mathrm{SM}(m)| = 2^m$.

3. The Algorithm

Several variants of Constraint-By-Constraint algorithms, as well as proofs of their validity, were presented in Adler *et al.* (1986). We shall briefly state the Parametric-CBC (PCBC) algorithm here, and refer the reader to Adler *et al.* (1986) for details and proofs.

For the Linear Program min $c^T x$, $Mx \geq v$, where $M \in R^{n \times d}$ $(m = n - d)$, define

$$X^{(k)} := \{x \in R^d \mid M_i x \geq v_i,\ i = 1, \ldots, d + k\}, \qquad k = 0, 1, \ldots, m$$
$$X := X^{(m)}.$$

Stage 0: Let $\bar{x}$ be the unique vertex of $X^{(0)}$. Choose $\bar{c} \in R^d$ such that the unique minimum of $\bar{c}^T x$ in $X^{(0)}$ is at $\bar{x}$. Go to Stage 1.

Stage k $(1 \leq k \leq m)$: Starting at $\bar{x}$ which minimizes $\bar{c}^T x$ in $X^{(k-1)}$, use the parametric objective algorithm to solve $\min\{\bar{c}^T x - \theta M_k x \mid x \in X^{(k-1)}\}$. Stop at the first point $\hat{x}$ along the path satisfying $M_k \hat{x} \geq v_k$. Set $\bar{x} := \hat{x}$ and go to stage $k + 1$. (If there is no such point along the path—Stop. $X = \varphi$).

Stage $m + 1$: Starting at $\bar{x}$ which minimizes $\bar{c}^T x$ in X, use the parametric objective algorithm to solve $\min\{\bar{c}^T x + \theta c^T x \mid x \in X\}$. The endpoint of the path yields the required solution. It may be either an optimal vertex or a ray demonstrating that the solution is unbounded.

If the problem includes nonnegativity constraints, we shall consider them the first d constraints, so that $X^{(0)} = R_+^d$. In that case we may choose in step 0 as the starting objective any $\bar{c} \in R^d$ satisfying $\bar{c} \geq 0$. Similarly if the sign constraints $Jx \geq 0$ where $J \in \mathrm{SM}(d)$ are included, we can choose any Je with $e > 0$ as the starting objective. To prove our result we shall use the objective

$$e := e(\varepsilon) := (\varepsilon, \varepsilon^2, \ldots, \varepsilon^d),$$

where ε is positive and sufficiently small.

For a linear program in form (P), we shall choose to solve either the primal or the dual problem, so that the algorithm is always performed with $d \leq m$.

4. The Probabilistic Model

We assume that the data (A, b, c) are generated according to a probability distribution satisfying the following properties:

(a) For fixed (A, b, c), all sign combinations of the inequalities

$$A_{i\cdot}x\ (\leq \text{ or } \geq)\ b_i, \qquad i = 1, \ldots, m$$
$$x_j (\leq \text{ or } \geq)\ 0, \qquad j = 1, \ldots, d$$

are equiprobable. In other words, 2^{d+m} equiprobable instances $J = [J^1, J^2]$ are generated according to $J^1 \in \mathrm{SM}(d)$, $J^2 \in \mathrm{SM}(m)$ all having the form

$$\begin{aligned} &\min c^T x \\ &J^1 x \geq 0 \\ &J^2 A x \geq J^2 b. \end{aligned}$$

Note that this condition is equivalent to the statement that all instances generated by inverting signs of rows of $[A, b]$ and columns of $\begin{bmatrix} c^T \\ A \end{bmatrix}$ are equiprobable.

(b) With probability one, both

$$\begin{bmatrix} c^T \\ A \\ I \end{bmatrix} \qquad \text{and} \qquad [I, A, b]$$

are strongly nondegenerate. Smale (1983a) assumes that the data are obtained from a spherically symmetric distribution. Since every such distribution satisfies properties (a) and (b) our results will hold in particular for his model. Our model, however, need not assume that the distribution is continuous. In fact, it may generate a finite set of 2^{m+d} linear program instances, all corresponding to the same strongly non-degenerate data (A, b, c).

The advantage of the model described above is that it enables one to reduce the probabilistic analysis to combinatorial analysis. The same kind of model was used by May and Smith (1982) for investigating random polytopes, and by Adler and Berenguer (1981, 1983) for investigating several issues in random linear programs.

5. Analysis

Consider data (A, b, c) satisfying our probabilistic model assumptions. These data induce 2^{m+d} equiprobable instances, corresponding to all sign combinations of the inequalities. In stage $k + 1$ of the Parametric-CBC algorithm, $d + k$ constraints are present, and they induce 2^{k+d} equiprobable instances. An instance may be represented by the sign combination $J = [J^1, J^2]$ it uses, namely

$$J^1 x \geq 0$$
$$J^2 \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix} \geq J^2 \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix},$$

where $J^1 \in \mathrm{SM}(d)$, $J^2 \in \mathrm{SM}(k)$. Denote those instances by $J_1, \ldots, J_{2^{d+k}}$.

All instances in stage $k + 1$ use $-A_{k+1}$ as the co-objective. The starting objective $\bar{c}$ used in instance $J_k = [J_k^1, J_k^2]$ was determined in stage 0 to be $\bar{c}(J_k) := J_k^1 e$. Denote these objectives by $e_1, \ldots, e_{2^d}$.

Consider a fixed basis S, corresponding to a vertex $\bar{x}$ in stage $k + 1$. Define

$$F(J_i, e_j, S) := \begin{cases} 1 & \text{if basis } S \text{ is } (e_j, -A_{k+1})\text{-co-optimal in instance } J_i. \\ 0 & \text{otherwise} \end{cases}$$

We are interested in the number of pivots actually performed by the PCBC algorithm in all the instances in stage $k + 1$. An upper bound to this number is

$$G(k) := \sum_S \sum_{J_i} F(J_i, J_i^1 e, S).$$

That is, for every basis S we wish to count a pivot in instance J_i only if the objective prescribed by the algorithm to be used in instance J_i makes S co-optimal.

Adler (1983) and Haimovich (1983) proved that for every basis S and for every objective e_j, assuming nondegeneracy,

$$\sum_{J_i} F(J_i, e_j, S) = d + 1.$$

From this we get the basic result

$$G(k) \leq \sum_S \sum_{e_j} \sum_{J_i} F(J_i, e_i, S) = \binom{k+d}{d} 2^d(d+1),$$

which was used in by Adler *et al.* to obtain the bound $(d+1)2^{d+1}$. We shall now improve upon that result using the special structure of the vector $e = e(\varepsilon)$.

We say that a basis S is of *type r* if $\{1, \ldots, r\} \subset S$, $\{r+1\} \notin S$, in other words, if the tight inequalities for that basis include $x_1 = 0, \ldots, x_r = 0$, but do not include $x_{r+1} = 0$. The main observation we shall use is the following:

THEOREM. *If S is of type r and there exists e_j such that $F(J_i, e_j, S) = 1$ then $\sum_{e_j} F(J_i, e_j, S) \geq 2^{d-r-1}$ for sufficiently small ε.*

We leave the proof to the end of this section. Let us first show how we use this theorem to get our main result: Define

$$G(J_i, S) := \begin{cases} 1 & \text{if } \sum_{e_j} F(J_i, e_j, S) \geq 1 \\ 0 & \text{otherwise;} \end{cases}$$

clearly

$$F(J_i, J_i^! e, S) \leq G(J_i, S) \qquad \forall\, J_i; S.$$

If $\hat{S}$ is of type r, then the theorem implies that

$$G(J_i, \hat{S}) \leq \left\{\sum_{e_j} F(J_i, e_j, \hat{S})\right\} 2^{-(d-r-1)}.$$

Now let us consider separately each type of S in the sum generating $G(k)$:

$$\begin{aligned} G(k) &= \sum_{r=0}^{d} \sum_{\substack{S \\ \text{of type } r}} \sum_{J_i} F(J_i, J_i^! e, S) \\ &\leq \sum_{r=0}^{d} \sum_{\substack{S \\ \text{of type } r}} \sum_{J_i} G(J_i, S) \end{aligned}$$

$$\leq \sum_{r=0}^{d} \sum_{\substack{S \\ \text{of type } r}} \sum_{J_i} \left\{ \sum_{e_j} F(J_i, e_j, S)\, 2^{-(d-r-1)} \right\}$$

$$= \sum_{r=0}^{d} \sum_{\substack{S \\ \text{of type } r}} \left\{ \sum_{e_j} \sum_{J_i} F(J_i, e_j, S) \right\} 2^{-(d-r-1)}$$

$$= \sum_{r=0}^{d} \binom{k+d-r-1}{d-r} \{2^d(d+1)\}\, 2^{-(d-r-1)}$$

$$= 2^d(d+1) \sum_{t=0}^{d} \binom{k+t-1}{t} 2^{-(t-1)}.$$

Since all 2^{k+d} instances in stage $k+1$ are equiprobable, we get that the average number of pivots required to solve an $m \times d$ problem using the PCBC is

$$\rho(m, d) \leq \sum_{k=1}^{m+1} G(k)\, 2^{-(k+d)} \leq (d+1) \sum_{k=1}^{m+1} \sum_{t=0}^{d} \binom{k+t-1}{t} 2^{-(k+t-1)}$$

$$= (d+1) \sum_{t=0}^{d} \sum_{k=1}^{m+1} \binom{k+t-1}{t} 2^{-(k+t-1)}$$

$$\leq (d+1) \sum_{t=0}^{d} \sum_{j=t}^{\infty} \binom{j}{t} 2^{-j} = 2(d+1)^2.$$

This result was proved for fixed strongly nondegenerate data, and is independent of those data. Since our model generates strongly nondegenerate data with probability one, the main result follows.

Before proving the theorem, let us first prove two short lemmas:

LEMMA 1. *The Strong Nondegeneracy property is preserved under pivotal transformations.*

Proof. Since a pivotal transformation is nonsingular, the rank of every $d \times d$ submatrix is preserved under such transformations. So if every $d \times d$ submatrix of the original matrix is nonsingular, the same is true after the pivotal transformation. ∎

COROLLARY 1. *In the nonbasic part of the transformed matrix described in Lemma* 1, *every $l \times l$ submatrix is nonsingular, where* $l = 1, \ldots, d$.

Proof. For $l = d$ this follows directly from Lemma 1. Let Q be an $l \times l$ submatrix of the nonbasic part of the transformed matrix, with $l < d$. By adding some unit columns from the basic part (and reordering rows and columns if necessary), a $d \times d$ matrix of the form $\begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix}$ can be generated. By Lemma 1 this matrix is nonsingular, hence Q is nonsingular. ■

LEMMA 2. *Let* $P(\varepsilon) := \varepsilon^p(\alpha + \sum_{i=1}^{k} \alpha_i \varepsilon^i)$ with $\alpha \neq 0$. If for $\varepsilon > 0$ sufficiently small $P(\varepsilon) > 0$, then for every polynomial $\bar{P}(\varepsilon)$ obtained from $P(\varepsilon)$ by changing the signs of some of the coefficients α_i, also $\bar{P}(\varepsilon) > 0$.

Proof. Since $\varepsilon > 0$, also $\alpha + \sum_{i=1}^{k} \alpha_i \varepsilon^i > 0$. Since that is true for sufficiently small ε, then by continuity at $\varepsilon = 0$ we get $\alpha \geq 0$. Since $\alpha \neq 0$ we get $\alpha > 0$. Hence also $\alpha + \sum_{i=1}^{k} (\pm\alpha_i)\varepsilon^i > 0$ for ε sufficiently small, for every possible sign combination. ■

Proof of Theorem

To prove the theorem, we shall show that the following holds for sufficiently small ε:

Let S be of type r and let J_k be any instance. Let $\bar{e}$ be any objective out of $e_1, \ldots, e_{2d}$. If S is $(\bar{e}, a)$-co-optimal in J_k, then for every $\tilde{e} = (\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_{r+1}, J(\bar{\varepsilon}_{r+2}, \ldots, \varepsilon_d))$, $J \in \mathrm{SM}(d - r - 1)$, S is also $(\tilde{e}, a)$-co-optimal in J_k.

Since $|\mathrm{SM}(d - r - 1)| = 2^{d-r-1}$ this will complete the proof.

For S of type r, the corresponding basis takes the form

$$\begin{array}{rcc|cl} & & r & d - r & \\ B = & \Big[& \begin{array}{c} I \\ \hline D^T \end{array} & \begin{array}{c} 0 \\ \hline E^T \end{array} & \Big] \begin{array}{l} r \\ d - r \end{array} \end{array}$$

(The numbers above and to the right of the matrix are the dimensions of the corresponding submatrices.) In order for S to be $(\bar{e}, a)$-co-optimal there must be some $\theta \in R$ satisfying

$$(B^T)^{-1}(\bar{e} + \theta a) > 0,$$

where the strict inequalities are implied by the nondegeneracy of the data.

Let us express the inverse explicitly:

$$(B^T)^{-1} = \left[\begin{array}{c|c} I & D \\ \hline 0 & E \end{array}\right]^{-1} = \begin{array}{cc} r & d-r \\ \left(\begin{array}{c|c} I & -DE^{-1} \\ \hline 0 & E^{-1} \end{array}\right) & \end{array}$$

$$=: \begin{array}{ccc|l} r & 1 & d-r-1 & \\ \left(\begin{array}{c|c|c} I & v & F \\ \hline 0 & w & G \end{array}\right) & & & \begin{array}{l} r \\ d-r \end{array} \end{array}$$

The last equality is introduced to define the first non-unit column of the inverse matrix. Define also $u := (B^T)^{-1}a$. So the condition

$$\exists\, \theta \text{ s.t. } (B^T)^{-1}(\bar{e} + \theta a) > 0$$

is equivalent to

$$\exists\, \theta \text{ s.t. } \left(\begin{array}{c|c|c} I & v & F \\ \hline 0 & w & G \end{array}\right) \bar{e} + \theta u > 0$$

or, redefining $\bar{e} =: (\bar{e}^1, \bar{e}_{r+1}, \bar{e}^2)$, where $\bar{e}^1 \in R^r$, $\bar{e}_{r+1} \in R$, $\bar{e}^2 \in R^{d-r-1}$,

$$\exists\, \theta \text{ s.t. } \begin{pmatrix} \bar{e}^1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \bar{e}_{r+1} \begin{pmatrix} v \\ w \end{pmatrix} + \begin{pmatrix} F \\ G \end{pmatrix} \bar{e}^2 + \theta u > 0. \tag{1}$$

Using the Fourier–Motzkin Elimination method, the system (1) has a solution θ if and only if the following system has a solution

$$\frac{1}{u_i}\left[\begin{pmatrix} \bar{e}^1 \\ 0 \end{pmatrix} + \bar{e}_{r+1} \begin{pmatrix} v \\ w \end{pmatrix} + \begin{pmatrix} F \\ G \end{pmatrix} \bar{e}^2\right]_i - \frac{1}{u_j}\left[\begin{pmatrix} \bar{e}^1 \\ 0 \end{pmatrix} + \bar{e}_{r+1} \begin{pmatrix} v \\ w \end{pmatrix} + \begin{pmatrix} F \\ G \end{pmatrix} \bar{e}^2\right]_j > 0$$

$$\forall_i \text{ s.t. } u_i > 0$$

$$\forall_j \text{ s.t. } u_j < 0 \tag{2.1}$$

$$\left[\begin{pmatrix} \bar{e}^1 \\ 0 \end{pmatrix} + \bar{e}_{r+1} \begin{pmatrix} v \\ w \end{pmatrix} + \begin{pmatrix} F \\ G \end{pmatrix} \bar{e}^2\right]_k > 0 \qquad \forall_k \text{ s.t. } u_k = 0. \tag{2.2}$$

Now we want to show that if (2) holds for sufficiently small ε, every change in the signs of the coordinates of the vector $\bar{e}^2$ will still keep the system (2) valid. By the equivalence of the systems (1) and (2) this will complete the proof.

Since the data are strongly nondegenerate, Corollary 1 (with $l = 1$) implies $u_k \neq 0$ for all k, so all the inequalities in (2) are of the form (2.1). These can be partitioned into two types:

Type 1. $i \leq r$ or $j \leq r$. The inequalities here take the form

$$\alpha\varepsilon^k + \varepsilon^{k+1}P(\varepsilon) > 0,$$

where $k = \min(i, j) \leq r$, $\alpha \neq 0$, and $P(\varepsilon)$ is a polynomial in ε. By Lemma 2 the result follows.

Type 2. $i, j > r$. These inequalities have the form

$$\pm\varepsilon^{r+1}\left[\frac{w_i}{u_i} - \frac{w_j}{u_j}\right] + \varepsilon^{r+2}P(\varepsilon) > 0.$$

By the strong nondegeneracy of the data, Corollary 1 (with $l = 2$) implies $w_i/u_i - w_j/u_j \neq 0$. So again the conditions of Lemma 2 hold and the result follows. ∎

6. Concluding Remarks

1. The PCBC algorithm can be implemented by a lexicographic rule, without any explicit use of ε. The reasoning behind this is as follows:

Let the current co-optimal basis in stage $k + 1$ be B^T. The co-objective is $a := -A_{k+1}$ and without loss of generality assume the objective is e. Let $\bar{a} := B^{-1}a$, $\bar{e} := B^{-1}e$. According to the Parametric Objective Algorithm, the next variable to leave the basis is determined by the ratio test:

$$\min_{i \in N}\left\{\frac{\bar{e}_i}{-\bar{a}_i}\right\}, \qquad \text{where } N := \{i \mid \bar{a}_i < 0\}.$$

Recall that

$$\bar{e}_i = (B^{-1}e)_i = B_i^{-1}e = B_{i,1}^{-1}\varepsilon + B_{i,2}^{-1}\varepsilon^2 + \cdots + B_{i,d}^{-1}\varepsilon^d.$$

Because B lies on the efficient path $\bar{e}_i > 0 \ \forall_{i \in N}$. Let $l(i)$ be the index of the first nonzero element in row B_i^{-1}. If $\varepsilon > 0$ and is sufficiently small, this can happen only if, for all $i \in N$, $B_{i,l(i)}^{-1} > 0$. In other words, the matrix $B_{N\cdot}^{-1}$ is *lexico-positive*. Again, for $\varepsilon > 0$ sufficiently small, if $i, j \in N$ and $l(i) > l(j)$ then j cannot be the minimizing index in the ratio test. So only indices i with $l(i) = t := \max\{l(j) | j \in N\}$ are candidates to win the ratio test. Since they all have the form $\bar{e}_i = B_{i,t}^{-1}\varepsilon^t + \cdots + B_{i,d}^{-1}\varepsilon^d$, we need to perform the ratio test:

$$\min_{\substack{i \in N \\ l(i) \text{ maximal}}} \left\{ \frac{B_{i,l(i)}^{-1}}{-\bar{a}_i} \right\}.$$

(The strong nondegeneracy guarantees that there will be no ties in this test, so we never need to go beyond the leading elements to compare $B_{i,l(i)+1}^{-1}/(-\bar{a}_i)$ and so on.) So the leaving variable is the one which *lexicographically minimizes*

$$\left\{ \frac{B_i^{-1}}{-\bar{a}_i} \,\middle|\, i \in N \right\},$$

and no ε is involved in the actual implementation.

2. The result obtained here holds also for other forms of linear programs. In Adler *et al.* (1986) we showed that presenting the linear program with nonnegativity constraints is immaterial. The essential requirement is sign invariance with respect to each of the constraints present in the problem. In general, for a problem with n arbitrary constraints and d variables, under our model assumptions the expected number of pivots is no more than $2(\min(d, n - d) + 1)^2$. By duality this also yields similar results for linear programs with equality constraints.

3. A desirable extension of our analysis is to relax the strong nondegeneracy assumption. This may allow improvement over the results obtained in Adler *et al.* (1986) for structured linear programs which yield sparse matrices, e.g., ones arising from transportation problems.

4. The crucial property for our analysis is sign invariance, which may be interpreted as a special kind of symmetry around the origin. However, in Smale (1983b), it is stated that the results proven in Smale (1983a) can be obtained on the weaker assumption of invariance under coordinate permutations. Blair (1986) proceeded to show that indeed the requirement that the distribution be continuous can also be removed and still the permutation-invariance assumption yields essentially the same results as Smale's. However, that result required that $m \gg d$. It may be interesting to extend our model in that direction.

5. A troublesome point which we discussed in Adler *et al.* (1986), and which was mentioned in several previous papers, is the behavior of the sign-invariant model in the case where the ratio of the dimensions is very far from 1: For $m \gg d$, all but a vanishing fraction of the linear programs generated by the model will be infeasible. Similarly, for $m \ll d$, almost all problems generated will be feasible and unbounded. Both of these categories of linear programs appear to be easier to solve than problems which

are feasible and bounded. That may indicate that this model is inadequate for obtaining meaningful results in such a situation.

This does not seem to be a severe limitation of our results, since they are also meaningful when m and d are approximately equal. In particular, in the case $m = d$ for which Klee and Minty (1972) demonstrated a worst-case behavior which is exponential in d, we get that the average-case behavior of the algorithm is at most quadratic in d. Also, when $m = d$ the proportion of cells containing optimal solutions out of all instances is $\Omega(d^{-1/2})$. So the expectation of the number of pivots per instance, conditioning on the instance having an optimal solution, is $O(d^{2.5})$.

Still, investigating a model which generates only feasible and bounded problems seems an interesting next step.

6. Can similar results be achieved for other variants of the Simplex method, besides the one investigated by Todd (1983), by Adler and Megiddo (1983), and by this work? (It is interesting to note that most of the probabilistic results obtained so far—perhaps with the exception of those of Adler *et al.* (1986) and Blair (1986)—have used parametric variants of the Simplex method.)

7. Another interesting open question is obtaining higher moments of the random variable investigated, especially obtaining the variance of the number of pivots.

Except for rough bounds (like $Pr[\text{no. of pivots} \geq a2(d + 1)^2] \leq 1/a$, obtained from Markov's inequality), it seems that new proof techniques and perhaps stronger probabilistic assumptions are required for such analysis.

8. By more refined analysis, our bounds on the number of pivots can be reduced by a constant factor. However, a reduction of the degree of the polynomial is impossible. This is a consequence of a recent study of Adler and Megiddo (1985): Under stronger probabilistic assumptions they have extended their work on the Self-Dual Simplex variant (1983) to obtain a *quadratic lower bound* on the average number of pivots. Since their Simplex variant was proven in Megiddo (1985) to be identical to the PCBC algorithm with lexicographic initialization vectors, this implies that the upper bound for the weaker model discussed here cannot be subquadratic. The computation of a tight lower bound under the weaker, more general model when using the PCBC algorithm is still to be carried out.

ACKNOWLEDGMENT

We thank Nimrod Megiddo for several stimulating discussions, and in particular for drawing to our attention the possibility of using lexicographic pivot rules.

REFERENCES

ADLER, I. (1983), "The Expected Number of Pivots Needed to Solve Parametric Linear Programs and the Efficiency of the Self-Dual Simplex Method," Working Paper, Department of Industrial Engineering and Operations Research, University of California, Berkeley, June.

ADLER, I., AND BERENGUER, S. E. (1981), "Random Linear Programs," Operations Research Center Report No. 81-4, University of California, Berkeley.

ADLER, I., AND BERENGUER, S. E. (1983), Generating random linear programs, manuscript, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

ADLER, I., KARP, R. M., AND SHAMIR, R. (1986), A family of simplex variants solving an $m \times d$ linear program in expected number of pivots depending on d only, *Math. Oper. Res.* **11,** No. 4 (Nov.), 570–590.

ADLER, I., AND MEGIDDO, N. (1983), A simplex-type algorithm solves linear programs of order $m \times n$ in only $O((\min(m, n))^2)$ steps on the average, preliminary report, Nov.

ADLER, I., AND MEGIDDO, N. (1985), A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension, *J. Assoc. Comput. Mach.* **32,** No. 4, 871–895.

BLAIR, C. (1986), Random linear programs with many variables and few constraints, *Math. Programming* **34,** No. 1, 62–71.

BORGWARDT, K. H. (1982a), Some distribution-independent results about the asymptotic order of the average number of pivot steps of the simplex method, *Math. Oper. Res.* **7,** No. 3 (Aug.), 441–462.

BORGWARDT, K. H. (1982b), The average number of pivot steps required by the simplex-method is polynomial, *Z. Oper. Res.* **26,** 157–177.

DANTZIG, G. B. (1963), "Linear Programming and Extensions," Princeton Univ. Press, Princeton, NJ.

GASS, S. I., AND SAATY, E. L. (1955), The computational algorithm for the parametric objective function, *Naval Res. Logist. Quart.* **2,** No. 1 (June), 39–45.

HAIMOVICH, M. (1983), The simplex algorithm is very good!—On the expected number of pivot steps and related properties of random linear programs, Columbia University, New York, April.

JEROSLOW, R. G. (1973), The simplex algorithm with the pivot rule of maximizing improvement criterion, *Discrete Math.* **4,** 367–378.

KARMARKAR, N. (1984), A new polynomial-time algorithm for linear programming, *Combinatorica* **4,** 373–395.

KLEE, V., AND MINTY, G. J. (1972), How good is the simplex algorithm? *in* "Proceedings, 3rd Symposium on Inequalities" (O. Shisha, Ed.), pp. 159–175, Academic Press, New York.

KUHN, H. W., AND QUANDT, R. E. (1963), An experimental study of the simplex method, *in* "Proceedings, 15th Symposium in Applied Mathematics" (Metropolis *et al.*, Eds.), pp. 107–124, Amer. Math. Soc., Providence, RI.

MAY, J. H., AND SMITH, R. L. (1982), Random polytopes: Their definition, generalization and aggregate properties, *Math. Programming* **24,** 39–54.

MEGIDDO, N. (1985), A note on the generality of the self-dual algorithm with various starting points, *Methods Oper. Res.* **49,** 271–275.

MEGIDDO, N. (1986), Improved asymptotic analysis of the average number of steps performed by the self dual simplex algorithm, *Math. Programming* **35,** No. 2, 140–172.

Smale, S. (1983a), On the average speed of the simplex method of linear programming, *Math. Programming* **27**, 241–262.

Smale, S. (1983b), The problem of the average speed of the simplex method, *in* "Mathematical Programming: The State of the Art" (Bachem, Grotschel, and Korte, Eds.), pp. 530–539, Springer-Verlag, Berlin/New York.

Todd, M. J. (1983), "Polynomial Expected Behaviour of a Pivoting Algorithm for Linear Complementary and Linear Programming Problems," Technical Report No. 595, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, Nov.