Probabilistic Analysis of an Infeasible-Interior-Point Algorithm for Linear Programming

Kurt M. Anstreicher[1], Jun Ji[2], Florian A. Potra[3], and Yinyu Ye[4]

Final Revision June, 1998

Abstract

We consider an infeasible-interior-point algorithm, endowed with a finite termination scheme, applied to random linear programs generated according to a model of Todd. Such problems have degenerate optimal solutions, and possess no feasible starting point. We use no information regarding an optimal solution in the initialization of the algorithm. Our main result is that the expected number of iterations before termination with an exact optimal solution is $O(n \ln(n))$.

Keywords: Linear Programming, Average-Case Behavior, Infeasible-Interior-Point Algorithm.

Running Title: Probabilistic Analysis of an LP Algorithm

---

[1] Dept. of Management Sciences, University of Iowa. Supported by an Interdisciplinary Research Grant from the Center for Advanced Studies, University of Iowa.

[2] Dept. of Mathematics, Valdosta State University. Supported by an Interdisciplinary Research Grant from the Center for Advanced Studies, University of Iowa.

[3] Dept. of Mathematics, University of Iowa. Supported by an Interdisciplinary Research Grant from the Center for Advanced Studies, University of Iowa.

[4] Dept. of Management Sciences, University of Iowa. Supported by NSF Grant DDM-8922636.

1. Introduction

A number of recent papers have attempted to analyze the probabilistic behavior of interior point algorithms for linear programming. Ye (1994) showed that a variety of algorithms, endowed with the finite termination scheme of Ye (1992) (see also Mehrotra and Ye 1993), obtain an exact optimal solution with "high probability" (probability approaching one as $n \to \infty$) in no more than $O(\sqrt{n}\ln(n))$ iterations. Here n is the number of variables in a standard form primal problem. Several subsequent works - Huang and Ye (1991), Anstreicher, Ji, and Ye (1992), and Ji and Potra (1992) - then obtained bounds on the expected number of iterations until termination, using various algorithms and termination methods. The analysis in each of these latter papers is based on a particular random linear programming model from Todd (1991) (Model 1 with $\hat{x} = \hat{s} = e$, see Todd 1991, p.677), which has a known initial interior solution for the primal and dual problems, and is nondegenerate with probability one. Unfortunately, we eventually realized that these three papers all suffer from a fatal error in conditional probability, and consequently do *not* provide correct analyses of the probabilistic behavior of interior point algorithms. The error is basically the following: Todd (1991, Theorem 3.6) determines the distribution of the components of a primal basic feasible solution for this case of his Model 1, and similar analysis can be used to obtain the distribution of the components of a dual basic feasible solution. What is required in the probabilistic analysis is the distribution of the positive components of the primal and dual optimal solutions. However, conditioning on optimality is equivalent to conditioning on primal and dual feasibility, and these are *not* independent of one another. (Theorem 3.6 of Todd (1991) itself contains an error which will be addressed in a forthcoming erratum to that paper, and which is further discussed in Section 4.)

A variant of Todd's Model 1 which allows for degeneracy is given in Todd (1991, Section 4). Throughout the paper we will refer to this model as "Todd's degenerate model." Todd's degenerate model controls the degree of degeneracy by specifying optimal primal and dual solutions, but provides no feasible starting point. This presents a difficulty for most interior point methods, which require feasible primal and/or dual solutions for initialization. One way around this difficulty is to use a combined primal-dual feasibility problem, as in Ye (1994). Another approach would be to use an artificial variable, with "M" objective coefficient, and increase M as necessary to insure feasibility. Interior point algorithms which employ such a strategy have been suggested by Ishihara and Kojima (1993), and Kojima, Mizuno, and Yoshise (1993). In fact, for Todd's degenerate model the required value of M could be inferred from the known optimal dual solution, but the use of such information is clearly "cheating," since a general linear programming algorithm cannot take as input properties of a (usually unknown) optimal solution. Finally, one could attempt a probabilistic analysis of a combined Phase I - Phase II algorithm, for example Anstreicher (1989, 1991) or Todd (1992, 1993).

In practice, another algorithm, the primal-dual "infeasible-interior-point" method, has been very successful for problems which have no initial feasible solution (see for example Lustig, Marsten, and Shanno 1989). A theoretical analysis of this method proved to be elusive for many years. Finally Zhang (1994) showed that a version of the infeasible-interior-point algorithm is globally convergent, and is actually an $O(n^2L)$ iteration (hence polynomial time) method if properly initialized. Here L is the bit size of a linear program with integer data. Unfortunately, however, this "polynomial time" initialization requires essentially the value of M which would be needed if an artificial variable were added to the problem. Mizuno (1994) subsequently obtained an $O(n^2L)$ bound for the infeasible-interior-point algorithm of Kojima, Megiddo, and Mizuno (1993), while Mizuno (1994) and Potra (1994, 1996) obtain an improved $O(nL)$ iteration result for infeasible-interior-point predictor-corrector algorithms.

The purpose of this paper is to obtain a probabilistic result for an infeasible-interior-point algorithm, endowed with the finite termination scheme of Ye (1992), applied to instances of Todd's degenerate model. As mentioned above, an infeasible-interior-point algorithm is a natural solution technique for instances of the degenerate model since these problems possess no initial feasible solution. A very important feature of our analysis is that we use no information regarding an optimal solution in the initialization of the algorithm. In particular, because the optimal solution is known for instances of the model, it would be easy to use a "polynomial time" initialization which would greatly simplify our analysis. However, as mentioned in the discussion of M above, such an approach is clearly cheating. Instead, we use a "blind" initialization of the algorithm, which could be applied to any linear program. In the initial version of the paper, our main result was that for Zhang's

(1994) algorithm applied to Todd's degenerate model, the expected number of iterations before termination with an exact optimal solution is $O(n^2 \ln(n))$. For the final version of the paper we have modified our original analysis to obtain an improved $O(n \ln(n))$ bound, using the infeasible-interior-point predictor-corrector algorithm of Potra (1994) in place of Zhang's method. At the end of the paper we also describe how our analysis can be applied to other infeasible-interior-point methods. The methodology used to obtain these results is relatively complex, for a number of reasons. First, the analysis of finite termination is complicated by the infeasibility of the iterates. Second, properties of the initial solution, such as "gap" and amount of infeasibility, are random variables. Third, due to our blind initialization, the global linear rate of improvement for the algorithm is itself a random variable. Fourth and finally, this random rate of improvement is dependent on other random variables connected with the initial solution, and finite termination criterion, resulting in product terms which cannot be simply factored (as would be the case with independence) in the expected value computation.

Subsequent to the initial version of this paper, an $O(\sqrt{n}L)$ infeasible-interior-point algorithm was devised by Ye, Todd and Mizuno (1994). The method of Ye, Todd, and Mizuno is based on an ingenious "homogenous self-dual" formulation for LP problems. The resulting algorithm is "infeasible" in the sense that iterates are infeasible for the original LP being solved, but is fundamentally different from the other infeasible-interior-point algorithms discussed above because the iterates *are* feasible for the homogenous self-dual problem. Anstreicher et al. (1992a) uses a number of results from this paper to obtain a bound of $O(\sqrt{n}\ln(n))$ for the expected number of iterations before termination with an exact optimal solution, for the algorithm of Ye, Todd, and Mizuno (1994) applied to instances of Todd's degenerate model.

2. The Infeasible-Interior-Point Algorithm

In this section we describe the main features of the infeasible-interior-point algorithm of Potra (1994). We assume familiarity with Potra's paper, and give major theoretical results concerning the algorithm without proof. Our notation generally follows Potra's, with a few minor changes to avoid conflicts with notation used in our later analysis. Throughout the paper, if $x \in \mathbf{R}^n$, then X is used to denote the diagonal matrix $X = \mathrm{diag}(x)$; similarly for s

and S, etc.. We use e to denote a vector of varying dimension with each component equal to one, and $\|\cdot\|$ to denote $\|\cdot\|_2$. Consider then primal and dual linear programs:

$$\begin{aligned}
&\text{(LP)} \qquad \min\{c^\top x \mid Ax = b,\ x \geq 0\},\\
&\text{(LD)} \qquad \max\{b^\top y \mid A^\top y + s = c,\ s \geq 0\}
\end{aligned}$$

where A is an $m \times n$ matrix with independent rows. We assume throughout that $n \geq 2$. A problem equivalent to the linear programs LP and LD, is then the LCP: $F(x, s, y) = 0$, $x \geq 0$, $s \geq 0$, where

$$F(x, s, y) = \begin{pmatrix} Xse \\ Ax - b \\ A^\top y + s - c \end{pmatrix}.$$

The algorithm is initiated with a point (x^0, s^0, y^0), $(x^0, s^0) > 0$, which is not assumed to satisfy the equality constraints of LCP. The algorithm generates a solution sequence (x^k, s^k, y^k) with $(x^k, s^k) > 0$ for $k \geq 0$. On iteration k, in the predictor step, a predictor-direction vector (u, v, w) is obtained by solving the Newton system

$$F'(x^k, s^k, y^k) \begin{pmatrix} u \\ v \\ w \end{pmatrix} = -F(x^k, s^k, y^k).$$

A step is then taken to a new point $(\bar{x}, \bar{s}, \bar{y})$, where for $0 \leq \bar{\theta}_k \leq 1$

$$(\bar{x}, \bar{s}, \bar{y}) = (x^k, s^k, y^k) + \bar{\theta}_k (u, v, w).$$

In the corrector step, we first find the solution $(\bar{u}, \bar{v}, \bar{w})$ of the linear system

$$F'(\bar{x}, \bar{s}, \bar{y}) \begin{pmatrix} \bar{u} \\ \bar{v} \\ \bar{w} \end{pmatrix} = -\begin{pmatrix} Uv \\ 0 \\ 0 \end{pmatrix},$$

and define $(\tilde{x}, \tilde{s}, \tilde{y}) = (\bar{x}, \bar{s}, \bar{y}) + \bar{\theta}_k^2 (\bar{u}, \bar{v}, \bar{w})$, $\tilde{\mu} = \tilde{x}^\top \tilde{s}/n$. Next we find the solution $(\tilde{u}, \tilde{v}, \tilde{w})$ of the linear system

$$F'(\bar{x}, \bar{s}, \bar{y}) \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{pmatrix} = \begin{pmatrix} \tilde{\mu} e - \tilde{X} \tilde{s} \\ 0 \\ 0 \end{pmatrix},$$

and finally set

$$(x^{k+1}, s^{k+1}, y^{k+1}) = (\tilde{x}, \tilde{s}, \tilde{y}) + (\tilde{u}, \tilde{v}, \tilde{w}).$$

Note that the two linear systems solved in the corrector step have the same coefficient matrix, so that only one matrix factorization is needed for the corrector step. Potra's predictor-corrector algorithm is a generalization of the Mizuno-Todd-Ye (1993) predictor-corrector algorithm, designed so that both "optimality" and "feasibility" are improved at the same rate, in the sense that with $p^k = b - Ax^k$ and $q^k = c - A^\top y^k - s^k$, the algorithm obtains

$$(x^{k+1})^\top s^{k+1} = (1-\bar{\theta}_k)(x^k)^\top s^k, \quad p^{k+1} = (1-\bar{\theta}_k)p^k, \quad q^{k+1} = (1-\bar{\theta}_k)q^k. \tag{2.1}$$

Given constants α and β such that

$$0 < \beta^2/(2\sqrt{2}(1-\beta)) < \alpha < \beta < 1, \tag{2.2}$$

the steplength $\bar{\theta}_k$ is chosen by a specific rule (see Potra 1994) that guarantees that

$$\begin{aligned} \| \bar{X}(\theta)\bar{s}(\theta) - \bar{\mu}(\theta) \| \le \beta\bar{\mu}(\theta), \qquad 0 \le \theta \le \bar{\theta}_k, \\ \| X^{k+1}s^{k+1} - \mu^{k+1}e \| \le \alpha\mu_{k+1}. \end{aligned} \tag{2.3}$$

In (2.3), $\bar{x}(\theta)$ and $\bar{s}(\theta)$ represent the predictor step parameterized by the steplength θ, $\bar{\mu}(\theta) = \bar{x}(\theta)^\top \bar{s}(\theta)/n$, and $\mu_{k+1} = {x^{k+1}}^\top s^{k+1}/n$. The parameters α and β in (2.3) enforce centering conditions on all iterates of the algorithm, i.e., all iterates are forced to lie in two cones around the central path. Clearly $\alpha = 0.25$ and $\beta = 0.5$ satisfy (2.2). Throughout the remainder of the paper, we will use this choice of α and β so as to simplify the exposition . We will also assume throughout that the initial solution has the form $(x^0, s^0, y^0) = (\rho e, \rho e, 0)$, for a scalar $\rho \ge 1$. Note that (2.3) implies that $(x^{k+1}, s^{k+1}) > 0$, unless the steplength $\bar{\theta}_k = 1$ leads directly to a solution of LCP. Suppose LP and LD have optimal solutions, say $\hat{x}$ and $(\hat{y}, \hat{s})$. Potra's analysis uses several scalar parameters, which for the particular (x^0, s^0, y^0), α, and β considered here specialize to:

$$
\begin{aligned}
\sigma_p &= 1 + \| A^+ b \|_\infty / \rho, \\
\sigma_d &= 1 + \| c \|_\infty / \rho, \\
\eta_p &= \frac{2}{\sqrt{3}} [2n + (\| \hat{x} \|_1 + \| \hat{s} \|_1)/\rho] \sigma_p, \\
\eta_d &= \frac{2}{\sqrt{3}} [2n + (\| \hat{x} \|_1 + \| \hat{s} \|_1)/\rho] \sigma_d, \\
\gamma_7 &= \frac{4}{3} \left((1 + n^{-.5}) + \sqrt{(1 + n^{-.5})^2 + 3} \right), \\
\lambda &= \frac{\sqrt{2}}{2} \left\{ \frac{n}{2} + 2\eta_p \eta_d + \sqrt{2}(\eta_p^2 + \eta_d^2) + (\sqrt{2} + 1)\sqrt{n(\eta_p^2 + \eta_d^2)} \right\},
\end{aligned}
\tag{2.4}
$$

where $A^+ = A^\top (AA^\top)^{-1}$. A major component in Potra's analysis of global convergence is the following result, which follows from his Lemmas 3.2 and 3.3, specialized for the particular case considered here:

Proposition 2.1. If $\hat{x}$ and $\hat{s}$ are optimal solutions of LP and LD, then there is a feasible steplength $\bar{\theta}_k \geq \delta$, where

$$\delta = \min \left\{ \frac{2}{1 + \sqrt{1 + 4\gamma_7 \lambda}}, \frac{0.321}{\sqrt{\lambda}} \right\}.$$

From Proposition 2.1, and the fact that $\gamma_7 = O(1)$, it is clear that the key quantity in the analysis of the algorithm is λ. In general λ is a fixed finite number, implying that the algorithm globally converges with a linear rate. Now let $\hat{\rho} = \| \hat{x} + \hat{s} \|$, for an optimal solution $(\hat{x}, \hat{s})$. Note that $\| \hat{x} \|_1 + \| \hat{s} \|_1 = \| \hat{x} + \hat{s} \|_1 \leq \sqrt{n} \| \hat{x} + \hat{s} \| = \sqrt{n}\hat{\rho}$. It is then immediate that if the parameter ρ that defines the starting point (x^0, s^0) is big enough, in the sense that

$$\rho \geq \max \left\{ \| A^+ b \|_\infty, \| c \|_\infty, \hat{\rho}/\sqrt{n} \right\},$$

then $\lambda = O(n^2)$, implying that $\delta = \Omega(1/n)$ and therefore the algorithm attains $O(nL)$ polynomial time complexity. Unfortunately, however, specifying ρ in this manner requires knowledge of $\hat{\rho}$, which is tantamount to knowledge of the required value of M when LP is solved by simply adding an artificial variable. Our analysis of the algorithm will not require such knowledge, but will instead use the fact that (2.4) implies that so long as $\rho \geq \max \{ \| A^+ b \|_\infty, \| c \|_\infty, 1 \}$,

$$\eta_p = O(n + \sqrt{n}\hat{\rho}), \quad \eta_d = O(n + \sqrt{n}\hat{\rho}), \quad \lambda = O([n + \sqrt{n}\hat{\rho}]^2), \quad \delta = \Omega(1/[n + \sqrt{n}\hat{\rho}]). \tag{2.5}$$

3. Finite Termination

In this section we consider the issue of finite termination of the infeasible-interior-point algorithm of Section 2, using the projection termination scheme of Ye (1992) (see also Mehrotra and Ye 1993). As in Ye (1992), our analysis requires the assumption that optimal solutions of LP and LD exist. We require a careful derivation of the technique, modified to deal with infeasibility of the iterates, for our probabilistic analysis in Section 5. The bounds obtained in this section are not necessarily the simplest, or tightest, possible, but are specifically chosen for applicability in our probabilistic analysis.

To begin, let $(\hat{x}, \hat{s}, \hat{y})$ be an optimal strictly complementary solution of LP/LD, that is, $\hat{x} + \hat{s} > 0$, and let $\hat{\rho} = \| \hat{x} + \hat{s} \|$. Let $\hat{\xi} = \min_j \{\hat{x}_j + \hat{s}_j\}$, $\hat{\sigma} = \{j \mid \hat{x}_j > 0\}$. We refer to $\hat{\sigma}$ as the "optimal partition." As in the previous section, we assume that $(x^0, s^0, y^0) = (\rho e, \rho e, 0)$, where $\rho \geq 1$. Our goal is to use the iterates (x^k, s^k) of the infeasible primal-dual algorithm to eventually identify the optimal partition, and generate exact optimal solutions of LP and LD. To begin, we characterize at what point the algorithm can correctly identify $\hat{\sigma}$. In the following analysis it is convenient to define $\psi_k = \prod_{i=0}^{k-1}(1 - \bar{\theta}_i)$, where $\bar{\theta}_k$ is the steplength used on the predictor step of the algorithm in iteration k.

Lemma 3.1. In order to obtain $s_j^k < x_j^k$, $j \in \hat{\sigma}$, and $x_j^k < s_j^k$, $j \notin \hat{\sigma}$, it suffices to have

$$x^{k\top} s^k \leq \frac{1}{3n} \left(\frac{\hat{\xi}}{(1 + \hat{\rho}/\sqrt{n})} \right)^2 . \tag{3.1}$$

Proof: From (2.1) we have $(Ax^k - b) = \psi_k(Ax^0 - b)$ and $(c - A^\top y^k - s^k) = \psi_k(c - A^\top y^0 - s^0)$, from which it follows that

$$A(x^k - \psi_k x^0)/(1 - \psi_k) = b, \qquad A^\top(y^k - \psi_k y^0)/(1 - \psi_k) + (s^k - \psi_k s^0)/(1 - \psi_k) = c.$$

Then $A\hat{x} = b$ and $A^\top \hat{y} + \hat{s} = c$ together imply that

$$(\psi_k x^0 + (1 - \psi_k)\hat{x} - x^k)^\top (\psi_k s^0 + (1 - \psi_k)\hat{s} - s^k) = 0,$$

which can be re-written as

$$(1 - \psi_k)(\hat{x}^\top s^k + \hat{s}^\top x^k) = n\psi_k^2 \rho^2 + x^{k\top} s^k + \rho\psi_k(1 - \psi_k)(e^\top \hat{x} + e^\top \hat{s}) - \rho\psi_k(e^\top x^k + e^\top s^k).$$

Using the facts that $(x^k, s^k) \geq 0$, $e^\top \hat{x} + e^\top \hat{s} \leq \sqrt{n}\hat{\rho}$, and ${x^k}^\top s^k = \psi_k {x^0}^\top s^0 = n\psi_k \rho^2$, we then obtain

$$\hat{x}^\top s^k + {x^k}^\top \hat{s} \leq ({x^k}^\top s^k)\left(\frac{1+\psi_k}{1-\psi_k} + \frac{\hat{\rho}}{\sqrt{n}\rho}\right).$$

Now assume that (3.1) holds. Note that $\hat{\rho} \geq \sqrt{n}\hat{\xi}$, so

$$\psi_k = {x^k}^\top s^k/(\rho^2 n) < 1/(3n^2) \leq 1/12,$$

for $\rho \geq 1$ and $n \geq 2$, and therefore

$$\hat{x}^\top s^k + \hat{s}^\top x^k < {x^k}^\top s^k \left(\frac{13}{11} + \frac{\hat{\rho}}{\rho\sqrt{n}}\right) < 1.2(1 + \hat{\rho}/\sqrt{n}){x^k}^\top s^k. \tag{3.2}$$

From (3.1) and (3.2), for $j \in \hat{\sigma}$ we then have

$$\begin{aligned} \hat{\xi} s_j^k \leq \hat{x}_j s_j^k &< \frac{.4}{n}\left(\frac{\hat{\xi}^2}{1+\hat{\rho}/\sqrt{n}}\right) \\ s_j^k &< \frac{.4\hat{\xi}}{n(1+\hat{\rho}/\sqrt{n})}. \end{aligned} \tag{3.3}$$

On the other hand, (3.2) implies that

$$(\hat{x}_j/x_j^k)(x_j^k s_j^k) < 1.2(1+\hat{\rho}/\sqrt{n}){x^k}^\top s^k.$$

Applying (2.3),

$$x_j^k > \frac{\hat{x}_j(x_j^k s_j^k)}{1.2(1+\hat{\rho}/\sqrt{n}){x^k}^\top s^k} \geq \frac{(1-\alpha)\hat{\xi}}{1.2n(1+\hat{\rho}/\sqrt{n})} > \frac{.6\hat{\xi}}{n(1+\hat{\rho}/\sqrt{n})}. \tag{3.4}$$

Combining (3.3) and (3.4), we have $x_j^k > s_j^k$, $j \in \hat{\sigma}$. The argument for $j \notin \hat{\sigma}$ is similar. $\square$

Next we consider the problem of generating an exact optimal solution to LP. (The analysis for obtaining a solution to LD is similar, and is omitted in the interest of brevity.) Given an iterate (x^k, s^k), let $B = B^k$ denote the columns of A having $x_j^k \geq s_j^k$, and let x_B denote the corresponding components of x. Similarly let N and x_N denote the remaining columns of A and components of x. The projection technique of Ye (1992) attempts to generate an optimal solution of LP by solving the primal projection problem

$$\text{(PP)} \qquad \min \quad \| x_B - x_B^k \| \qquad Bx_B = b.$$

A similar projection problem can be defined for the dual. Clearly if B corresponds to the optimal partition $\hat{\sigma}$, and the solution x_B^* of PP is nonnegative, then $(x_B, x_N) = (x_B^*, 0)$ is an optimal solution of LP. In what follows, we will choose k large enough so that, by Lemma 3.1, B does in fact correspond to the optimal partition $\hat{\sigma}$. Let B_1 be any set of rows of B having maximal rank ($B_1 = B$ if the rows of B are independent). Let N_1, A_1, and b_1 denote the corresponding rows of N and A, and components of b. Let B_{11} denote any square, nonsingular submatrix of B_1.

Theorem 3.2. The solution of PP generates an optimal solution of LP whenever

$$x^{k\top} s^k \leq \frac{\hat{\xi}^2}{3n(1+\hat{\rho}/\sqrt{n})^3(1+\sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \|)},$$

where A_{1j} denotes the jth column of A_1.

Proof: Note that if the assumption of the theorem is satisfied, then B corresponds to the optimal partition $\hat{\sigma}$ by Lemma 3.1. Clearly PP is equivalent to the problem

$$\begin{aligned} \min \quad & \| x_B - x_B^k \| \\ & B_1(x_B - x_B^k) = b_1 - B_1 x_B^k. \end{aligned}$$

The solution to PP, x_B^*, then satisfies

$$\begin{aligned} x_B^* - x_B^k &= B_1^\top (B_1 B_1^\top)^{-1}(b_1 - B_1 x_B^k) \\ &= B_1^\top (B_1 B_1^\top)^{-1}(N_1 x_N^k + b_1 - B_1 x_B^k - N_1 x_N^k) \\ &= B_1^\top (B_1 B_1^\top)^{-1} N_1 x_N^k + B_1^\top (B_1 B_1^\top)^{-1}(b_1 - A_1 x^k). \end{aligned}$$

$$\| x_B^* - x_B^k \| \leq \| B_1^\top (B_1 B_1^\top)^{-1} N_1 x_N^k \| + \| B_1^\top (B_1 B_1^\top)^{-1}(b_1 - A_1 x^k) \|. \tag{3.5}$$

Next we consider the two terms in (3.5). First, we have

$$\begin{aligned} \| B_1^\top (B_1 B_1^\top)^{-1} N_1 x_N^k \| &\leq \| B_{11}^{-1} N_1 x_N^k \| \\ &\leq \max_{j\notin\hat{\sigma}} \{x_j^k\} \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \| \\ &\leq 1.2\, x^{k\top} s^k \frac{(1+\hat{\rho}/\sqrt{n})}{\hat{\xi}} \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \|, \end{aligned} \tag{3.6}$$

where the first inequality uses the fact that $u^\top(B_1B_1^\top)^{-1}u \le u^\top(B_{11}B_{11}^\top)^{-1}u = \| B_{11}^{-1}u \|^2$ for any conforming vector u, and the last inequality uses (3.2) as in the proof of Lemma 3.1. To bound the second term of (3.5), we use $x^0 = \rho e,\ s^0 = \rho e,\ b_1 = B_1\hat{x}_B$, and the fact that $b - Ax^k = \psi_k(b - Ax^0)$ for all iterates k, to obtain

$$\begin{aligned}
\| B_1^\top(B_1B_1^\top)^{-1}(b_1 - A_1x^k) \| &= \psi_k\| B_1^\top(B_1B_1^\top)^{-1}(b_1 - A_1x^0) \| \\
&= \psi_k\| B_1^\top(B_1B_1^\top)^{-1}(B_1\hat{x}_B - B_1x_B^0 - N_1x_N^0) \| \\
&= \psi_k\| B_1^\top(B_1B_1^\top)^{-1}(B_1\hat{x}_B - \rho B_1e - \rho N_1e) \| \\
&\le \psi_k(\| B_1^\top(B_1B_1^\top)^{-1}B_1(\hat{x}_B - \rho e) \| + \rho\| B_1^\top(B_1B_1^\top)^{-1}N_1e \|) \\
&\le \psi_k(\| \hat{x}_B - \rho e \| + \rho\| B_{11}^{-1}N_1e \|) \\
&\le \psi_k(\hat{\rho} + \sqrt{n}\rho + \rho\sum_{j\notin\hat{\sigma}} \| B_{11}^{-1}A_{1j} \|).
\end{aligned}$$

Using $\psi_k = {x^k}^\top s^k/(\rho^2 n)$, we then certainly have

$$\| B_1^\top(B_1B_1^\top)^{-1}(b_1 - A_1x^k) \| \le {x^k}^\top s^k \left((1 + \hat{\rho}/\sqrt{n}) + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1}A_{1j} \| \right) . \tag{3.7}$$

Substituting (3.6) and (3.7) into (3.5), using $\hat{\rho} \ge \hat{\xi}\sqrt{n}$, we obtain

$$\| x_B^* - x_B^k \| \le 1.2\, {x^k}^\top s^k \frac{(1 + \hat{\rho}/\sqrt{n})^2}{\hat{\xi}} \left(1 + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1}A_{1j} \| \right) . \tag{3.8}$$

Finally (3.4) implies that if

$$\| x_B^* - x_B^k \| \le \frac{.6\,\hat{\xi}}{n(1 + \hat{\rho}/\sqrt{n})}, \tag{3.9}$$

then $x_B^* > 0$, as desired. But (3.8) and the hypothesis of the theorem imply (3.9), and the proof is complete. $\square$

4. Random Linear Programs

In this section we describe the random linear programming model to be used in our probabilistic analysis. We also describe an alternative version of the model, and briefly discuss the technical problems that arise if an analysis using the second version is attempted.

Todd's Degenerate Model, Version 1 (TDMV1): Let $A = (B, N)$, where B is $m \times n_1$, N is $m \times n_2$, $n_1 + n_2 = n$, $1 \leq n_1 \leq n - 1$, and each component of A is i.i.d. from the $N(0,1)$ distribution. Let

$$\hat{x} = \begin{pmatrix} \hat{x}_B \\ 0 \end{pmatrix}, \qquad \hat{s} = \begin{pmatrix} 0 \\ \hat{s}_N \end{pmatrix},$$

where the components of $\hat{x}_B$ and $\hat{s}_N$ are i.i.d. from the $|\,N(0,1)\,|$ distribution. Let $b = A\hat{x} = B\hat{x}_B$, and $c = \hat{s} + A^\top \hat{y}$, where the components of $\hat{y}$ are i.i.d. from any distribution with $O(1)$ mean and variance.

TDMV1 is a special case of Model 1 from Todd (1991). The simplest choice for $\hat{y}$ in the model is $\hat{y} = 0$. Note that in any case $\hat{x}_B > 0$ and $\hat{s}_N > 0$ with probability one, so $(\hat{x}, \hat{s})$ is an optimal, strictly complementary solution for LP/LD. If $n_1 = m$, then LP and LD are nondegenerate with probability one, but $n_1 < m$ results in a degenerate optimal solution for LP, and $n_1 > m$ results in a degenerate optimal solution for LD.

In the sequel we will analyze the behavior of the IIP algorithm of Section 2 applied to problems generated according to TDMV1, using the finite termination scheme of Section 3. In preliminary versions of the paper we also considered the following degenerate version of Todd's Model 1.

Todd's Degenerate Model, Version 2 (TDMV2): Let $A = (\hat{A}_1, \hat{A}_2, \hat{A}_3)$, where $\hat{A}_i$ is $m \times n_i$, $0 < n_1 \leq m$, $m \leq n_1 + n_2 < n$, $n_1 + n_2 + n_3 = n$, and each component of A is i.i.d. from the $N(0,1)$ distribution. Let

$$\hat{x} = \begin{pmatrix} \hat{x}_1 \\ 0 \\ 0 \end{pmatrix}, \qquad \hat{s} = \begin{pmatrix} 0 \\ 0 \\ \hat{s}_3 \end{pmatrix},$$

where the components of $\hat{x}_1$ and $\hat{s}_3$ are i.i.d. from the $|\,N(0,1)\,|$ distribution. Let $b = A\hat{x} = \hat{A}_1\hat{x}_1$, $c = \hat{s} + A^\top \hat{y}$, where the components of $\hat{y}$ are i.i.d. from any distribution with $O(1)$ mean and variance.

TDMV2 is described in Todd (1991, Section 4). Note that in TDMV2, $(\hat{x}, \hat{s})$ are clearly optimal solutions for LP/LD , but are *not* strictly complementary. Since our analysis of the finite termination scheme of Section 3 is based on a strictly complementary solution, to analyze the performance of our IIP algorithm on an instance of TDMV2 we would first need to characterize the properties of a strictly complementary solution (x^*, s^*). One approach to this problem, based on Section 7 of Ye (1994), proceeds as follows. As in Section 3, let B denote the columns of A corresponding to the optimal partition $\hat{\sigma}$, and let N denote the remaining columns of A. From Todd (1991, Proposition 4.2) we have either $B = \hat{A}_1$, or $B = (\hat{A}_1, \hat{A}_2)$, with probability one. Consider the case of $B = (\hat{A}_1, \hat{A}_2)$. Then the system

$$\hat{A}_1 x_1 + \hat{A}_2 x_2 = 0, \qquad x_2 \geq 0, x_2 \neq 0 \tag{4.1}$$

is feasible, and with probability one has a solution with $x_2 > 0$. By adjusting the signs of columns of $\hat{A}_1$ to form a new matrix $\tilde{A}_1$, we can assume that the system

$$\tilde{A}_1 x_1 + \hat{A}_2 x_2 = 0, \qquad x_1 \geq 0, x_2 \geq 0, (x_1, x_2) \neq 0 \tag{4.2}$$

is feasible, and with probability one has a solution with $(x_1, x_2) > 0$. In Ye (1994, Lemma 2) it is shown that if (4.2) is feasible then (4.2) must have a certain "basic feasible partition." Moreover, using a result of Todd (1991), the distribution of a solution to (4.2) given by a basic feasible partition can easily be determined (see the proof of Ye 1994, Theorem 4). Such a solution can then be used to construct an x^* so that $(x^*, \hat{s})$ are strictly complementary solutions to LP/LD. Unfortunately it was eventually pointed out to us by Mike Todd (private communication) that the above line of reasoning is incorrect, for a rather subtle reason. Essentially the problem is that taking a given basic partition for (4.2), and conditioning on that partition's feasibility, does *not* provide a valid distribution for a solution to (4.2) conditional on (4.2) being feasible. A similar problem occurs in a simpler context in Todd (1991, Theorem 3.6), and will be described in a forthcoming erratum to that paper.

Because of the above, references to results in earlier versions of this paper using TDMV2, in Anstreicher et al. (1992a) and Ye (1997), are incorrect. In particular, Proposition 4.1 of Anstreicher et al. (1992a), which is the basis of the probabilistic analysis in that paper, is invalid. However, it is very easy to modify the statement and proof of Lemma 4.2 of Anstreicher et al. (1992a) to apply using TDMV1 instead of TDMV2. As a result, Theorem 4.3, the main result of Anstreicher et al. (1992a), holds exactly as stated if "Todd's degenerate model" in the statement of the theorem is taken to be TDMV1, rather than TDMV2. Similarly the analysis of TDMV2 in Section 7 of Ye (1994) is incorrect, but Theorem 6, the main result of that section, can easily be shown to hold using TDMV1 in place of TDMV2.

5. Probabilistic Analysis

In this section we consider the performance of the infeasible-interior-point algorithm of Section 2, equipped with the finite termination criterion of Section 3, applied to the random linear program TDMV1 of Section 4. Given an instance of LP, we first obtain A^+b, the minimum norm solution of $Au = b$, a procedure which requires $O(n^3)$ total operations. We then set $\rho = 1 + \| A^+b \|_\infty + \| c \|_\infty$, ensuring $\rho \geq \max\{\| A^+b \|_\infty, \| c \|_\infty, 1\}$, and set $(x^0, s^0, y^0) = (\rho e, \rho e, 0)$. The algorithm is then applied until the projection technique of Section 3 yields an exact optimal solution of LP. Let

$$\epsilon = \frac{\hat{\xi}^2}{3n^2(1+\hat{\rho}/\sqrt{n})^3(1+\sum_{j\notin\hat{\sigma}} \| B_{11}^{-1}A_{1j} \|)}\,. \tag{5.1}$$

From Theorem 3.2, the algorithm will certainly terminate once $\mu_k = {x^k}^\top s^k/n \leq \epsilon$. Moreover, from Proposition 2.1, $\mu_k \leq (1-\delta)^k \mu_0 = \rho^2(1-\delta)^k$, so to obtain $\mu_k \leq \epsilon$ it certainly suffices to have

$$\begin{aligned}(1-\delta)^k\rho^2 &\leq \epsilon\\ k\ln(1-\delta) &\leq \ln(\epsilon) - 2\ln(\rho)\\ k &\geq [2\ln(\rho) - \ln(\epsilon)]/\delta,\end{aligned}$$

where the last inequality uses $\ln(1-\delta) \leq -\delta$. Finally, from (2.5) we have $1/\delta = O(n+\sqrt{n}\hat{\rho}) = O(n+\hat{\rho}^2)$, so termination of the algorithm definitely occurs on some iteration K, with

$$K = O\left((n+\hat{\rho}^2)(\ln(\rho) - \ln(\epsilon))\right). \tag{5.2}$$

By (5.2), to obtain bounds on $E[K]$ we require bounds on $E[(n+\hat{\rho}^2)\ln(\rho)]$, and $E[-(n+\hat{\rho}^2)\ln(\epsilon)]$. We obtain these bounds via a series of lemmas, below. Throughout we use $\| A \|$ to denote the Frobenius norm of a matrix A: $\| A \| = \| A \|_F = (\sum_{i,j} a_{ij}^2)^{1/2}$. It is then well known that for any matrix A and conforming vector x, $\| Ax \| \leq \| A \| \| x \|$. We also use $\chi^2(d)$ to denote a χ^2 random variable with d degrees of freedom.

Lemma 5.1. For an instance of TDMV1, $E[\hat{\rho}^2 \ln(1+\hat{\rho}^2)] = O(n \ln(n))$.

Proof: Note that $\hat{\rho}^2 \sim \chi^2(n)$, with mean n and variance $2n$. Let Q denote a random variable with the $\chi^2(n)$ distribution. Then

$$E[Q \ln(1+Q)] \leq E[Q \ln(n+Q)] = \ln(n)E[Q] + E[Q \ln(1+Q/n)] \leq \ln(n)E[Q] + E[Q^2]/n,$$

where the last inequality uses the fact that $\ln(1+a) \leq a$ for $a \geq 0$. The proof is completed by noting that $E[Q] = n$, $E[Q^2] = n^2 + 2n$. $\square$

Lemma 5.2. For an instance of TDMV1, $E[(n+\hat{\rho}^2)\ln(\rho)] = O(n \ln(n))$.

Proof: Note that $A\hat{x} = b$, so $\| A^+ b \| \leq \| \hat{x} \|$. Moreover $c = \hat{s} + A^\top \hat{y}$, so $\| c \| \leq \| \hat{s} \| + \| A^\top \hat{y} \|$. Since $\rho = 1 + \| A^+ b \|_\infty + \| c \|_\infty$, we immediately have

$$\rho \leq 1 + \| \hat{x} \| + \| \hat{s} \| + \| A^\top \hat{y} \| \leq 1 + 2\hat{\rho} + \| \hat{y} \| \, \| A \| \leq 2 + \hat{\rho}^2 + \| \hat{y} \| \, \| A \| \, .$$

Finally, we use the fact that $\ln(1+a+b) \leq \ln(1+a) + \ln(1+b)$ for $a \geq 0$, $b \geq 0$ to obtain

$$\ln(\rho) \leq \ln(2) + \ln(1+\hat{\rho}^2) + \ln(1 + \| \hat{y} \| \, \| A \|). \tag{5.3}$$

Now $\hat{\rho}^2 \sim \chi^2(n)$, so $E[\hat{\rho}^2] = n$, and $E[\hat{\rho}^2 \ln(1+\hat{\rho}^2)] = O(n \ln(n))$, by Lemma 5.1. Also $\| A \|^2 \sim \chi^2(mn)$, so $E[\| A \|^2] = mn \leq n^2$, $E[\| A \|] \leq n$, and furthermore $E[\| \hat{y} \|^2] = O(n)$, $E[\| \hat{y} \|] = O(\sqrt{n})$. Finally $\hat{\rho}^2$, $\| \hat{y} \|$, and $\| A \|$ are independent of one another. Combining all these facts with (5.3), and using $E[\ln(X)] \leq \ln(E[X])$ for any random variable X, we obtain $E[(n+\hat{\rho}^2)\ln(\rho)] = O(n \ln(n))$. $\square$

Lemma 5.3. For an instance of TDMV1, $E[-\hat{\rho}^2 \ln(\hat{\xi})] = O(n \ln(n))$.

Proof: By definition we have

$$\begin{aligned} E[-\hat{\rho}^2 \ln(\hat{\xi})] &= E[\hat{\rho}^2] \; E[-\ln(\hat{\xi})] - \mathrm{Cov}[\hat{\rho}^2, \ln(\hat{\xi})] \\ &\leq E[\hat{\rho}^2] \; E[-\ln(\hat{\xi})] + \sqrt{\mathrm{Var}[\hat{\rho}^2] \; \mathrm{Var}[\ln(\hat{\xi})]}. \end{aligned}$$

Since $\hat{\rho}^2 \sim \chi^2(n)$, we have $E[\hat{\rho}^2] = n$, $\mathrm{Var}[\hat{\rho}^2] = 2n$. It is also easily shown (see Lemma A.2 of the Appendix) that $E[-\ln(\hat{\xi})] = O(\ln(n))$, and $\mathrm{Var}[\ln(\hat{\xi})] = O(n)$. The lemma follows immediately. $\square$

Lemma 5.4. For an instance of TDMV1, $E[\ln(1 + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \|)] = O(\ln(n))$.

Proof: An application of the Cauchy-Schwarz inequality results in

$$\begin{aligned} 1 + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \| &\leq \sqrt{n}[1 + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \|^2]^{1/2} \\ &\leq \sqrt{n}[\sum_{j\notin\hat{\sigma}} (1 + \| B_{11}^{-1} A_{1j} \|^2)]^{1/2}. \end{aligned} \tag{5.4}$$

Results of Girko (1974) and Todd (1991) imply that for each $j \notin \hat{\sigma}$, we may write

$$\| B_{11}^{-1} A_{1j} \|^2 = \frac{\eta_j}{\nu_j^2},$$

where $\eta_j \sim \chi^2(m_1)$, and $\nu_j \sim | N(0,1) |$. Therefore

$$\sum_{j\notin\hat{\sigma}} (1 + \| B_{11}^{-1} A_{1j} \|^2) \leq n \max_{j\notin\hat{\sigma}} \frac{\nu_j^2 + \eta_j}{\nu_j^2} = n \max_{j\notin\hat{\sigma}} \frac{\tilde{\eta}_j}{\nu_j^2} \leq n \frac{\hat{\eta}}{\hat{\nu}^2}, \tag{5.5}$$

where $\tilde{\eta}_j \sim \chi^2(m_1 + 1)$, $\hat{\eta} = \max_{j\notin\hat{\sigma}} \tilde{\eta}_j$, $\hat{\nu} = \min_{j\notin\hat{\sigma}} \nu_j$. Combining (5.4) and (5.5) we obtain

$$\ln(1 + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \|) \leq \ln(n) + \frac{\ln(\hat{\eta})}{2} - \ln(\hat{\nu}).$$

However, in Lemmas A.2 and A.3 of the Appendix it is shown that $E[\ln(\hat{\eta})] = O(\ln(n))$, and $E[-\ln(\hat{\nu})] = O(\ln(n))$, and therefore $E[\ln(1 + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \|)] = O(\ln(n))$. $\square$

Lemma 5.5. For an instance of TDMV1, $E[-(n + \hat{\rho}^2)\ln(\epsilon)] = O(n\ln(n))$.

Proof: From (5.1), we have

$$-\ln(\epsilon) = \ln(3) + 2\ln(n) + 3\ln(1 + \hat{\rho}/\sqrt{n}) + \ln(1 + \sum_{j\notin\hat{\sigma}} \| B_{11}^{-1} A_{1j} \|) - 2\ln(\hat{\xi}). \tag{5.6}$$

Note that $1+\hat{\rho}/\sqrt{n} \le (1+\hat{\rho})^2 = O(1+\hat{\rho}^2)$, $E[\hat{\rho}^2] = n$, $E[\ln(1+\hat{\rho}^2)] = O(\ln(n))$, and $E[\hat{\rho}^2 \ln(1+\hat{\rho}^2)] = O(n\ln(n))$, from Lemma 5.1. Furthermore $E[-\ln(\hat{\xi})] = O(\ln(n))$, and $E[-\hat{\rho}^2 \ln(\hat{\xi})] = O(n\ln(n))$, from Lemma 5.3. Finally $E[\ln(1+\sum_{j\notin\hat{\sigma}} \| B_{11}^{-1}A_{1j} \|)] = O(\ln(n))$, from Lemma 5.4, and moreover $\sum_{j\notin\hat{\sigma}} \| B_{11}^{-1}A_{1j} \|$ is independent of $\hat{\rho}$. Combining these facts with (5.6) we immediately obtain $E[-(n+\hat{\rho}^2)\ln(\epsilon)] = O(n\ln(n))$. □

Combining Lemmas 5.2 and 5.5 with (5.2), we arrive at the major result of the paper:

Theorem 5.6. Assume that the infeasible-interior-point algorithm of Section 2, equipped with the finite termination technique of Section 3, is applied to an instance of TDMV1. Then the expected number of iterations before termination with an exact optimal solution of LP is $O(n\ln(n))$.

Note that our analysis of $E[K]$ for our IIP algorithm applied to TDMV1 is complicated by dependencies between $\hat{\rho}$ and ρ, and between $\hat{\rho}$ and ϵ. These dependencies would not affect a simpler "high probability" analysis (see for example Ye 1994), since if a fixed collection of events each holds with high probability, then the joint event also holds with high probability, regardless of dependencies. (The events of interest here are that $\hat{\rho}$, $\ln(\rho)$, and $\ln(\epsilon)$, satisfy certain bounds.) In the interest of brevity we omit the details of a high probability analysis of K, which also obtains a bound of $O(n\ln(n))$ iterations using TDMV1.

6. Application to Other Algorithms

A large literature on the topic of infeasible-interior-point methods for linear programming, and related problems, has developed since this paper was first written. See for example Bonnans and Potra (1994) for a discussion of the convergence properties for a broad class of such methods. In this section we describe the key features of Potra's (1994) algorithm that are exploited in our probabilistic analysis, and discuss the extent to which our analysis can be applied to a number of other infeasible-interior-point methods.

To begin, as described in Section 2, the algorithm of Potra (1994) satisfies

$$\begin{gathered} {x^k}^\top s^k = \psi_k {x^0}^\top s^0, \quad p^k = \psi_k p^0, \quad q^k = \psi_k q^0, \\ \| X^k s^k - \mu_k e \| \le \alpha\mu_k, \end{gathered} \tag{6.1}$$

where $p^k = b - Ax^k$, $q^k = c - A^\top y^k - s^k$, $\mu_k = {x^k}^\top s^k/n$, and $\psi_k = \prod_{i=0}^{k-1}(1-\bar\theta_i)$. However, it is straightforward to verify that the analysis of finite termination, in Section 3, continues to hold if the conditions in (6.1) are relaxed to

$$\begin{gathered} {x^k}^\top s^k = \psi_k {x^0}^\top s^0, \quad p^k = \bar\psi_k p^0, \quad q^k = \bar\psi_k q^0, \quad \bar\psi_k \le \psi_k, \\ x_i^k s_i^k \ge (1-\alpha)\mu_k, \quad i = 1, \ldots, n. \end{gathered} \tag{6.2}$$

The conditions in (6.2) are satisfied by almost all primal-dual infeasible-interior-point algorithms, and consequently the analysis in Section 3 applies very generally to these methods. (For simplicity we used $\alpha = .25$ throughout the paper, but obviously the analysis in Section 3 can be adapted to other α.) In Section 5, the important feature of Potra's (1994) algorithm, for our purposes, is that if

$$(x^0, s^0, y^0) = \rho(e, e, 0), \quad \text{where} \tag{6.3}$$

$$\rho = 1 + \| A^+ b \|_\infty + \| c \|_\infty, \tag{6.4}$$

then on each iteration k we have

$$\bar\theta_k \ge \Omega\left(\frac{1}{n + \sqrt{n}\hat\rho}\right), \tag{6.5}$$

where $\hat\rho = \| \hat x + \hat s \|$, and $(\hat x, \hat s)$ are any primal and dual optimal solutions. An initialization of the form (6.3) is quite standard for primal-dual infeasible-interior-point methods. The exact relationship between the initial normalization (6.4), and the lower bound on the steplength (6.5), does not carry over immediately to other methods. However, similar relationships between ρ and the convergence rate *do* hold for many other algorithms. For example, in the original version of this paper we used the fact that if $\rho \ge \| u^0, v^0 \|$, where $Au^0 = b$, $Fv^0 = Fc$ for any F whose rows span the nullspace of A, and $\rho = \Omega(1)$, then Zhang's (1994) method achieves

$$\bar\theta_k \ge \Omega\left(\frac{1}{n^2 + n\hat\rho^2}\right), \tag{6.6}$$

where $\bar\theta_k$ is the steplength used on iteration k. (The analysis of Zhang's algorithm is actually complicated somewhat by the fact that his proofs are based on the decrease of a "merit function" $\phi^k = {x^k}^\top s^k + \sqrt{\| p^k \|^2 + \| q^k \|^2}$, rather than decrease in the individual components ${x^k}^\top s^k$, $\| p^k \|$, and $\| q^k \|$. Consequently a lower bound on the steplength must be translated into a lower bound on the decrease in the merit function.) Note that here again an initialization similar to (6.4) fits the analysis well, since we can take $\rho = 1 + \| (u^0, v^0) \|$, where $u^0 = A^+ b$, $v^0 = c$. The difference between (6.5) and (6.6) results in a bound of $O(n^2 \ln(n))$ on the expected number of iterations before termination when Zhang's (1994) algorithm is applied to Todd's degenerate model. Our analysis could similarly be used to obtain an $O(n^2 \ln(n))$ expected iteration bound for the algorithms of Wright (1994), Zhang and Zhang (1994), and Wright and Zhang (1996). These three papers modify the method of Zhang (1994) to add asymptotic superlinear convergence to the algorithm. (The paper of Wright and Zhang obtains superquadratic convergence.) It is worth noting that applying the probabilistic analysis devised here to these methods ignores their improved asymptotic behavior. An interesting line of further research would attempt to exploit the superlinear convergence of these algorithms in the probabilistic analysis.

Our analysis could also be applied to the algorithms of Mizuno (1994), whose work is based on the infeasible-interior-point method of Kojima, Megiddo, and Mizuno (1993). Mizuno's algorithms include a termination condition that halts the iterative process if it can be proved that

$$\| (x^*, s^*) \|_\infty > \rho \tag{6.7}$$

for all optimal solutions (x^*, s^*). With a "polynomial time" initialization, involving a very large ρ, (6.7) has no effect when the algorithm is applied to a problem having an optimal solution. However, to perform an analysis similar to the one here one would need to bound the probability of termination due to (6.7), and possibly consider restarting the algorithm with a larger ρ, following termination(s) due to (6.7), until the finite projection technique yielded exact optimal solutions. A probabilistic analysis involving such restarts is undoubtedly possible, but we have not attempted to work out the details. Infeasible-interior-point potential reduction algorithms are devised by Mizuno, Kojima, and Todd (1995). These methods are quite similar to other primal-dual infeasible-interior-point methods, except that a potential function is used to motivate the search directions, and prove convergence. The algorithms developed by Mizuno, Kojima, and Todd (1995) also use the added termination condition (6.7). As a result, to apply our probabilistic analysis to these methods one would again need to bound the probability of termination due to (6.7), and possibly consider a "restart" strategy as described above. In addition, Algorithms II and III of Mizuno, Kojima, and Todd (1995) do *not* explicitly enforce the "feasibility before optimality" condition $\bar{\psi}^k \le \psi^k$ in (6.2), and therefore our analysis of finite termination, in Section 3, would not immediately apply to these methods.

Freund (1996) devises an infeasible-interior-point method that uses search directions based on a primal barrier function, as opposed to the primal-dual equations used by all the methods mentioned above. Freund's complexity analysis is also given in terms of explicit measures of the infeasibility and nonoptimality of the starting point. A probabilistic analysis of this algorithm would require substantial modifications of the techniques used here. Potra (1994) also shows that the complexity of his method can be improved if the infeasibility of the initial point is sufficiently small, but our analysis based on the initialization (6.3)-(6.4) ignores this refinement.

Acknowledgement

We are very grateful to the referees for their careful readings of the paper, and numerous comments which substantially improved it. We are indebted to Mike Todd for pointing out the error in our original analysis using the second version of his degenerate model for linear programming.

Appendix

In this appendix we provide several simple probability results which are required in the analysis of Section 5. Throughout, $x \in \mathbf{R}^k$ is a random vector whose components are *not* assumed to be independent of one another. We begin with an elementary proposition whose proof is omitted.

Proposition A.1. Let x_i, $i = 1, \ldots, k$ be continuous random variables, with sample space $[0, \infty)$. Define the new random variables $\xi = \min(x) = \min_i\{x_i\}$, and $\zeta = \max(x) = \max_i\{x_i\}$. Then for any $u \ge 0$, $f_\xi(u) \le \sum_{i=1}^k f_{x_i}(u)$, and $f_\zeta(u) \le \sum_{i=1}^k f_{x_i}(u)$, where $f_X(\cdot)$ is the p.d.f. of a random variable X.

Lemma A.2. Let $x \in \mathbf{R}^k$, where each $x_j \sim |\,N(0,1)\,|$, $j = 1,\dots,k$. Let $\xi = \min(x)$, $\zeta = \max(x)$. Then $E[-\ln(\xi)] = O(\ln(k))$, $\mathrm{Var}[\ln(\xi)] = O(k)$, and $E[\ln(\zeta)] = O(\ln(k))$.

Proof:

$$\begin{aligned}
E[\ln(\xi)] &= \int_0^\infty \ln(u) f_\xi(u)\,du \\
&= \int_0^{\frac{1}{k}} \ln(u) f_\xi(u)\,du + \int_{\frac{1}{k}}^\infty \ln(u) f_\xi(u)\,du \\
&\geq -\ln(k) - \int_0^{\frac{1}{k}} |\ln(u)\,| f_\xi(u)\,du.
\end{aligned}$$

Applying Proposition A.1, with each x_j having p.d.f $f(u) = \sqrt{2/\pi}\exp(-u^2/2)$, we obtain

$$\begin{aligned}
\int_0^{\frac{1}{k}} |\ln(u)\,| f_\xi(u)\,du &\leq k\sqrt{2/\pi}\int_0^{\frac{1}{k}} |\ln(u)\,| \exp(-u^2/2)\,du \\
&\leq k\sqrt{2/\pi}\int_0^{\frac{1}{k}} |\ln(u)\,|\,du \\
&= k\sqrt{2/\pi}(1+\ln(k))/k \\
&< 1 + \ln(k),
\end{aligned}$$

and therefore $E[-\ln(\xi)] = O(\ln(k))$. Using Proposition A.1 in a similar way,

$$\begin{aligned}
\mathrm{Var}[\ln(\xi)] &= E[\ln^2(\xi)] - (E[\ln(\xi)])^2 \\
&\leq \int_0^\infty \ln^2(u) f_\xi(u)\,du \\
&\leq k\sqrt{2/\pi}\int_0^\infty \ln^2(u)\exp(-u^2/2)\,du \\
&= O(k).
\end{aligned}$$

Finally, from Jensen's inequality and Proposition A.1,

$$E[\ln(\zeta)] \leq \ln(E[\zeta]) = \ln\left(\int_0^\infty u f_\zeta(u)\,du\right) \leq \ln\left(k\int_0^\infty u f(u)\,du\right) = O(\ln(k)).$$

$\square$

Lemma A.3. Let $x \in \mathbf{R}^k$, where each $x_j \sim \chi^2(d)$, $j = 1,\dots,k$. Let $\zeta = \max(x)$. Then $E[\ln(\zeta)] \leq \ln(k) + \ln(d)$.

Proof: This follows from the same analysis used to bound $E[\ln(\zeta)]$ in Lemma A.2, but letting $f(\cdot)$ be the p.d.f. of the $\chi^2(d)$ distribution, and recalling that the expected value of a $\chi^2(d)$ random variable is d. $\square$

References

Anstreicher, K. M. (1989). A combined phase I-phase II projective algorithm for linear programming. *Math. Programming* **43** 209-223.

Anstreicher, K. M. (1991). A combined phase I-phase II scaled potential algorithm for linear programming. *Math. Programming* **52** 429-439.

Anstreicher, K. M., J. Ji, F. A. Potra, Y. Ye (1992a). Average performance of a self-dual interior point algorithm for linear programming. In *Complexity in Numerical Optimization*, P. Pardalos (ed.), World Scientific, Singapore, 1-15.

Anstreicher, K. M., J. Ji, Y. Ye (1992b). *Average performance of an ellipsoid termination criterion for linear programming interior point algorithms,* Working paper 92-1, Dept. of Management Sciences, University of Iowa, Iowa City, IA.

Bonnans, J. F., F. A. Potra (1994). *Infeasible path following algorithms for linear complementarity problems,* Reports on Computational Mathematics 63, Dept. of Mathematics, University of Iowa, Iowa City, IA.

Freund, R. M. (1996). An infeasible-start algorithm for linear programming whose complexity depends on the distance from the starting point to the optimal solution. *Ann. Oper. Res.* **62** 29-58.

Girko, V. L. (1974). On the distribution of solution of systems of linear equations with random coefficients. *Theor. Probability and Math. Statist.* **2** 41-44.

Huang, S., Y. Ye (1991). *On the average number of iterations of the polynomial interior-point algorithms for linear programming,* Working paper 91-16, Dept. of Management Sciences, University of Iowa, Iowa City, IA.

Ishihara, T., M. Kojima (1993). On the big $\mathcal{M}$ in the affine scaling algorithm. *Math. Programming* **62** 85-93.

Ji, J., F. A. Potra (1992). *On the average complexity of finding an ϵ-optimal solution for linear programming,* Reports on Computational Mathematics No. 25, Dept. of Mathematics, University of Iowa, Iowa City, IA.

Kojima, M., N. Megiddo, S. Mizuno (1993). A primal-dual infeasible-interior-point algorithm for linear programming. *Math. Programming* **61** 263-280.

Kojima, M., S. Mizuno, A. Yoshise (1993). A little theorem of the big $\mathcal{M}$ in interior point algorithms. *Math. Programming* **59** 361-376.

Lustig, I. J., R. E. Marsten, D. F. Shanno (1991). Computational experience with a primal-dual interior point method for linear programming. *Linear Algebra and its Applications* **152** 191-222.

Mehrotra, S., Y. Ye (1993). Finding an interior point in the optimal face of linear programs. *Math. Programming* **62** 497-515.

Mizuno, S. (1994). Polynomiality of infeasible-interior-point algorithms for linear programming. *Math. Programming* **67** 109-119.

Mizuno, S., M. Kojima, M. J. Todd (1995). Infeasible-interior-point primal-dual potential-reduction algorithms for linear programming. *SIAM J. Optim.* **5** 52-67.

Mizuno, S., M. J. Todd, Y. Ye (1993). On adaptive-step primal-dual interior-point algorithms for linear programming. *Math. Oper. Res.* **18** 964-981.

Potra, F. A. (1994). A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points. *Math. Programming* **67** 383-406.

Potra, F. A. (1996). An infeasible interior-point predictor-corrector algorithm for linear programming. *SIAM J. Optim.* **6** 19-32.

Todd, M. J. (1991). Probabilistic models for linear programming. *Math. Oper. Res.* **16** 671-693; erratum to appear.

Todd, M. J. (1992). On Anstreicher's combined phase I-phase II projective algorithm for linear programming. *Math. Programming* **55** 1-15.

Todd, M. J. (1993). Combining phase I and phase II in a potential reduction algorithm for linear programming. *Math. Programming* **59** 133-150.

Wright, S. J. (1994). An infeasible interior point algorithm for linear complementarity problems. *Math. Programming* **67** 29-52.

Wright, S. J., Y. Zhang (1996). A superquadratic infeasible-interior-point algorithm for linear complementarity problems. *Math. Programming* **73** 269-289.

Ye, Y. (1992). On the finite convergence of interior-point algorithms for linear programming. *Math. Programming* **57** 325-335.

Ye, Y. (1994). Towards probabilistic analysis of interior-point algorithms for linear programming. *Math. Oper. Res.* **19** 38-52.

Ye, Y. (1997). *Interior Point Algorithms: Theory and Analysis.* Wiley-Interscience, New York.

Ye, Y., M. J. Todd, S. Mizuno (1994). An $O(\sqrt{n}L)-$iteration homogeneous and self-dual linear programming algorithm. *Math. Oper. Res.* **19** 53-67.

Zhang, Y. (1994). On the convergence of a class of infeasible interior-point algorithms for the horizontal linear complementarity problem. *SIAM J. Optim.* **4** 208-227.

Zhang, Y., D. Zhang (1994). Superlinear convergence of infeasible-interior-point methods for linear programming. *Math. Programming* **66** 361-377.