A Simplex Algorithm Whose Average Number of Steps Is Bounded between Two Quadratic Functions of the Smaller Dimension

ILAN ADLER

University of California, Berkeley, California

AND

NIMROD MEGIDDO

IBM Almaden Research Center, San Jose, California and Tel Aviv University, Tel Aviv, Israel

Abstract. It has been a challenge for mathematicians to confirm theoretically the extremely good performance of simplex-type algorithms for linear programming. In this paper the average number of steps performed by a simplex algorithm, the so-called *self-dual method*, is analyzed. The algorithm is not started at the traditional point $(1, \ldots, 1)^T$, but points of the form $(1, \epsilon, \epsilon^2, \ldots)^T$, with ϵ sufficiently small, are used. The result is better, in two respects, than those of the previous analyses. First, it is shown that the expected number of steps is bounded between two quadratic functions $c_1(\min(m, n))^2$ and $c_2(\min(m, n))^2$ of the *smaller* dimension of the problem. This should be compared with the previous two major results in the field. Borgwardt proves an upper bound of $O(n^4 m^{1/(n-1)})$ under a model that implies that the zero vector satisfies all the constraints, and also the algorithm under his consideration solves only problems from that particular subclass. Smale analyzes the self-dual algorithm starting at $(1, \ldots, 1)^T$. He shows that for any fixed m there is a constant $c(m)$ such the expected number of steps is less than $c(m)(\ln n)^{m(m+1)}$; Megiddo has shown that, under Smale's model, an upper bound $C(m)$ exists. Thus, for the first time, a polynomial upper bound with no restrictions (except for nondegeneracy) on the problem is proved, and, for the first time, a nontrivial lower bound of precisely the same order of magnitude is established. Both Borgwardt and Smale require the input vectors to be drawn from spherically symmetric distributions. In the model in this paper, invariance is required only under certain

---

This paper extends our previous report entitled "A Simplex-Type Algorithm Solves Linear Programs of Order $m \times n$ in Only $O((\min(m, n))^2)$ steps on the Average," November 1983. The present paper adds the lower bounding part of the result. A result similar to our earlier upper bounding part was independently obtained in "Polynomial Expected Behavior of a Pivoting Algorithm for Linear Complementarity and Linear Programming Problems," by M. J. Todd, Tech. Rep. No. 595, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, November 1983. Also related is Report UCB CSD 83/158, "A Simplex Variant Solving an $m \times d$ Linear Program in $O(\min(m^2, d^2))$ Expected Number of Pivot Steps," by Adler, Karp, and Shamir, Computer Science Division, University of California, Berkeley, December 1983. An earlier version of the present paper appeared in *Proceedings of the 16th Annual ACM Symposium on Theory of Computing* (Washington, DC, Apr. 30–May 2). ACM, New York, 1984, pp. 312–323.

This research was done while the second author was visiting Stanford University and XEROX Palo Alto Research Center. The work of the second author was supported in part by the National Science Foundation under grants MCS-83-00984, ECS-81-21741, and ECS-82-18181.

Authors' addresses: I. Adler, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720; N. Megiddo, IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120-6099.

reflections or permutations and not under every possible rotation. The fact that ϵ has to be sufficiently small raises no difficulties whatsoever. The algorithm can either determine the correct value while solving the problem, or simply operate on ϵ symbolically, using "lexicographic" rules.

Categories and Subject Descriptors: F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems—*computations on matrices*; G.1.6 [**Numerical Analysis**]: Optimization—*linear programming*

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Lexicographic pivoting, probabilistic analysis of algorithms, simplex algorithms

1. *Introduction*

The "simplex" algorithm for linear programming, which was developed by Dantzig [9], is not just a single algorithm but, as matter of fact, a class of algorithms whose common feature is that they iteratively change the basis of a linear system of equations, until they reach an "optimal" basis or a basis that exhibits that no optimal solution exists. For a linear programming problem with n (nonnegative) variables and m constraints, the number of bases is $\binom{m+n}{m}$; hence this quantity is an obvious upper bound on the number of steps that any simplex-type algorithm can perform. However, the vast computational experience accumulated to date has shown that the number of steps is usually much smaller. This has been observed while solving practical problems, as well as those generated in a laboratory. It has been a challenge to confirm these findings theoretically. Tremendous effort has been made in the direction of studying properties of convex polyhedra that are related to linear programming. However, it is known that many simplex-type algorithms may require exponential number of steps in the worst case. The first example to this effect was given by Klee and Minty [11], and Murty [16] provided an example in the context of the self-dual method. Similar examples are known for several other variants of the simplex method.

Borgwardt [6, 7] and Smale [17, 18] have recently provided probabilistic analyses of simplex-type algorithms. We note that an analysis of this type requires a specification of algorithms to which it applies, as well as probabilistic distributions of inputs. Both Borgwardt's and Smale's models assume that the vectors generating the problem are sampled from spherically symmetric distributions; however, Smale actually obtains his results under a weaker model of symmetry with respect to permutations of coefficients within rows. Borgwardt analyzes different variations on the radial part of his distributions, while under Smale's model the radial part is immaterial. Both of these analyses deal with "parametric" simplex algorithms, and this is, apparently, a key property for carrying out a probabilistic analysis.

In order to understand the contribution of the present paper, we first state the results of Borgwardt and Smale. Borgwardt considers the problem in the form

$$\begin{aligned} &\text{maximize} && c^{\mathrm{T}}x \\ &\text{subject to} && Ax \leq e \end{aligned}$$

(where $x, c, \in R^n$, $A \in R^{m\times n}$ and $e = (1, \ldots, 1)^{\mathrm{T}} \in R^m$). The rows of A, as well as the vector c, are distributed spherically symmetrically over their respective spaces. Under this model the zero vector satisfies the inequalities. Note that under this model *every* subproblem, determined by a subset of the columns, has to be feasible. Indeed, every problem, which is given together with a feasible solution, can easily

be transformed into Borgwardt's form, but the probabilistic assumptions can hardly be justified afterward. The algorithm is a certain parametric simplex method, with a special initialization procedure that is necessary only for the mathematical reasoning, and capitalizes on the fact that the zero vector is feasible. Therefore, the algorithm as a whole solves only problems from this particular subclass. It cannot explain the so-called Phase 1 of linear programs. Under this model, Borgwardt shows that the expected number of steps, $\rho^B(m, n)$, satisfies

$$\rho^B(m, n) \leq cn^4 m^{1/(n-1)},$$

where c is a certain constant. We note that this upper bound tends to infinity when either m or n tends to infinity.

Smale considers the problem in the form

$$\begin{array}{ll} \text{minimize} & c^{\mathrm{T}}x \\ \text{subject to} & Ax \geq b \\ & x \geq 0 \end{array}$$

(where $x, c \in R^n$, $A \in R^{m \times n}$, and $b \in R^m$). Under his model, the matrix A is distributed spherically symmetrically over $R^{m \times n}$ and the vector (b, c) is distributed (independently) spherically symmetrically over R^{m+n}. However, a weaker model is actually used for obtaining the result. The algorithm is the so-called self-dual simplex algorithm [9] (which can be viewed as a special case of "Lemke's algorithm" for the linear complementarity problem [12]). Lemke's algorithm requires a specification of a starting point in the positive orthant of R^{m+n}. Traditionally, as well as in Smale's model, the starting point is taken as $(1, \ldots, 1)^{\mathrm{T}}$, reproducing the "self-dual" version. Under this model, Smale shows that the expected number of steps, $\rho(m, n)$, satisfies the following condition: For every fixed m there exists a constant $c(m)$ such that for every n,

$$\rho(m, n) \leq c(m)(\ln n)^{m(m+1)}.$$

Obviously, this upper bound tends to infinity with n. Blair [5] proves that the expected number of undominated columns under an even more general model is less than $c(m)(\ln n)^{(m+1)\ln(m+1)+1}$, which implies an upper bound of $c(m)(\ln n)^{m(m+1)\ln(m+1)+m}$ on the number of pivot steps for a wider class of algorithms. We remark that bounds like those of Smale and Blair can be derived by estimating expected numbers of extreme points of the primal or the dual polytope. Obviously, the efficiency of the simplex method does not stem from a small number of extreme points, but rather from the fact that usually only few of these points occur on the path followed by the algorithm. Megiddo [14] has shown that, under Smale's model, for every m there is a constant $c(m)$ such that for all n

$$\rho(m, n) \leq c(m).$$

An upper bound on $c(m)$ depended exponentially on m.

In this paper we improve upon the previous results considerably. We confirm the observed phenomenon that the average number of steps is *polynomial in the smaller dimension* of the problem. We analyze the average number of pivot steps performed by the self-dual simplex algorithm with a different starting point. Instead of the point $(1, \ldots, 1)^{\mathrm{T}}$, we start the algorithm at $(1, \epsilon, \epsilon^2, \ldots)^{\mathrm{T}}$ with $\epsilon > 0$ sufficiently small. The powers of ϵ can be assigned arbitrarily. This algorithm is closely related to an algorithm proposed by van der Heyden [19]. It can operate on ϵ symbolically, or can, alternately, be stated with "lexicographic" rules. The

actual determination of ϵ does not raise any difficulties whatsoever. Incidentally, the algorithm itself can determine what is a sufficiently small ϵ by maintaining upper bounds under which the "lexicographic" arithmetics holds. The choice of the different starting point yields a better bound on the average number of steps, $\rho^{\epsilon}(m, n)$. We show that this number is bounded between two quadratic functions of the *minimum* of the two dimensions:

$$c_1(\min(m, n))^2 \leq \rho^{\epsilon}(\mathrm{m}, \mathrm{n}) \leq \mathrm{c}_2(\min(m, n))^2.$$

Thus, we obtain a nontrivial lower bound that seems to be in conflict with the common belief that the simplex algorithm performs, on the average, only linearly many steps. There has not been much experience with the particular variants discussed in this paper and that may account for the difference in the results. However, the reason why these variants behave quadratically is transparent in view of the present paper.

Our analysis for the upper bound is carried out under a model that is weaker than those of Borgwardt and Smale in the sense explained below. Instead of complete spherical symmetry, we require only symmetry with respect to certain reflections, together with a certain regularity condition on the matrix; this condition holds with probability one if the problem is sampled from any continuous distribution. For the lower bound we need a stronger model in which the entries are independent, identically distributed (symmetrically with respect to zero) random variables.

We also note that the *conditional* expectation of the number of steps, given that the problem has an optimal solution, can now be bounded from above by a low-order polynomial in the case usually considered most difficult, that is, $m = n$. The probability that the problem has an optimal solution is

$$\binom{m+n}{m} 2^{-m-n}$$

(see [1]). In case $m = n$ this is of order $m^{-1/2}$. Thus, the conditional expectation of the number of steps in this case is $O(m^{2.5})$. Also, an obvious consequence of our result is that the probability that the algorithm will require an exponential number of steps is exponentially decreasing to zero. However, we expect a stronger result to be obtained by a more careful look into the distribution of the number of steps.

We discuss the model in Section 2. The algorithm is described in Section 3. In Section 4 we describe the four cases to be distinguished in the analysis of the probability of a basis to occur in the solution process. The upper bounds for these cases are then analyzed in two pairs in Sections 5 and 6. In Section 7 we prove the lower bound result. The specific upper and lower bounds are summarized in Section 8.

2. *The Probabilistic Model*

For an "average-case" analysis, with results different from the "worst case," one has to make some assumptions on the distribution of problems. A probabilistic analysis does not have to assume a unique distribution of problems. It is more desirable to be able to prove good bounds that are valid for *any* distribution in a wide class. Notice that, under the model proposed by Smale, any spherically symmetric distribution has the same average-case complexity. However, one should

seek wider classes such that the average case is not necessarily the same for all the members of the class, but yet each satisfies some good bound.

It is natural to consider models with some symmetry assumptions. Very roughly, the hope is that, in a symmetric set of instances, if one is bad, then others should be good, so that the average over the set should not be bad. More specifically, suppose we have a group of symmetries and consider the equivalence classes of instances that are invariant under the group. Suppose the average over each equivalence class is bounded nicely. Then, regardless of how a class is picked, provided an instance is adequately selected from the class, the overall average will be nicely bounded. Subject to this terminology, it is desirable to have the "classes" as small as possible, that is, the group of symmetries as small as possible. Under the spherically symmetric model, two instances (A_1, b_1, c_1) and (A_2, b_2, c_2) are in the same equivalence-class if (i) the matrix A_2 can be obtained from A_1 by an orthogonal transformation (of $R^{m\times n}$) followed by a multiplication by a positive constant, and (ii) the vectors (b_1, c_1) and (b_2, c_2) are related in a similar fashion. Obviously, each class contains a continuum of instances.

Under our model the classes are finite. Given an instance (A, b, c), it is convenient in the present section to consider an $(m + 1) \times (n + 1)$ matrix A^* such that $A^*_{ij} = A_{ij}$ $(i = 1, \ldots, m, j = 1, \ldots, n)$, $A^*_{m+1,j} = c_j$ $(j = 1, \ldots, n)$, $A^*_{i,n+1} = b_i$ $(i = 1, \ldots, m)$, and $A^*_{m+1,n+1} = 0$. Obviously, if A^* is sampled from any continuous distribution (over the subspace of $R^{(m+1)\times(n+1)}$ characterized by $A^*_{m+1,n+1} = 0$), then every submatrix of A^* (except for the entry $A^*_{m+1,n+1}$) is nonsingular with probability one. It is thus convenient for us to make this assumption explicitly, even though for our proofs not all the submatrices have to be nonsingular. Indeed, matrices that do not satisfy our regularity assumption do arise in practice, and the simplex algorithms handle them efficiently. However, it seems that generalizing our proofs, using arguments of infinitesimal perturbations, would not shed much more light on the problem.

The more important feature of the probabilistic model is the statement of the group of symmetries. In fact, for the lower-bound result, we need a model stronger than the one required for the upper bound result. We first describe the weaker model. Under the weaker model the group is generated by the $m + n$ transformations of multiplying either one of the first n columns or one of the first m rows of the matrix A^* by -1. This group has 2^{m+n} members, giving rise to the same number of instances in each equivalence class. We assume that all the members of a class are equally probable, that is, given that the class was picked, each member has the same probability to be selected from the class. We note than an equivalent description of the model can be given as follows. Instead of fixing the direction of the inequalities $Ax \leq b$ and $x \geq 0$ and letting columns and rows be multiplied by -1, we can fix the matrix A^* and then choose the direction of each of the $m + n$ inequalities independently at random. Closely related models have been considered by Adler and Berenguer [1–4], Buck [8], Haimovich [10], and May and Smith [13]. We note that none of these papers analyzes a complete algorithm for the general linear programming problem, even though some interesting expected values of certain parameters of random polytopes are derived. It turns out that for many parameters, like numbers of faces of any dimension, the probability of a polytope being nonempty, the probability of a polytope being unbounded, and more, the weak model we have described suffice for determining the exact average value of the parameter. However, this is not the case with respect to the average number of steps performed by the self-dual algorithm, as we argue later.

It is interesting to mention that the number of symmetries cannot be subexponential if we are to prove a polynomial upper bound on the average number of steps, since in the worst case the number is exponential.

The stronger model, under which we are able to prove the lower bound result, requires that all the entries of A^* (except for $A^*_{m+1,n+1}$) be independent, identically distributed random variates, whose common distribution is symmetric with respect to zero. We believe that a weaker model would suffice for the same result, but may on the other hand be cumbersome to state. Of course, it follows from the stronger model result that there exist distributions that satisfy the weaker assumptions, relative to which the lower bound holds.

3. *The Algorithm*

We now explain the self-dual method. Consider the following linear programming problem:

$$\begin{array}{ll} \text{maximize} & c^{\mathrm{T}}x \\ \text{subject to} & Ax \leq b \\ & x \geq 0 \end{array}$$

(where $x, c \in R^n$, $A \in R^{m\times n}$, and $b \in R^m$). The dual problem is the following

$$\begin{array}{ll} \text{minimize} & y^{\mathrm{T}}b \\ \text{subject to} & y^{\mathrm{T}}A \geq c^{\mathrm{T}} \\ & y \geq 0. \end{array}$$

The complementary slackness conditions state that two vectors, x (such that $Ax \leq b$ and $x \geq 0$) and y (such that $y^{\mathrm{T}}A \geq c^{\mathrm{T}}$ and $y \geq 0$) are optimal (for their respective problems) if and only if

$$y^{\mathrm{T}}(Ax - b) = 0$$

and

$$(y^{\mathrm{T}}A - c^{\mathrm{T}})x = 0.$$

Letting

$$M = \left[\begin{array}{c|c} & A^{\mathrm{T}} \\ \hline -A & \end{array}\right]$$

and $q = (-c, b)^{\mathrm{T}}$, the problem amounts to finding two vectors z and w in R^{m+n} such that

$$-Mz + w = q, \qquad z^{\mathrm{T}}w = 0, \quad z \geq 0, \quad \text{and} \quad w \geq 0.$$

A useful observation can be made in terms of a piecewise linear mapping

$$F\colon R^{m+n} \to R^{m+n},$$

where

$$F(x) = -Mx^+ - x^-.$$

(We denote $x_i^+ = \max(x_i, 0)$ and $x_i^- = \min(x_i, 0)$.) Here, x^+ plays the role of z, whereas $-x^-$ plays the role of w. Solving the primal and the dual problems amounts to finding an inverse image $F^{-1}(q)$.

The self-dual algorithm starts from any positive vector q_0 and attempts to find solutions for every point on the line segment determined by q_0 and q. Thus, it looks at points of the form $(1 - t)q_0 + tq$. For $t = 0$ there is an obvious solution, namely,

$$z = 0 \qquad \text{and} \qquad w = q_0.$$

The algorithm increases the value of t continuously and follows the inverse image of the point $(1 - t)q_0 + tq$ under the mapping F. Although the inverse image stays within an orthant of R^{m+n}, it varies linearly therein. Every orthant is represented by a *prebasis*, namely, a set of vectors $\{b^1, \ldots, b^{m+n}\} \subseteq R^{m+n}$, where b^i is equal either to the ith column of $-M$ or to the ith unit vector e^i. A prebasis whose vectors are linearly independent is called a *basis*. We identify a basis with an $(m + n) \times (m + n)$ matrix B whose columns are the vectors of the basis. A necessary condition for a prebasis B to be a basis is that equal numbers of unit vectors from the sets $\{\mathbf{e}^1, \ldots, \mathbf{e}^n\}$ and $\{\mathbf{e}^{1+n}, \ldots, \mathbf{e}^{m+n}\}$ are *not* in B. Under the regularity assumption stated in Section 2 (which holds with probability one whenever the matrix A is sampled from a continuous distribution), this condition is also sufficient.

It is well known that the self-dual method solves the linear programming problem under the nondegeneracy assumptions; the algorithm reaches a point q if and only if the linear programming problem has an optimal solution. Otherwise, it discovers an infinite ray that implies that the problem is either infeasible or feasible but unbounded.

The number of pivot steps performed by the algorithm is equal to the number of bases occurring in the path following process, minus one. A basis B occurs in the process if and only if for some t $(0 \le t \le 1)$,

$$B^{-1}((1 - t)q_0 + tq) \ge 0.$$

We note that the algorithm itself is deterministic, so all the probabilistic statements regard the distribution from which the instance (A, b, c) is taken. Denoting by $\Pr(B)$ the probability that the basis B occurs in the process, we note that the expected number $\rho(m, n; q_0)$ of pivot steps corresponding to the starting point q_0 is

$$\rho(m, n; q_0) = \sum_B \Pr(B) - 1.$$

An alternative way to represent $\rho(m, n; q_0)$ (which is called the *facet form* in Smale's papers) is as follows. First, define an *artificial basis* to be a matrix $B_{/i}$ obtained from a basis B by replacing its ith column by the column $-q_0$. Let $\Pr(B_{/i})$ denote the probability that q is in the cone spanned by the columns of $B_{/i}$.

Under these conditions,

$$\rho(m, n; q_0) = \sum_{B,i} \Pr(B_{/i}).$$

We shall estimate the probabilities $\Pr(B_{/i})$.

It turns out that the exact value of $\rho(m, n; q_0)$ depends on the particular distribution and may not be the same for different distributions that satisfy our conditions. The precise value also seems difficult to evaluate. However, for vectors of the form $q_0 = (1, \epsilon, \epsilon^2, \ldots)$, the limits of $\rho(m, n; q_0)$ (as ϵ tends to zero) are close for many distributions, and moreover, they are much easier to estimate. We note that for a fixed distribution the limit of $\rho(m, n; q_0)$ does not necessarily equal the expected number relative to the limit of the starting points, that is, $\rho(m, n; e^1)$.

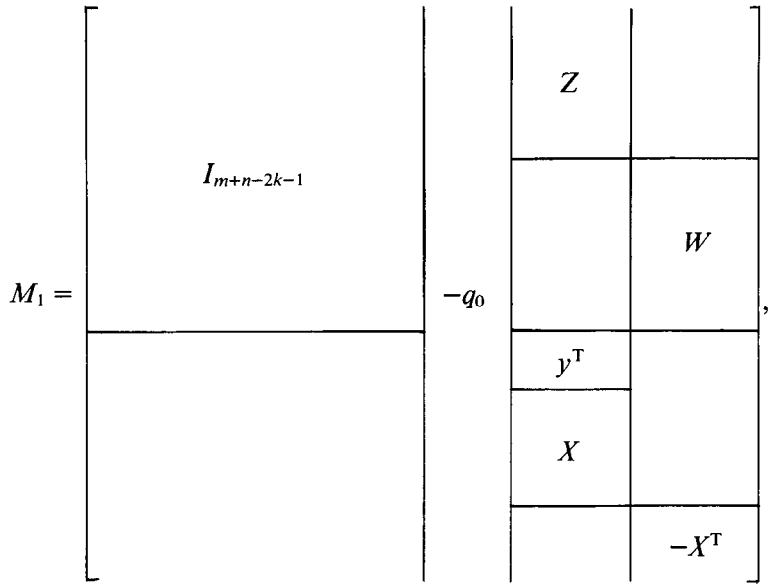
It is very important at this point to clarify the issue of the value of ϵ. For any fixed value of ϵ, the algorithm is well defined (subject to nondegeneracy). The progress of the algorithm, that is, the sequence of bases that it produces, depends of course on ϵ. Obviously, there are only a finite number of intervals of ϵ values such that over each interval, the algorithm produces the same sequence of bases. The latter follows from the fact that the progress depends on comparisons between polynomials of bounded degree in ϵ. It follows that there is $\epsilon_0 > 0$ such that for all ϵ, $0 < \epsilon < \epsilon_0$, the progress of the algorithm is the same. The actual choice of ϵ does not have to be made in advance. In fact, the value of ϵ_0 can be determined by the algorithm itself.

The question of what is the best starting point for solving linear programming problems on the average is still open. However, we know that linear complementarity problems, usually the point $(1, \ldots, 1)^T$, is the worst, while $(1, \epsilon, \epsilon^2, \ldots)^T$ is best in the positive orthant [15]. The effect of the starting point is much easier to study in the context of the linear complementarity problem (see [15]).

4. *Four Types of Artificial Bases*

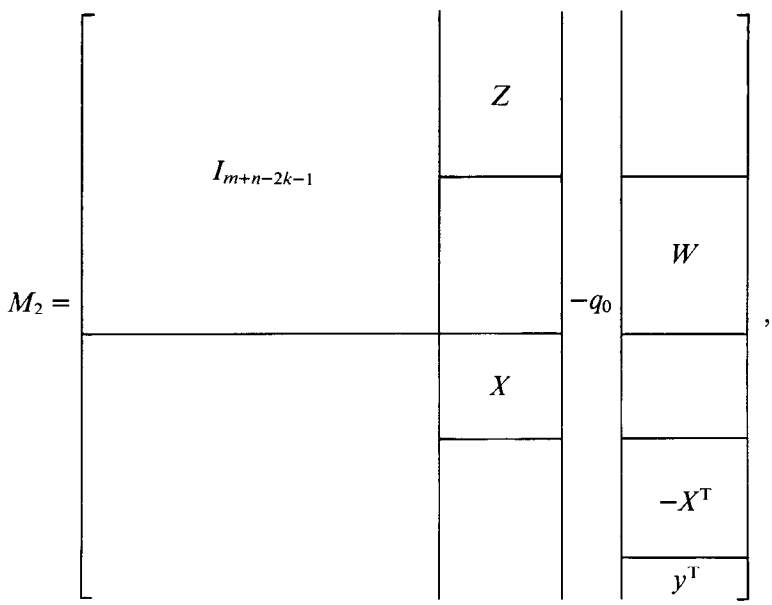
There are four types of artificial bases, $B_{/i}$, depending on the type of basis column that is replaced by $-q_0$: (i) a unit column representing a dual slack, (ii) a unit column representing a primal slack, (iii) a column of M representing a dual variable, (iv) a column of M representing a primal variable. We note that these four cases may be viewed as two pairs of symmetric ones via the primal–dual symmetry. However, the vector q_0 is not symmetric in this respect. We henceforth assume that $q_0 = (1, \epsilon, \epsilon^2, \ldots, \epsilon^{m+n-1})^T$. It is interesting to mention at this point that a different assignment of powers, depending on whether $m \leq n$, or vice versa, yields a slightly better upper bound when the larger dimension tends to infinity while the other is fixed. This issue will be discussed later. Notice that the first n columns of a basis correspond either to primal variables or to dual slacks, whereas the last m columns correspond to either dual variables or primal slacks. It is also convenient to assume, without loss of generality, that $m \leq n$. However, when we represent an artificial basis by an $(m + n) \times (m + n)$ matrix $B_{/i}$, we usually change the order of columns and rows so as to exhibit how a solution to the linear system $B_{/i}x = q$ is obtained. Specifically, we find it convenient to rearrange the matrix so that it has an identity submatrix in the upper-left-hand corner. In the following matrices we use different letters to denote submatrices of the rearranged matrix. These notations do not necessarily reflect the relationships with submatrices of the original matrix. For example, an artificial basis of type (i) can be represented by a matrix

of the form



where $X \in R^{k\times k}$, $Z \in R^{(n-k-1)\times k}$, $W \in R^{(m-k)\times k}$, and $y \in R^k$ $(0 \leq k \leq \min(m, n-1))$. It is very essential to understand at this point what powers of ϵ can arise in the different rows of this matrix. To that end, observe that the rows of Z and the rows of X, together with y^T, constitute segments of the first n rows of the matrix M. Hence, the components of q_0 corresponding to these rows are powers ϵ^j where $0 \leq j \leq n-1$. On the other hand, in rows corresponding to W and $-X^T$ we find in q_0 powers ϵ^j with $n \leq j \leq m+n-1$.

We now describe briefly the other three types of artificial bases. The second type of matrix is of the form



where $X \in R^{k\times k}$, $Z \in R^{(n-k)\times k}$, $W \in R^{(m-k-1)\times k}$, and $y \in R^k$, $(0 \le k \le m-1 \le n-1)$. Here the components of q_0 corresponding to the rows of Z and X are the powers ϵ^j, with $0 \le j \le n-1$, whereas those corresponding to the rows of W, $-X^T$, and y are those with $n \le j \le m+n-1$.

The third type of matrix is of the form

$$M_3 = \left[\begin{array}{c|c|c|c} & & Z & \\ I_{m+n-2k} & & & \\ & -q_0 & & W \\ & & X & \\ & & & y^T \\ & & & -X^T \end{array}\right],$$

where $X \in R^{k\times(k-1)}$, $Z \in R^{(n-k)\times(k-1)}$, $W \in R^{(m-k)\times k}$, and $y \in R^k$ $(1 \le k \le m \le n)$.

The fourth type of matrix is of the form

$$M_4 = \left[\begin{array}{c|c|c|c} & Z & & \\ I_{m+n-2k} & & & W \\ & & -q_0 & \\ & y^T & & \\ & X & & \\ & & & -X^T \end{array}\right],$$

where $X \in R^{(k-1)\times k}$, $Z \in R^{(n-k)\times k}$, $W \in R^{(m-k)\times(k-1)}$, and $y \in R^k$ $(1 \le k \le m \le n)$.

For each of the four types we estimate the probability that, when the vector q is represented as a linear combination of the columns of M_i, all the coefficients are nonnegative. It turns out that types (i) and (iii) are very similar in this respect, whereas types (ii) and (iv) are very similar to each other, but different from (i) and (iii). The reason will become transparent later.

5. *Upper Bounds for Types (i) and (iii)*

In the present section we estimate the limit of the probability $\Pr(B_{/i})$ as ϵ tends to zero, where $B_{/i}$ is an artificial basis of type (i) (see the matrix M_1 in Section 4). We then estimate the expected number of bases of type (i) that occur in the solution process. The analysis of type (iii) is essentially the same with a change of the value of one index as we show later.

The effect of the powers of ϵ in q_0 is illustrated by the following observation. For any $k \times k$ matrix A, let $\Pr(A)$ denote the probability that a random unit k vector v, sampled from a certain continuous distribution over the unit sphere in R^k, is in the cone spanned by the columns of A. Obviously, if A is singular, then $\Pr(A) = 0$. It is interesting to observe the following. Suppose that $A^1, A^2, \ldots$ is a sequence of $k \times k$ matrices, converging to a matrix A^0. If A^0 is nonsingular, then $\lim \Pr(A^n) = \Pr(A^0)$. On the other hand, if A^0 is singular, then $\Pr(A^0) = 0$, but $\lim \Pr(A^n)$ may be positive. For example, consider the case where the columns of A^n are $(1, 0)^T$ and $(-1, 1/n)^T$. Many of our matrices converge to singular matrices when ϵ tends to zero, but we can still estimate the (positive) limit of their probabilities $\Pr(A^n)$.

Our assumptions about the distribution imply that the components of the vector q are nonzeros, and all the 2^{m+n} possible sign patterns have the same probability. In other words, q belongs to any orthant of R^{m+n} with the same probability of $2^{-(m+n)}$. Consider the linear system $M_1 x = q$. It is easy to see that the coefficients of the last $2k + 1$ columns of M_1 (in a representation of q as a linear combination of the columns of M_1) are determined by a smaller system of equations. Let

$$M_1^* = \begin{bmatrix} & y^T & \\ -q_0' & X & \\ & & -X^T \end{bmatrix}$$

($M_1^* \in R^{(2k+1)\times(2k+1)}$), where q_0' is the restriction of q_0 to the components corresponding to the rows of X, $-X^T$ and y. Now consider the system

$$M_1^*(\lambda, \alpha, \beta)^T = q',$$

where q' is the restriction of q to the rows described above, λ is a real number, and α and β are k-vectors. Obviously, the vector $(\lambda, \alpha, \beta)^T$ consists of the coefficients of the last $2k + 1$ columns of M_1 in a representation of q as a linear combination of the columns of M_1. We estimate the probability that $(\lambda, \alpha, \beta)^T \geq 0$. First, we prove a fundamental lemma.

LEMMA 1. *Let $Y \in R^{(k+1)\times(k+1)}$ and let $u \in R^k$. Denote by Y^* a $(k + 2) \times (k + 1)$ matrix such that $Y_{ij}^* = Y_{ij}$ $(i = 1, \ldots, k + 1, j = 1, \ldots, k + 1)$, $Y_{k+2,j}^* = u_j$ $(j = 1, \ldots, k)$ and $Y_{k+2,k+1}^* = 0$. Assume that Y^* satisfies the assumptions of our model, that is, every submatrix of Y^* (except for the entry $Y_{k+2,k+1}^*$) is nonsingular, and the distribution from which Y^* is picked is invariant under multiplication of columns and rows by -1. Let $X \in R^{k\times k}$ be the submatrix obtained from Y by deleting the last row and the last column. Also, let i, $1 \leq i \leq k + 1$, be fixed. Under*

these conditions, the probability that the unit vector $\mathbf{e}^i$ *is in the cone spanned by the columns of* Y, *while* u^T *is in the cone spanned by the rows of* X, *is equal to* $2^{-(2k+1)}$.

PROOF. For any $S \subseteq \{1, \ldots, k+1\}$ and any matrix D, denote by SD a matrix obtained from D by multiplying each row of D, whose index is in S, by -1. Similarly, let DS denote a matrix obtained from D by multiplying each column of D, whose index is in S, by -1. Thus, the objects SY, YS, SX, XS, Se^i, and u^TS are well defined. Let $T \subseteq \{1, \ldots, k\}$ be any subset such that $i \notin T$. Now, consider events as follows. Let E_S denote the event in which $\mathbf{e}^i$ is in the cone spanned by the columns of YS, and let F_T denote the event in which u^T is in the cone spanned by the rows of TX. Obviously, E_S occurs if and only if Te^i is in the cone spanned by TYS, and F_T occurs if and only if u^TS is in the cone spanned by the rows of TXS. It is easy to see that $S_1 \neq S_2$ implies $\Pr(E_{S_1} \cap E_{S_2}) = 0$ and $T_1 \neq T_2$ implies $\Pr(F_{T_1} \cap F_{T_2}) = 0$. By our symmetry assumptions, it follows that the quadruple (TYS, TXS, Te^i, u^TS) has the same joint distribution as the quadruple (Y, X, e^i, u^T). (Recall that $i \notin T$.) Let $G_{ST} = E_S \cap F_T$ and consider the union of the events G_{ST} ($S \subseteq \{1, \ldots, k+1\}$, $T \subseteq \{1, \ldots, k\}$, $i \notin T$). We have already argued that these events have the same probability. Moreover, the intersection of any two of them is empty by the nonsingularity assumption or, alternately, measures zero under any continuous distribution. If $i = k+1$, then the union is the entire sample space. In this case, we have 2^{2k+1} events and the probability of each is hence $2^{-(2k+1)}$. Otherwise ($i \leq k$), we have only 2^{2k} events. On the other hand, the union of these events is not the entire space. In fact, the union is the event in which the coefficient of the ith row of X, in a representation of u^T as a linear combination of the rows of X, is nonnegative. The probability of this event is obviously $\frac{1}{2}$. Thus, the probability in this case is, again, $2^{-(2k+1)}$. □

As a result we get the following

LEMMA 2. *The probability that the last* $2k+1$ *coefficients,* λ, α, *and* β, *are nonnegative tends to* $2^{-(2k+1)}$, *as* ϵ *tends to zero.*

PROOF. We first remark that for any fixed data, the coefficients λ, α, and β tend to (possibly infinite) limits as ϵ tends to zero; however, the α's tend to finite limits. Also, recall that all square submatrices arising are nonsingular with probability one by our assumptions. As a matter of fact, the values of λ and α are determined by a smaller system, corresponding to the square submatrix of order $(k+1) \times (k+1)$ in the upper-left-hand corner of M_1^*, consisting of X, y and a portion of $-q_0$. It follows by arguments similar to those of Lemma 1, that the probability that λ and α are nonnegative is $2^{-(k+1)}$. Furthermore, the asymptotic behavior of λ (as ϵ tends to zero) depends only on the smallest power of ϵ in the portion of q_0 corresponding to rows of X and y. The latter follows from Cramer's formula for the solution of linear equations, under the assumption that the minor corresponding to this power of ϵ does not vanish. Let j denote this smallest power ($0 \leq j \leq n-1$) and assume X, y, and q have been fixed. Then, λ is asymptotically proportional to ϵ^{-j}. This enables us to estimate the probability that also β is nonnegative.

Let q^β and q_0^β denote, respectively, the portions of q and q_0 corresponding to the rows of $-X^T$. It is easy to see that

$$\beta = (-X^T)^{-1}(q^\beta + \lambda q_0^\beta).$$

Recall that all the ϵ^i's participating in q_0^β are with $i \geq n$. It follows that for any fixed data, λq_0^β tends to zero with ϵ. Thus, the probability that β is nonnegative tends to

the probability that $(-X^T)^{-1}q^\beta$ is nonnegative. The latter is obviously equal to 2^{-k}. However, we have to evaluate the intersection of the events "λ and α are nonnegative" and "β is nonnegative." A priori, these are not known to be independent since both depend on the matrix X. However, it follows by arguments similar to those of Lemma 1 that these events are *asymptotically* independent, and the probability of their intersection tends to $2^{-(2k+1)}$. □

LEMMA 3. *Let M_1 be an artificial basis of type (i) and let j be the largest index such that $e^1, \ldots, e^j$ belong to M_1. (If e^1 is not in M_1, then $j = 0$.) Under these conditions, $Pr(M_1)$ tends to $2^{-(m+n-j)}$.*

PROOF. For the proof, we need to consider the rest of the coefficients, that is, those of the $m + n - 2k - 1$ unit vectors. These unit vectors can be classified as primal slacks and dual slacks. A dual slack has a unity in a row in which q_0 has an ϵ^i with $0 \leq i \leq n - 1$, whereas a primal slack has a unity in a row in which q_0 has an ϵ^i with $n \leq i \leq m + n - 1$. Note that, by the definition of the index j, the smallest power of ϵ, in the portion of q_0 corresponding to X and y, is precisely ϵ^j (since e^i corresponds to ϵ^{i-1}). Consider a primal slack e^i $(n + 1 \leq i \leq m + n)$. Let W_i denote the row of W corresponding to the unity of the primal slack, and let q_i denote the component of q in that row. Obviously, the coefficient of e^i is $q_i - W_i\alpha + \lambda\epsilon^{i-1}$. Since $i - 1 > j$, it follows that $\lambda\epsilon^{i-1}$ tends to zero with ϵ. However, by our assumptions, $q_i - W_i\alpha$ is nonzero with probability one. Thus, the probability that the coefficient of e^i is nonnegative tends to the probability that $q_i - W_i\alpha$ is nonnegative. Consider the 2^{m-k} different ways of multiplying rows of W, each augmented with the corresponding coordinate from q, by -1. It follows that the probability that the coefficients of the primal slacks are all nonnegative is equal to $2^{-(m-k)}$.

Now, consider the dual slacks, that is, unit vectors e^i with $1 \leq i \leq n$. The arguments here are similar to those of the previous case, except that $i - 1$ may now be smaller than j. In such a case, the probability that the coefficient of e^i is nonnegative (given that λ is positive) tends to 1, since $\lambda\epsilon^{i-1}$ tends to infinity. If, on the other hand, $i - 1 > j$, then the probability that the coefficient of e^i is nonnegative tends to $\frac{1}{2}$. We can now summarize our findings about the probability that all the coefficients are nonnegative. Each ϵ^i with $i > j$ contributes a factor of $\frac{1}{2}$, while every other ϵ^i contributes a factor of 1. The limit of the probability thus depends only on the value of j, and is equal to $2^{-(m+n-j)}$. □

COROLLARY 4. *The expected number of bases of type (i) occurring in the solution process is less than $m/2 + 1$.*

PROOF. The number of artificial bases of type (i), containing the unit vectors $e^1, \ldots, e^j$ and not containing e^{j+1}, is calculated as follows. For every k $(k = 0, \ldots, \min(m, n - j - 1))$, we can choose the k dual variables in $\binom{m}{k}$ ways. We can choose the $k + 1$ dual slacks to be dropped from the basis (and replaced by k primal variables together with the column $-q_0$) in $\binom{n-j-1}{k}$ different ways, since e^{j+1} must be dropped. Then, the particular choice of which of these will actually be replaced by $-q_0$ can be made in $k + 1$ different ways. To summarize, the number of such bases is

$$\sum_{k=0}^{\min(m,n-j-1)} (k + 1)\binom{m}{k}\binom{n - j - 1}{k}.$$

It follows that the expected number of these bases occurring in the solution process is

$$\sum_{k=0}^{\min(m,n-1)} \left\{(k+1)\binom{m}{k}2^{-m}\sum_{j=0}^{n-k-1}\binom{n-j-1}{k}2^{-(n-j)}\right\}$$
$$= \sum_{k=0}^{\min(m,n-1)} \left\{(k+1)\binom{m}{k}2^{-m-1}\sum_{i=k}^{n-1}\binom{i}{k}2^{-i}\right\}.$$

Now, observe that for $|x| < 1$,

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i$$

and

$$\frac{d^k}{dx^k}\left(\frac{1}{1-x}\right) = \frac{k!}{(1-x)^{k+1}} = \sum_{i=k}^{\infty} k!\binom{i}{k}x^{i-k},$$

so that, for $x = \frac{1}{2}$, we obtain the well-known identity for all k,

$$\sum_{i=k}^{\infty}\binom{i}{k}2^{-i} = 2.$$

It follows that, for any n, the expected number of artificial bases of type (i) occurring in the process is less than

$$2\sum_{k=0}^{m}(k+1)\binom{m}{k}2^{-m-1} = \frac{m}{2} + 1. \qquad \square$$

COROLLARY 5. *The expected number of bases of type (iii) occurring in the solution process is less than $m/2$.*

PROOF. The number of artificial bases of type (iii), containing the unit vectors $e^1, \ldots, e^j$ and not containing e^{j+1}, is

$$\sum_{k=1}^{\min(m,n-j)} k\binom{m}{k}\binom{n-j-1}{k-1}.$$

In order to deal with type (iii), we need a lemma analogous to Lemma 1. More specifically, consider a matrix Y of order $k \times (k+1)$ augmented into a matrix Y^* by an additional row such that Y^* satisfies our usual assumptions. Here X is of order $k \times (k-1)$ and obtained from Y by deleting the first and the last columns. We are interested in the probability of the event in which a unit vector e^i is in the cone spanned by the first k columns of Y, while, in the matrix consisting of the last k columns of Y^*, the bottom row is in the cone spanned by the other rows. By arguments of equally probable events with intersection of measure zero (like those we have used in the proof of Lemma 1), it follows that the probability of this event tends to 2^{-2k}. It then follows that $\Pr(M_3)$ tends to $2^{-(m+n-j)}$. Now the expected number of these bases occurring in the solution process is

$$\sum_{k=1}^{m}\left\{k\binom{m}{k}2^{-m}\sum_{j=0}^{n-k}\binom{n-j-1}{k-1}2^{-(n-j)}\right\}$$
$$= \sum_{k=1}^{m}\left\{k\binom{m}{k}2^{-m-1}\sum_{i=k-1}^{n-1}\binom{i}{k-1}2^{-i}\right\},$$

from which it follows that the expected number of bases of type (iii) is less than $m/2$. □

6. *Upper Bounds for Types (ii) and (iv)*

The analysis of types (ii) and (iv) is slightly more complicated than that of types (i) and (iii). This is due to the fact that, in the case of (ii) and (iv), the coefficient λ of the column $-q_0$ is essentially determined by a row in which the power of ϵ is greater than $n-1$, while smaller powers are present in the submatrix in the lower-right-hand corner of the matrix (see Section 4). However, this situation can still be handled. We consider type (ii) in detail. Type (iv) can then be treated analogously.

LEMMA 6. *Let $Y \in R^{(k+1)\times(k+1)}$ be a random matrix from a distribution like in Lemma 1, that is, the distribution is invariant under multiplication of rows and columns by -1, and every submatrix of Y is nonsingular. Let $X \in R^{k\times k}$ be the submatrix obtained from Y by deleting the last row and the last column. Let $v \in R^{k+1}$ be a unit vector with the unity in the first position and let $u \in R^k$ be a unit vector with the unity in the first position. Under these conditions, the probability that v is in the cone spanned by the columns of Y and $-u^T$ is in the cone spanned by the rows of X is not greater than 2^{-2k}.*

PROOF. We use the notation of Lemma 1, so that the objects SY, YS, SX, XS, Sv, and u^TS are well defined. For $S \subseteq \{2, \cdots, k+1\}$ and $T \subseteq \{2, \cdots, k\}$, consider events as follows. Let E_S denote the event in which v is in the cone spanned by the columns of YS, and let F_T denote the event in which $-u^T$ is in the cone spanned by the rows of TX. Obviously, E_S occurs if and only if Tv is in the cone spanned by TYS, and F_T occurs if and only if $-u^TS$ is in the cone spanned by the rows of TXS. It is easy to see that $S_1 \neq S_2$ implies $\Pr(E_{S_1} \cap E_{S_2}) = 0$ and $T_1 \neq T_2$ implies $\Pr(F_{T_1} \cap F_{T_2}) = 0$. By our symmetry assumptions, it follows that the quadruple $(TYS, TXS, Tv, -u^TS)$ has the same joint distribution as the quadruple $(Y, X, v, -u^T)$. (Recall that $1 \notin S \cup T$.) Let $G_{ST} = E_S \cap F_T$ and consider the union of the events G_{ST} ($S \subseteq \{2, \cdots, k+1\}$, $T \subseteq \{2, \cdots, k\}$). We have already argued that these events have the same probability and, moreover, the intersection of any two of them measures zero. The union of these events is the intersection of the following two events. First is the event in which the coefficient c_u, of the first row of X, in a representation of $-u^T$ as a linear combination of the rows of X, is nonnegative. Second is the event in which the coefficient c_v, of the first column of Y, in a representation of v as a linear combination of the columns of Y, is nonnegative. The probability of this intersection is of course not greater than the probability of each of the events, which is equal to $\frac{1}{2}$. Since this is a union of 2^{2k-1} equally probable events, G_{ST}, it follows that each G_{ST} has a probability not greater than 2^{-2k}. □

It is interesting to point out that our weak model does not allow us to prove a stronger result. Consider the case of $k = 1$ with the matrix Y sampled uniformly from the equivalence class of the following matrix:

$$\begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix},$$

that is, Y can be obtained from this matrix by arbitrary multiplications of rows and columns by -1. It follows that the coefficients c_u and c_v have the same sign for any Y in the class, and the probability that both are positive is $\frac{1}{2}$. Under stronger models (see Section 7), the events are *negatively* correlated, so the probability of

the intersection is less than $\frac{1}{4}$. For example, if Y is a 2×2 matrix whose four entries are sampled independently from the same symmetric (with respect to zero) distribution, then it follows that the probability of both c_u and c_v being positive is precisely $\frac{1}{8}$. We elaborate on these issues in Section 7.

By symmetry it follows that Lemma 6 applies to any positions (not necessarily identical) of the unities in the vectors u and v, except for the last position in v. It turns out that in this latter case the two events are in fact *independent* and a stronger estimate, $2^{-(2k+1)}$, can be proved by similar arguments.

LEMMA 7. *Let M_2 be an artificial basis of type (ii). Let i be an index such that the unit vectors $e^1, \ldots, e^i$ are in the basis, while e^{i+1} is not. Similarly, let j be the index such that the unit vectors $e^{n+1}, \cdots, e^{n+j}$ are in the basis, while e^{n+j+1} is not. Under these conditions, the probability that M_2 occurs in the solution process tends to a limit not greater than $2^{-(m+n-i-j-1)}$.*

PROOF. Let $\alpha \in R^k$, λ and $\beta \in R^k$ denote the coefficients of the last $2k + 1$ columns of M_2 in a representation of a random vector as a linear combination of the columns of M_2. As in the case of M_1, they are determined by a smaller system, corresponding to the square submatrix M_2^*, of order $(2k + 1) \times (2k + 1)$ in the lower-right-hand corner of M_2. Actually, λ and β are determined by an even smaller submatrix in the lower-right-hand corner of M_2^*, consisting of $-X^T$, y and a portion of $-q_0$. By arguments similar to those of Lemma 1, the probability that λ and β are nonnegative tends to $2^{-(k+1)}$. Furthermore, the asymptotic behavior of λ (as ϵ tends to zero) depends only on the smallest power of ϵ, in the portion of q_0 corresponding to rows of $-X^T$ and y. The latter follows from Cramer's formula for the solution of linear equations, under the assumption that the minor, corresponding to this power of ϵ, does not vanish. Our choice of indices implies that this power is precisely ϵ^{n+j}. It follows that λ is asymptotically proportional to $\epsilon^{-(n+j)}$. This enables us to estimate the probability that α is also nonnegative.

Let q^α and q_0^α denote, respectively, the portions of q and q_0 corresponding to the rows of X. It is easy to see that

$$\alpha = X^{-1}(q^\alpha + \lambda q_0^\alpha).$$

Recall that all the ϵ^i's participating in q_0^α are with $i \leq n - 1$. It follows that for any fixed data, each component of λq_0^α tends to infinity when ϵ tends to zero. However, the *direction* of $q^\alpha + \lambda q_0^\alpha$ simply tends to the direction of e^{i+1}, since the smallest power of ϵ in that portion of q_0 is the one that corresponds to this vector. Thus, given that λ and β are nonnegative, the probability that α is also nonnegative tends to the probability that e^{i+1} is in the cone spanned by the columns of X. We are under the conditions of Lemma 6, with some changes of indices. Actually, the case where the smallest power of ϵ, in the portion corresponding to $-X^T$ and y, occurs in the row of y is not covered. However, as indicated earlier, this case can be handled analogously, resulting in an even better bound, $2^{-(2k+1)}$. The conclusion in our case is that the probability of λ, β, and α being nonnegative tends to a limit not greater than 2^{-2k}.

We now need to consider the rest of the coefficients, that is, those of the $m + n - 2k - 1$ unit vectors, as in the proof of Lemma 3. Consider a primal slack e^ν $(n + 1 \leq \nu \leq m + n)$. Let W_ν denote the row of W corresponding to the unity of the primal slack and let q_ν denote the component of q in that row. Obviously, the coefficient of e^ν is $q_\nu - W_\nu\beta + \lambda\epsilon^{\nu-1}$. The probability that $q_\nu - W_\nu\beta$ is nonzero is one. If $\nu - 1 > n + j$, then $\lambda\epsilon^{\nu-1}$ tends to zero with ϵ. The probability of positivity

of the coefficient in this case tends to $\frac{1}{2}$. If $\nu - 1 < n + j$, then $\lambda\epsilon^{\nu-1}$ tends to infinity and the corresponding probability tends to 1.

Consider a dual slack, e^ν with $1 \le \nu \le n$. The arguments here are similar to those of the previous case. It can be verified that, if $\nu - 1 < i$, then the probability that this coefficient of e^i is nonnegative (given that λ is positive) tends to one, although, if $\nu - 1 > i$, then this probability tends to $\frac{1}{2}$. We can now summarize our findings about the probability that all the coefficients are nonnegative. Each ϵ^ν with either $i < \nu \le n - 1$ or $\nu > n + j$ contributes a factor of $\frac{1}{2}$, while each of the other ϵ^ν's contributes a factor of 1. The limit of the probability is thus bounded from above by $2^{-(m+n-i-j-1)}$. □

COROLLARY 8. *The expected number of bases of type (ii) occurring in the solution process is not greater than* $m^2 + m$.

PROOF. Our calculations here are similar to those of Corollary 4. The number of bases, with indices i and j as defined in Lemma 7, is (using the convention $\binom{i}{-1} = 1$ for $i \ge -1$),

$$\sum_{k=0}^{\min(m-i-1,n-j)} (k+1)\binom{m-i-1}{k}\binom{n-j-1}{k-1}.$$

It follows that the expected number of these bases occurring in the solution process is not greater than

$$2\sum_{k=0}^{m-1}\left\{(k+1)\sum_{i=0}^{m-k-1}\binom{m-i-1}{k}2^{-(m-i)}\sum_{j=0}^{n-k}\binom{n-j-1}{k-1}2^{-(n-j)}\right\}$$

$$= \frac{1}{2}\sum_{k=0}^{m-1}\left\{(k+1)\sum_{i=k}^{m-1}\binom{i}{k}2^{-i}\sum_{j=k-1}^{n-1}\binom{j}{k-1}2^{-j}\right\}.$$

It follows that for any n, the expected number of artificial bases of type (ii) occurring in the process is not greater than

$$2\sum_{k=0}^{m-1}(k+1) = m^2 + m. \qquad \square$$

COROLLARY 9. *The expected number of bases of type (iv) occurring in the solution process is not greater than* $m^2 + m$.

PROOF. The arguments are very similar to those of the previous case. The probability can be shown to tend to a limit not greater than $2^{-(m+n-i-j-1)}$. The number of bases is

$$\sum_{k=1}^{\min(m-i,n-j)} k\binom{m-i-1}{k-1}\binom{n-j-1}{k-1}.$$

It follows that the upper bound in the present case is

$$2\sum_{k=1}^{m}\left\{k\sum_{i=0}^{m-k}\binom{m-i-1}{k-1}2^{-(m-i)}\sum_{j=0}^{n-k}\binom{n-j-1}{k-1}2^{-(n-j)}\right\} < 2\sum_{k=1}^{m}k = m^2 + m. \qquad \square$$

In view of the calculations made in this section and the preceding one, it turns out that there is room for some improvement. First, notice the following symmetries between types. Suppose we assign the powers ϵ^j with $0 \le j \le m - 1$ to the dual variables, and those with $m \le j \le m + n - 1$ to the primal variables, and suppose

we interchange the roles of primal and dual. Types (i) and (ii) are symmetric under this transformation, and so are types (iii) and (iv). Subject to the original assignment of powers of ϵ, types (i) and (iii) contribute linear terms, whereas types (ii) and (iv) contribute quadratic terms. A quadratic term arises when there are two critical indices i and j, whereas a linear term arises when there is only one. More specifically, the first critical index is the smallest power of ϵ that is present in the section according to which the coefficient of q_0 is determined. This power is associated with a dual variable or a primal variable (depending on the type) that is present in the basis. If it is associated with a primal variable, then the second critical power is the smallest that corresponds to a dual variable, which is present in the basis, and vice versa. However, the second critical power plays its role as critical only if it is smaller than the first one. As we show later, *any* assignment of powers of ϵ yields an algorithm with a quadratic upper bound. On the other hand, there is room for improvement in the linear term of the upper bound, in case one of the dimensions is substantially larger than the other one, which can be seen as follows.

Assume $m \le n$ and let us assign the powers ϵ^j with $0 \le j \le m - 1$ to the dual variables, and those with $m \le j \le m + n - 1$ to the primal variables. Essentially, assign the smaller powers to the smaller dimension and the larger ones to the larger dimension. It follows from our symmetry arguments that, in this case, type (i) contributes the number of steps contributed by type (ii) subject to the original assignment, that is, no more than $m^2 + m$ steps. Similarly, type (iii) behaves as type (iv) did in the original assignment. Of course, type (ii) also behaves like type (i) and type (iv) behaves like type (iii). However, we are able to prove a better upper bound for the latter two in case n tends to infinity. Consider type (ii). Let i denote the critical index; that is, the first i primal slacks are present in the basis, while the $(i + 1)$st one is not. Now the second critical index is not critical at all. The number of bases of type (ii) with critical index i is

$$\sum_{k=0}^{m-i-1} (k+1)\binom{m-i-1}{k}\binom{n}{k}.$$

The probability of each occurring is $2^{-(m+n-i)}$. Thus, the expected number of bases of type (ii) in this case is no more than

$$\sum_{k=0}^{m-1}\left\{(k+1)\binom{n}{k}2^{-n}\sum_{i=0}^{m-k-1}\binom{m-i-1}{k}2^{-(m-i)}\right\}$$

$$= \sum_{k=0}^{m-1}\left\{(k+1)\binom{n}{k}2^{-n}\frac{1}{2}\sum_{j=k}^{m-1}\binom{j}{k}2^{-j}\right\},$$

which is less than

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

The latter tends to *zero* when n tends to infinity while m is fixed. A similar bound can be obtained for the expected number of bases of type (iv). However, types (i) and (iii) contribute a quadratic expected number of bases, so this improvement is not a major one.

In general, any assignment of powers of ϵ yields a quadratic upper bound. The following is a sketch of proof of this statement. Consider the orders that are induced on the set of primal slacks and on the set of dual slacks by the assignment of powers of ϵ. Consider any artificial basis B and define two critical indices as follows. Let i

be the index such that the first i primal slacks (i.e., the ones with the i smallest powers of ϵ that are assigned to dual variables) are present in B, while the $(i+1)$st is not. Analogously, let j be the index such that the first j dual slacks (i.e., the ones with the j smallest powers of ϵ that are assigned to primal variables) are present in B while the $(j+1)$st is not. By arguments that have been repeated in this paper, regardless of the type of B, the probability that B occurs in the solution process can be shown not greater than $2^{-(m+n-i-j-1)}$. The technique of counting the bases is essentially the same for all the four types. It turns out that for each type the expected number of bases occurring in the process is $O((\min(m, n))^2)$.

7. *The Quadratic Lower Bound*

In this section we establish that the expected number of steps is indeed quadratic in the minimum of the two dimensions of the problem. To this end, it is of course sufficient to show that the expected number of bases of type (ii) is $\Omega((\min(m, n))^2)$. As in the previous sections, we continue to assume without loss of generality that $m \leq n$.

For the lower bound result we need a stronger probabilistic model. A convenient model is as follows. We simply assume the entires of A, b, and c to be independent, identically distributed random variates, with a common distribution that is symmetric with respect to zero. This assumption strengthens the symmetry under reflection conditions assumed earlier in this paper. On the other hand, the result implies that there is no asymptotically better upper bound under the weaker model. We also assume nonsingularities as before.

The following lemma complements Lemma 6, under the stronger model.

LEMMA 10. *Let $Y \in R^{(k+1)\times(k+1)}$ be a matrix whose entries are independent, identically distributed random variates whose common distribution is symmetric with respect to zero. Also, assume all the minors of Y to be nonzero. Let $X \in R^{k\times k}$ be the submatrix obtained from Y by deleting the last row and the last column. Let $v \in R^{k+1}$ be a unit vector with the unity in the first position and let $u \in R^k$ be a unit vector with the unity in the first position. Under these conditions, the probability that v is in the cone spanned by the columns of Y, and $-u^T$ is the cone spanned by the rows of X, is between 2^{-2k-2} and 2^{-2k-1}.*

It follows by symmetry that the same result holds for any positions (not necessarily identical) of the unities in the vectors u and v, except for the last position in v. As indicated earlier, in the latter case, the two events (namely, the inclusion of v and $-u^T$ in the respective cones) are independent, so that under an even weaker model the probability of the intersection is precisely 2^{-2k-1}.

For the proof of Lemma 10 we need several preparatory lemmas. The first is a fact of linear algebra.

LEMMA 11. *Let $X \in R^{k\times k}$ be any matrix and denote $\alpha = X_{11}$, $a = (X_{12}, \cdots, X_{1k})^T$ and $b = (X_{21}, \cdots, X_{k1})^T$. Also, let $U \in R^{(k-1)\times(k-1)}$ denote the lower-right-hand corner of X and suppose U is nonsingular. Under these conditions,*

$$det(X) = det(U)(\alpha - a^T U^{-1} b).$$

PROOF. Since

$$X = \begin{bmatrix} \alpha & a^T \\ b & U \end{bmatrix},$$

it follows by row operations that

$$\det X = \det \begin{bmatrix} \alpha - a^{\mathrm{T}}U^{-1}b & 0 \\ b & U \end{bmatrix},$$

which implies the lemma. □

The following lemma is again of linear algebra.

LEMMA 12. *Let $Y \in R^{(k+1)\times(k+1)}$ be any matrix and denote submatrices of Y as follows:*

(*i*) *Let $X \in R^{k\times k}$ be the submatrix of Y in the upper-left-hand corner.*
(*ii*) *Let $Z \in R^{k\times k}$ be the submatrix of Y in the lower-left-hand corner.*
(*iii*) *Let $W \in R^{k\times k}$ be the submatrix of Y in the upper-right-hand corner.*
(*iv*) *Let $V \in R^{k\times k}$ be the submatrix of Y in the lower-right-hand corner.*
(*v*) *Let $U \in R^{(k-1)\times(k-1)}$ be the center submatrix of Y (obtained by deleting both the first and the last row and both the first and the last column).*

Under these conditions,

$$det(Y)det(U) = det(X)det(V) - det(Z)det(W).$$

PROOF. Represent the matrix Y in the following form:

$$Y = \begin{bmatrix} \alpha & a^{\mathrm{T}} & \beta \\ b & U & c \\ \gamma & c^{\mathrm{T}} & \delta \end{bmatrix}.$$

Now, if U is nonsingular, then row operations give

$$\det Y = \det \begin{bmatrix} \alpha - a^{\mathrm{T}}U^{-1}b & 0 & \beta - a^{\mathrm{T}}U^{-1}c \\ b & U & c \\ \gamma & c^{\mathrm{T}} & \delta \end{bmatrix}.$$

Expanding along the first row gives

$$\det Y = (\alpha - a^{\mathrm{T}}U^{-1}b)\det V + (-1)^k(\beta - a^{\mathrm{T}}u^{-1}c)\det Z.$$

It follows from Lemma 11 that

$$\det X = (\alpha - a^{\mathrm{T}}U^{-1}b)\det U \quad \text{and} \quad \det W = (-1)^{k-1}(\beta - a^{\mathrm{T}}U^{-1}c)\det U$$

from which the result follows. If U is singular then there are nonsingular U''s arbitrarily close to U and the result follows by continuity of the determinant function. □

The following is a simple probabilistic lemma.

LEMMA 13. *Suppose u and v are independent, identically distributed random n vectors, and let $C \subseteq R^n$ be a random set (independent of u and v) from any probability space whose elementary events are measurable subsets of R^n. Under these conditions,*

$$Pr(\{u, v\} \subseteq C) \geq Pr(u \in C)Pr(v \in C).$$

PROOF. Obviously,

$$\Pr(u \in C) = \Pr(v \in C).$$

Denote by Pr* the probability of an event where C is *fixed.* It follows that

$$\Pr^*(\{u, v\} \subseteq C) = \Pr^*(u \in C)\Pr^*(b \in C) = (\Pr^*(u \in C))^2.$$

Let μ denote the probability measure corresponding to the sampling of C. It follows that

$$\begin{aligned}\Pr(\{u, v\} \subseteq C) &= \int \Pr^*(\{u, v\} \subseteq C)\, d\mu \\ &= \int \Pr^*(u \in C)\Pr^*(v \in C)\, d\mu \\ &= \int (\Pr^*(u \in C))^2\, d\mu \\ &\geq \left(\int \Pr^*(u \in C)\, d\mu\right)^2 \\ &= (\Pr(u \in C))^2 = \Pr(u \in C)\Pr(v \in C),\end{aligned}$$

where the inequality follows from Jensen's inequality. □

We now apply Lemmas 11 and 13 in a situation that involves random matrices.

LEMMA 14. *Let $Y \in R^{(k+1)\times(k+1)}$ be a matrix whose entries are independent, identically distributed random variates, such that their common distribution is continuous and symmetric with respect to zero. Let X, V, Z, and W be the four corner submatrices of Y of order $k \times k$ as defined in Lemma 12. Under these conditions,*

$$Pr(det(X)det(V)det(Z)det(W) \geq 0) \geq \tfrac{1}{2}.$$

PROOF. By the continuity assumption, all minors of Y are nonzero with probability one. Let α, a, b, and U be as in the previous lemmas, and also denote

$$\begin{aligned}\beta &= Y_{1,k+1}, \\ \gamma &= Y_{k+1,1}, \\ \delta &= Y_{k+1,k+1}, \\ c &= (Y_{k+1,2}, \ldots, Y_{k+1,k})^{\mathrm{T}}, \\ d &= (Y_{2,k+1}, \ldots, Y_{k,k+1})^{\mathrm{T}}.\end{aligned}$$

It follows from Lemma 11 that the product of the four determinants is equal to

$$(\det(U))^4(\alpha - a^{\mathrm{T}}U^{-1}b)(\beta - a^{\mathrm{T}}U^{-1}d)(\gamma - c^{\mathrm{T}}U^{-1}b)(\delta - c^{\mathrm{T}}U^{-1}d).$$

Obviously, it is sufficient to consider the sign of

$$(\alpha - a^{\mathrm{T}}U^{-1}b)(\beta - a^{\mathrm{T}}U^{-1}d)(\gamma - c^{\mathrm{T}}U^{-1}b)(\delta - c^{\mathrm{T}}U^{-1}d).$$

We now apply Lemma 11. Let $u = (Y_{11}, \ldots, Y_{1,k+1})^{\mathrm{T}}$, that is, $u = (\alpha, a^{\mathrm{T}}, \beta)^{\mathrm{T}}$, and $v = (Y_{k+1,1}, \ldots, Y_{k+1,k+1})^{\mathrm{T}}$, that is, $v = (\gamma, c^{\mathrm{T}}, \delta)^{\mathrm{T}}$. Given the values of Y_{ij} for $i = 2, \ldots, k$ and $j = 1, \ldots, k+1$, let C denote the set of all vectors $(\alpha, a^{\mathrm{T}}, \beta)^{\mathrm{T}}$ such that

$$(\alpha - a^{\mathrm{T}}U^{-1}b)(\beta - a^{\mathrm{T}}U^{-1}d) > 0.$$

We note that also

$$(\gamma - c^{\mathrm{T}}U^{-1}b)(\delta - c^{\mathrm{T}}U^{-1}d) > 0$$

if and only if $(\gamma, c^{\mathrm{T}}, \delta)^{\mathrm{T}} \in C$. Let C^{c} denote the complement of C and note that

under our model all the determinants are nonzero with probability one. Thus,

$$\begin{aligned}\Pr(\det(X)\det(V)\det(Z)\det(W) \geq 0) &= \Pr(\{\det(X)\det(W) > 0\} \cap \{\det(V)\det(Z) > 0\}) \\ &\quad + \Pr(\{\det(X)\det(W) < 0\} \cap \{\det(V)\det(Z) < 0\}) \\ &= \Pr(\{u, v\} \subseteq C) + \Pr(\{u, v\} \subseteq C^c) \\ &\geq (\Pr(u \in C))^2 + (\Pr(u \in C^c))^2 \geq \tfrac{1}{2}. \end{aligned}$$ □

We now return to questions that are more closely related to those we were dealing with in the previous sections. The following lemma constitutes the essence of Lemma 10.

LEMMA 15. *Let Y, X, u, and v be as in Lemma* 10. *Let c_u denote the coefficient of the first row of X in a representation of $-u^T$ as a linear combination of the rows of X, and let c_v denote the coefficient of the first column of Y in a representation of v as a linear combination of the columns of Y. Under these conditions, the probability that both c_u and c_v are positive is between $\frac{1}{8}$ and $\frac{1}{4}$.*

PROOF. Let U, V, W, and Z be as in the previous lemmas. Obviously,

$$c_u = -\frac{\det(U)}{\det(X)} \quad \text{and} \quad c_v = \frac{\det(V)}{\det(Y)}.$$

We are interested in the event E in which

$$\frac{\det(V)}{\det(Y)} > 0 > \frac{\det(U)}{\det(X)}.$$

First, note that, when the first row of Y is multiplied by -1, then the signs of c_u and c_v are reversed. Since the distribution of Y is invariant under this operation, it follows that

$$\Pr(E) = \tfrac{1}{2} \Pr(\det(X)\det(V)\det(Y)\det(U) < 0).$$

However, by Lemma 12,

$$\det(Y)\det(U) = \det(X)\det(V) - \det(Z)\det(W).$$

Consider the random variates $\xi = \det(X)\det(V)$ and $\eta = \det(Z)\det(W)$. Obviously, ξ and η are identically distributed. Moreover, the common distribution is symmetric with respect to zero since they change sign when the first column of Y is multiplied by -1. It follows that

$$\begin{aligned}\Pr(E) &= \tfrac{1}{2}(\Pr\{\eta < \xi < 0\} + \Pr\{\eta > \xi > 0\}) \\ &= \Pr(\eta < \xi < 0) \\ &= \Pr(\{\eta < 0\} \cap \{\xi < 0\})\Pr(\eta < \xi \mid \{\eta < 0\} \cap \{\xi < 0\}) \\ &= \tfrac{1}{2} \Pr(\{\eta < 0\} \cap \{\xi < 0\}) \leq \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4}. \end{aligned}$$

This establishes the upper bounding part of our lemma.

On the other hand,

$$\Pr(\{\eta < 0\} \cap \{\xi < 0\}) = \tfrac{1}{2} \Pr\{\eta\xi > 0\},$$

so, in view of Lemma 14, we also have

$$\begin{aligned}\Pr(E) &= \Pr\{\eta < \xi < 0\} \\ &= \tfrac{1}{2} \Pr\{\eta\xi > 0\}\Pr(\eta < \xi \mid \{\eta < 0\} \cap \{\xi < 0\}) \\ &= \tfrac{1}{4} \Pr\{\eta\xi > 0\} \geq \tfrac{1}{8}. \end{aligned}$$ □

We are now able to prove Lemma 10.

PROOF OF LEMMA 10. The proof follows directly from Lemmas 6 and 15. By Lemma 15, the union of the events G_{ST} has probability between $\frac{1}{4}$ and $\frac{1}{8}$. Since these are 2^{2k-1} equally probable events (and the intersection of every two of them measures zero), it follows that each has probability between 2^{-2k-1} and 2^{-2k-2}. This completes the proof of Lemma 10. □

We now have a result stronger than that of Lemma 7.

COROLLARY 16. *Under the conditions of Lemma* 7, *subject to a model in which the inputs are independent, identically distributed random variates (symmetric with respect to zero), the probability that* M_2 *occurs in the solution process tends to a limit between* $2^{-(m+n-i-j+1)}$ *and* $2^{-(m+n-i-j)}$.

PROOF. The proof is essentially the same as that of Lemma 7, taking advantage of the result of Lemma 10. □

Before stating the lower bound result, we need a combinatorial lemma.

LEMMA 17. *For every* k, $k = 1, 2, \ldots,$

$$\sum_{i=k}^{2k} \binom{i}{k} 2^{-i} = 1.$$

PROOF. The proof goes by induction on k. The lemma is obviously true for $k = 1$. The inductive step is as follows

$$\begin{aligned}
\sum_{i=k+1}^{2k+2} \binom{i}{k+1} 2^{-i} &= \sum_{i=k+1}^{2k+2} \binom{i-1}{k} 2^{-i} + \sum_{i=k+2}^{2k+2} \binom{i-1}{k+1} 2^{-i} \\
&= \frac{1}{2} \sum_{j=k}^{2k+1} \binom{j}{k} 2^{-j} + \frac{1}{2} \sum_{j=k+1}^{2k+1} \binom{j}{k+1} 2^{-j} \\
&= \frac{1}{2} \sum_{j=k}^{2k} \binom{j}{k} 2^{-j} + \frac{1}{2} \binom{2k+1}{k} 2^{-(2k+1)} \\
&\quad + \frac{1}{2} \sum_{j=k+1}^{2k+2} \binom{j}{k+1} 2^{-j} - \frac{1}{2} \binom{2k+2}{k+1} 2^{-(2k+2)}.
\end{aligned}$$

Notice that

$$\binom{2k+1}{k} = \frac{1}{2} \binom{2k+2}{k+1}.$$

The rest of the proof follows easily. □

Finally, we can prove a quadratic lower bound on the expected number of bases of type (ii) occurring in the solution process.

THEOREM 18. *The expected number of bases of type* (*ii*) *occurring in the solution process grows quadratically with* m.

PROOF. We rely on figures obtained in Corollary 8 and the lemmas of the present section. The number of bases, with indices i and j as defined in Lemma 7, is

$$\sum_{k=0}^{\min(m-i-1,n-j)} (k+1) \binom{m-i-1}{k} \binom{n-j-1}{k-1}.$$

By Corollary 16, the probability of a basis of this type to occur in the process is at least $2^{-(m+n-i-j+1)}$. It follows that the expected number of these bases occurring in the process is at least

$$\frac{1}{2}\sum_{k=1}^{m-1}\left\{(k+1)\sum_{i=0}^{m-k-1}\binom{m-i-1}{k}2^{-(m-i)}\sum_{j=0}^{n-k}\binom{n-j-1}{k-1}2^{-(n-j)}\right\}$$

$$=\frac{1}{8}\sum_{k=1}^{m-1}\left\{(k+1)\sum_{i=k}^{m-1}\binom{i}{k}2^{-i}\sum_{j=k-1}^{n-1}\binom{j}{k-1}2^{-j}\right\}.$$

The latter is greater than

$$\frac{1}{8}\sum_{k=1}^{\lfloor (m-1)/2\rfloor}\left\{(k+1)\sum_{i=k}^{2k}\binom{i}{k}2^{-i}\sum_{j=k-1}^{2k-2}\binom{j}{k}2^{-j}\right\}$$

$$=\frac{1}{8}\sum_{k=1}^{\lfloor (m-1)/2\rfloor}(k+1)>\frac{1}{64}m^2+\frac{1}{32}m-\frac{1}{8}. \qquad \square$$

We have not attempted to maximize the coefficient of m^2 in our lower bound for the expected total number of steps of the algorithm. The latter is obviously larger than $\frac{1}{64}$ since we also have the bases of type (iv) contributing a similar term. Also, we were quite generous in the proof, especially in taking the sum only up to $k = \lfloor (m-1)/2 \rfloor$.

8. *Conclusion*

We have estimated the expected number of artificial bases occurring in the solution process. It is interesting to mention that the self-dual algorithm can actually be implemented with only half the number of pivot steps described in this paper. This is due to the fact that every other orthant of R^{m+n}, which is met by the inverse image of the line segment $[q_0, q]$, corresponds to a singular prebasis (see Section 3). While the inverse image is crossing such an orthant, the point in the image space does not move at all. The implementation, as described in [9], maintains a tableau of dimension $m \times m$ ($m \leq n$) and has the characteristic that (under nondegeneracy) a pivot occurs only if the point in the image space moves. Subject to this observation, the expected number of steps, as estimated in this paper, is bounded from above by

$$m^2 + 1.5m + 0.5$$

(assuming $m \leq n$). A better bound is obtained if the smaller exponents of ϵ are assigned to the problem with the fewer variables (see Section 5). The result is that asymptotically, when n tends to infinity while m is fixed, the average number of steps is bounded from above by

$$m^2 + m,$$

but the previous bound prevails for any m and n. Under the stronger model of Section 7 the probabilities corresponding to types (ii) and (iv) are multiplied by $\frac{1}{2}$. This implies a uniform bound of

$$0.5m^2 + 1.5m + 0.5,$$

decreasing to

$$0.5m^2 + 0.5m$$

as n tends to infinity. Moreover, the expected number of steps under the stronger

model is bounded from below by $\frac{1}{64}m^2 + \frac{1}{32}m - \frac{1}{8}$. This lower bound can obviously be improved upon (since it is based on type (ii) only), but we have not attempted to do so in the present paper.

ACKNOWLEDGMENTS. We thank Mike Todd for providing the short proof of Lemma 12 and the referees for their numerous constructive suggestions.

REFERENCES

1. ADLER, I. The expected number of pivots needed to solve parametric linear programs and the efficiency of the self-dual simplex method. Dept. of Industrial Engineering and Operations Research, Univ. of Calif., Berkeley, Berkeley, Calif., June 1983.
2. ADLER, I., AND BERENGUER, S. E. Random linear programs. Tech. Rep. ORC 81-4, Operations Research Center, Univ. of Calif., Berkeley, Berkeley, Calif., 1981.
3. ADLER, I., AND BERENGUER, S. E. Duality theory and the random generation of linear programs. Manuscript. Dept. of Industrial Engineering and Operations Research, Univ. of Calif., Berkeley, Berkeley, Calif., 1981.
4. ADLER, I., AND BERENGUER, S. E. Generating random linear programs. Revised manuscript, Dept. of Industrial Engineering and Operations Research, Univ. of Calif., Berkeley, Berkeley, Calif., (1983).
5. BLAIR, C. Random linear programs with many variables and few constraints. Faculty Working Paper No. 946, College of Commerce and Business Administration, Univ. of Illinois at Urbana-Champaign, Urbana, Ill., April 1983.
6. BORGWARDT, K.-H. Some distribution-independent results about the asymptotic order of the average number of pivot steps of the simplex method. *Math. Oper. Res. 7*(1982), 441–462.
7. BORGWARDT, K.-H. The average number of steps required by the simplex method is polynomial. *Z. Oper. Res. 26*(1982), 157–177.
8. BUCK, R. C. Partition of space. *Am. Math. Monthly 50*(1943), 541–544.
9. DANTZIG, G. B. *Linear Programming and Extensions.* Princeton University Press, Princeton, N. J., 1963.
10. HAIMOVICH, M. 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 1983.
11. KLEE, V., AND MINTY, G. J. How good is the simplex algorithm? In *Inequalities*, vol. III. O. Shisha, Ed. Academic Press, Orlando, Fla., 1972, pp. 159–175.
12. LEMKE, C. E. Bimatrix equilibrium points and mathematical programming. *Manage. Sci. 11*(1965), 681–689.
13. MAY, J., AND SMITH, R. Random polytopes: Their definition, generation, and aggregate properties. *Math. Prog. 24*(1982), 39–54.
14. MEGIDDO, N. Improved asymptotic analysis of the average number of steps performed by the self-dual simplex algorithm. *Math. Prog. Study*, to appear.
15. MEGIDDO, N. On the expected number of linear complementarity cones intersected by random and semi-random rays. *Math. Prog. Study*, to appear.
16. MURTY, K. G. Computational complexity of parametric linear programming. *Math. Prog. 19*(1980), 213–219.
17. SMALE, S. On the average number of steps of the simplex method of linear programming. *Math. Prog. 27*(1983), 241–262.
18. SMALE, S. The problem of the average speed of the simplex method. In *Mathematical Programming: The State of the Art*, A. Bachem, M. Grötschel, and B. Korte, Eds. Springer-Verlag, New York, 1983, pp. 530–539.
19. VAN DER HEYDEN, L. A variable dimension algorithm for the linear complementarity problem. *Math. Prog. 19*(1980), 328–346.

RECEIVED DECEMBER 1983; REVISED MARCH 1985; ACCEPTED APRIL 1985