Available online at www.sciencedirect.com
ScienceDirect
Journal of Complexity 24 (2008) 214–227
Journal of COMPLEXITY
www.elsevier.com/locate/jco

On the probabilistic complexity of finding an approximate solution for linear programming

Jun Ji[a], Florian A. Potra[b,*]

[a] *Department of Mathematics, Kennesaw State University, Kennesaw, GA 30144, USA*
[b] *Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA*

Received 1 November 2006; accepted 8 May 2007
Available online 21 May 2007

Abstract

We consider the problem of finding an ε-optimal solution of a standard linear program with real data, i.e., of finding a feasible point at which the objective function value differs by at most ε from the optimal value. In the worst-case scenario the best complexity result to date guarantees that such a point is obtained in at most $O(\sqrt{n}|\ln\varepsilon|)$ steps of an interior-point method. We show that the expected value of the number of steps required to obtain an ε-optimal solution for a probabilistic linear programming model is at most $O(\min\{n^{1.5}, m\sqrt{n}\ln(n)\}) + \log_2(|\ln\varepsilon|)$.

Keywords: Linear programming; Interior-point algorithm; Average complexity; Approximate solution

1. Introduction

The advent of interior-point methods has revolutionized the field of mathematical programming. Most of the research has concentrated on the analysis of the worst-case behavior of interior-point methods. The state-of-the-art results show that a linear program with integer data can be solved in at most $O(\sqrt{n}L)$ iterations, where L is the length of a binary coding of the data. As each iteration requires at most $O(n^3)$ arithmetic operations, this implies that the corresponding interior-point methods are polynomial, in contrast with some variants of simplex method whose worst-case complexities are exponential. An inspection of the proof of the above mentioned complexity result shows that at most $O(\sqrt{n}L)$ iterations are required to obtain a strictly feasible (i.e., interior) point at which the objective function value differs by at most 2^{-L} from the optimal value. Then it is shown that by a rounding technique requiring at most $O(n^3)$ arithmetic operations it is possible to

* Corresponding author.
E-mail addresses: jji@kennesaw.edu (J. Ji), potra@math.umbc.edu (F.A. Potra).

0885-064X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jco.2007.05.001

obtain an (exact) optimal solution. The above mentioned complexity result has limited practical value, because for large scale linear programs the value of the parameter L is extremely large. Moreover, the above arguments do not apply to linear programs with real data. In many practical applications it is sufficient to obtain a feasible point x_ε such that the objective function value at x_ε differs by at most ε from the optimal value. Such a feasible point is called an ε-optimal solution. Standard results from the theory of interior-point methods imply that an ε-optimal solution can be obtained in at most $O(\sqrt{n}|\ln\varepsilon|)$ iterations for sufficiently small ε. It seems that this complexity result cannot be improved in the worst-case analysis setting. However, practical experiments show that interior-point algorithms work much better than shown by the worst-case complexity results. Different attempts to explain this phenomenon are contained in the works of Ye [22], Anstreicher et al. [2,3], and Huang [10] where a probabilistic analysis of the convergence of interior-point methods is presented. The analysis of [2,3] is performed on Todd's degenerate model, version 1 (TDMV1) introduced by Todd [17]. For Todd's non-degenerate model (Model 1 with $\hat{x} = \hat{s} = e$ [17, p. 677]), Huang and Ye [11], Anstreicher et al. [4], and Ji and Potra [14] obtained bounds on the average number of iterations needed by an interior-point method to find a solution of the linear programming problem, using various finite termination techniques. The analysis from [4,14] was based on a technical result of Huang and Ye [11]. Unfortunately, there is a subtle error in this technical result, which renders the proofs from [4,14] incorrect. For a detailed account of the history of the developments of this area, please refer to [3].

The paper [14] had already been accepted for publication when the error in [11] was discovered. Fortunately it was still possible to withdraw it from print. Recently, we found out that Huang [10] has managed to correct the error involved in the technical result of [11]. We have checked his new proof and we are convinced that it is correct. Since the result of Huang is published in a journal that is not easily accessible, we include a proof in the Appendix of this paper. In what follows we present a revised version of [14], based on the corrected technical result of Huang [10].

The complexity results of [2,3,10] are obtained not for the original interior-point methods, but for some hybrid algorithms consisting of an interior-point method combined with a finite termination criterion. In a more recent work by Todd et al. [18], a probabilistic analysis was performed for a primal-dual "layered-step" interior-point algorithm for linear programming [21]. In order to compare the average behavior of IPMs and the simplex methods, Huhn and Borgwardt [12,13] recently presented a probabilistic analysis on the rotation-symmetry model of Borgwardt [5,6]. Just like other aforementioned papers, the IPMs analyzed in [12,13] also involve a termination procedure.

In the present paper, we perform a probabilistic analysis for the "pure" interior-point method without using any "finite termination" criterion. We terminate the interior-point method whenever the primal-dual gap is less than ε. This is a very natural termination criterion and it is used by most practical implementations. In this way only an ε-optimal solution is obtained, but this is standard in most applications. We use the probabilistic linear programming model of Todd that was considered in the above mentioned paper of Huang [10]. We show that on this model the expected value of the number of iterations required by the "large step" predictor–corrector (P–C) algorithm of Mizuno et al. [15] is at most $O(\min\{n^{1.5}, m\sqrt{n}\ln(n)\}) + \log_2(|\ln\varepsilon|)$. Our proof uses some probabilistic results obtained in [10] as well as the quadratic convergence result obtained for the algorithm by Ye et al. [23].

Throughout the paper $\|\cdot\|$ denotes both the l_2-norm and the corresponding matrix norm. Also $X = \mathrm{diag}(x)$ for any vector x. Obviously, $\|X\| = \max\{|x_i| : i = 1,\ldots,n\} \neq \|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$.

2. The random LP model and a P–C algorithm

In [17] Todd introduced the following probabilistic linear programming model,

$$\text{(RLP)} \quad \min\{e^T x : Ax = b, x \geqslant 0\},$$

where $e = (1, \ldots, 1)^T \in \mathbf{R^n}$, $A \in \mathbf{R^{m\times n}}$ is a random matrix whose elements are identically independently drawn from $N(0, 1)$ and $b = Ae$. Let F be a matrix whose rows form a basis of $Ker(A)$, the null space of A. Then the dual problem can be written in terms of s alone (see [19]) as

$$\text{(RLD)} \quad \min\{e^T s : Fs = c, s \geqslant 0\},$$

where $c = Fe$ and the elements of F are also identically independently drawn from $N(0, 1)$. We note that the rows of A form a basis of $Range(A^T)$ which is orthogonal on $Ker(A)$. Given A, there are many methods for computing a basis of $Ker(A)$, but most of them will not lead to a matrix with elements being i.i.d. from $N(0, 1)$. However, from [17] it follows that there is a basis of $Ker(A)$ having this property.

In [17] it is shown that the feasible regions of both RLP and RLD are non-empty. Moreover, the problem is non-degenerate with probability one, and is endowed obviously with a natural initial starting feasible pair, i.e., $x^0 = e, s^0 = e$. In particular, this shows that the relative interiors of the feasible regions for RLP and RLD are non-empty, which allows application of interior-point methods.

In the present paper we consider the primal-dual predictor–corrector (P–C) method proposed by Mizuno et al. [15]. A typical iteration of the P–C algorithm begins with a pair (x, s) in a neighborhood of the central trajectory,

$$N(\alpha_1) = \left\{(x, s) \in \overset{\circ}{F} \,|\, \|Xs/\mu - e\| \leqslant \alpha_1,\ \mu = \frac{x^T s}{n}\right\}, \tag{1}$$

where α_1 is a constant between 0 and 1, and

$$\overset{\circ}{F} = \{(x, s) | Ax = b,\ Fs = c,\ x > 0,\ s > 0\}$$

denotes the set of all strictly feasible primal-dual feasible pairs. The predictor part of the P–C algorithm computes the primal-dual search directions u, v by solving a linear system of the form

$$\begin{aligned} Su + Xv &= -Xs, \\ Au &= 0, \\ A^T w + v &= 0 \end{aligned} \tag{2}$$

and a new pair is defined as

$$\begin{aligned} x(\theta) &= x + \theta u, \\ s(\theta) &= s + \theta v, \end{aligned} \tag{3}$$

for some $0 \leqslant \theta \leqslant 1$. Let $\mu(\theta) = x(\theta)^T s(\theta)/n$ and $X(\theta) = diag(x(\theta))$. The step-length is chosen by Mizuno et al. [15] to be the maximum value $\overline{\theta} \in (0, 1)$ such that

$$\|X(\theta)s(\theta)/\mu(\theta) - e\| \leqslant \alpha_2 < 1 \quad \text{for all } 0 \leqslant \theta \leqslant \overline{\theta}, \tag{4}$$

where α_2 is a constant bigger than α_1. The values used in [15,23] are actually $\alpha_1 = \frac{1}{4}$ and $\alpha_2 = \frac{1}{2}$. By a simple continuity argument it is proved that (4) implies $(x(\overline{\theta}), s(\overline{\theta})) \in N(\alpha_2)$ (see [15]). Let

us denote by $\hat{\theta}$ the largest value of $\overline{\theta} < 1$ such that (4) is satisfied. No closed-form expression for $\hat{\theta}$ has been given in the above mentioned papers. In fact, the results of [15] are obtained by using some coarse lower bound for $\hat{\theta}$. In the following lemmas we show that a simple closed-form expression for $\hat{\theta}$ can be obtained.

Lemma 2.1. *Let* $0 < \alpha_1 < \alpha_2 < 1$. *Suppose* $(x, s) \in N(\alpha_1)$, *let* u, v, w *be given by* (2), *and denote*

$$\delta = Uv/\mu, \tag{5}$$

(i) *if* $\delta = 0$, *then* $(x(\theta), s(\theta)) \in N(\alpha_2)$ *for all* $0 \leqslant \theta < 1$,
(ii) *if* $\delta \neq 0$, *then* (4) *holds if and only if* $\overline{\theta} \leqslant \hat{\theta}$, *where*

$$\hat{\theta} = \frac{-\hat{t} + \sqrt{\hat{t}^2 + 4\hat{t}}}{2}, \quad \hat{t} = \frac{-b + \sqrt{b^2 + ac}}{a}, \tag{6}$$

$$a = \|\delta\|^2, \quad b = q^T\delta, \quad c = \alpha_2^2 - \|q\|^2, \quad q = (Xs - \mu e)/\mu. \tag{7}$$

Proof. From the hypothesis, we have $\|q\| \leqslant \alpha_1$. Define the function $f : [0, 1) \longrightarrow \mathbf{R}$ as

$$f(\theta) = \|X(\theta)s(\theta)/\mu(\theta) - e\|^2,$$

where $\mu(\theta) = x(\theta)^T s(\theta)/n$. Since

$$X(\theta)s(\theta) = (1 - \theta)Xs + \theta^2 Uv \quad \text{and} \quad \mu(\theta) = (1 - \theta)\mu,$$

(see for example [15]) we have

$$f(\theta) = \|q + t\delta\|^2 = \|q\|^2 + 2tq^T\delta + t^2\|\delta\|^2 \equiv g(t),$$

where $t = h(\theta) \equiv \theta^2/(1 - \theta)$. Clearly, if $\delta = 0$, then $g(t) = \|q\|^2 \leqslant \alpha_1^2 \leqslant \alpha_2^2$ for all $0 \leqslant \theta < 1$. Otherwise, $f(\theta) \leqslant \alpha_2^2$ is satisfied if and only if $t \in [0, \hat{t}]$, where $\hat{t}$ given in (6) is the unique positive solution of $g(t) = \alpha_2^2$. From $t = h(\theta)$, it turns out that $f(\theta) \leqslant \alpha_2^2$ is satisfied if and only if $\theta \in [0, \hat{\theta}]$, where $\hat{\theta}$ given in (6) is the unique positive solution of $h(\theta) = \hat{t}$. Therefore, (4) holds if and only if $\overline{\theta} \leqslant \hat{\theta}$. □

It is easily seen that (6) and (7) implies

$$\hat{\theta} = \frac{2}{1 + \sqrt{1 + 4/\hat{t}}} \tag{8}$$

and

$$\begin{aligned} 1/\hat{t} &= \left(b + \sqrt{b^2 + ac}\right) \Big/ c \\ &= \left(q^T\delta + \sqrt{(q^T\delta)^2 + \|\delta\|^2(\alpha_2^2 - \|q\|^2)}\right) \Big/ (\alpha_2^2 - \|q\|^2) \\ &\leqslant (\|q\|\|\delta\| + \|\delta\|\alpha_2) \Big/ (\alpha_2^2 - \|q\|^2) \quad (\text{since } |q^T\delta| \leqslant \|q\|\|\delta\|) \\ &\leqslant \|\delta\|/(\alpha_2 - \alpha_1) \quad (\text{since } \|q\| \leqslant \alpha_1). \end{aligned} \tag{9}$$

By combining (8) with (9), we have

$$\hat{\theta} \geqslant \frac{2}{1+\sqrt{1+4\|\delta\|/(\alpha_2-\alpha_1)}}. \tag{10}$$

For $\alpha_1 = \frac{1}{4}$ and $\alpha_2 = \frac{1}{2}$ the above inequality leads to the bound used in [23]:

$$1-\hat{\theta} \leqslant 4\|\delta\|. \tag{11}$$

In what follows we will only consider the case when $\alpha_1 = \frac{1}{4}$ and $\alpha_2 = \frac{1}{2}$ so that (11) will always be satisfied.

It is clear that if $\delta = 0$ then $\theta = 1$ leads immediately to an optimal primal-dual solution. Consequently we will only consider the case when $\delta \neq 0$. Then, with $\hat{\theta}$ given by (6) we define the predictor pair

$$\hat{x} = x(\hat{\theta}) \quad \text{and} \quad \hat{s} = s(\hat{\theta}) \tag{12}$$

and the new pair (x^+, s^+) is obtained by solving the system

$$\begin{aligned} \hat{S}\hat{u} + \hat{X}\hat{v} &= \hat{\mu}e - \hat{X}\hat{s}, \\ A\hat{u} &= 0, \\ A^T\hat{w} + \hat{v} &= 0, \end{aligned} \tag{13}$$

with $\hat{\mu} = \hat{x}^T\hat{s}/n$, and by setting

$$x^+ = \hat{x} + \hat{u} \quad \text{and} \quad s^+ = \hat{s} + \hat{v}. \tag{14}$$

The following result can be proved as in [15, Lemma 3].

Lemma 2.2. *Suppose that the hypothesis of Lemma* 2.1 *is satisfied with* $\alpha_1 = \frac{1}{4}$ *and* $\alpha_2 = \frac{1}{2}$, (x^+, s^+) *is computed as in* (12)–(14) *and assume that* $\delta \neq 0$. *Then*

$$(x^+, s^+) \in N(\tfrac{1}{4}) \quad \text{and} \quad (x^+)^T s^+ = (1-\hat{\theta})x^T s. \tag{15}$$

The above lemma allows us to repeat the process and to obtain a sequence $(x^k, s^k) \in N(\frac{1}{4})$ such that

$$(x^{k+1})^T s^{k+1} = (1-\hat{\theta}_k)(x^k)^T s^k. \tag{16}$$

By using (10) and (16) one can prove (see [15]) that

$$(x^{k+1})^T s^{k+1} \leqslant \left(1 - \frac{C}{\sqrt{n}}\right)(x^k)^T s^k, \tag{17}$$

where $C = 8^{-.25} = 0.594603558\ldots$. The above relation implies that $(x^k)^T s^k \leqslant 2^{-L}$ for $k = O(\sqrt{n}L)$. Moreover, asymptotically the primal-dual gap $(x^k)^T s^k$ converges Q-quadratically to zero (cf. [23]).

In the next two sections, we will study the probabilistic behavior of the above P–C algorithm for Todd's random problems RLP and RLD. Our analysis will heavily depend on the upper bound in (11).

3. Technical results

In this section we assume that there is a unique non-degenerate optimal solution (x^*, s^*) to LP. This assumption holds with probability one for Todd's model described in the previous section. We denote $\{i : x_i^* > 0\}$ by β, and $\{i : s_i^* > 0\}$ by v. Since (x^*, s^*) is strictly complementary, we have

$$\beta \cap v = \emptyset \quad \text{and} \quad \beta \cup v = \{1, 2, \ldots, n\}.$$

Let the columns of A corresponding to the index set β form a matrix B, and the rest form a matrix N. Under our assumption the matrix B is non-singular with probability one. Given a vector $z \in \mathbf{R^n}$ and an index subset $W \subseteq \{1, \ldots, n\}$, we denote by z_W the vector of entries z_i, $i \in W$. Define

$$\xi_p = \min_{j\in\beta} x_j^*, \quad \xi_d = \min_{j\in v} s_j^*, \quad \text{and} \quad \xi = \min\{\xi_p, \xi_d\}.$$

Obviously, $\xi > 0$.

It has been shown that the sequence (x^k, s^k) generated by the P–C algorithms satisfies $(x^k, s^k) \in N(\frac{1}{4})$ (see Lemma 2.2). Therefore, we have

$$\frac{\min(X^k s^k)}{(x^k)^T (s^k)} \geqslant \frac{3}{4n}.$$

It is then straightforward to prove the following result of Güler and Ye [8]:

Lemma 3.1. *If the sequence (x^k, s^k) is obtained by the P–C algorithm, then*

$$x_j^k \geqslant \frac{3\xi}{4n} \quad \textit{for each } j \in \beta \quad \textit{and} \quad s_j^k \geqslant \frac{3\xi}{4n} \quad \textit{for each } j \in v.$$

Define $D = X^{1/2} S^{-1/2}$ and denote by P_L the orthogonal projection onto the linear subspace L of $\mathbf{R^n}$. It is easily seen (cf. [1]) that the solution vectors u and v corresponding to the system of linear equations (2) can be written as

$$u = -DP_{K(AD)}(XS)^{1/2} e \quad \text{and} \quad v = -D^{-1} P_{R(DA^T)}(XS)^{1/2} e, \tag{18}$$

where $K(AD)$ and $R(DA^T)$ denote the null space of the matrix AD and the range of the matrix DA^T, respectively.

Lemma 3.2. *If u and v are obtained from the linear system* (2), *then*

$$\|u_v\| \leqslant \frac{4n}{3\xi} x^T s \quad \textit{and} \quad \|v_\beta\| \leqslant \frac{4n}{3\xi} x^T s.$$

Proof. From (18), following the proof in [23], we obtain

$$\|(D^{-1}u)_v\| \leqslant \|P_{K(AD)}(XS)^{1/2} e\| \leqslant \|(XS)^{1/2} e\|.$$

Therefore,

$$\begin{aligned} \|u_v\| &= \|D_v D_v^{-1} u_v\| \leqslant \|D_v\| \|(D^{-1}u)_v\| \\ &\leqslant \|(XS)_v^{1/2} S_v^{-1}\| \|(XS)^{1/2} e\| \leqslant \|S_v^{-1}\| x^T s \leqslant \frac{4n}{3\xi} x^T s. \end{aligned}$$

In proving the last inequality we have used Lemma 3.1. The upper bound for $\|v_\beta\|$ can be proved similarly. □

Lemma 3.3. *If u and v are obtained from the linear system* (2), *then*

$$\|u_\beta\| \leqslant \frac{4n}{3\xi}\|B^{-1}N\|x^T s \quad \text{and} \quad \|v_v\| \leqslant \frac{4n}{3\xi}\|B^{-1}N\|x^T s.$$

Proof. From (2), we have $Bu_\beta = -Nu_v$. Therefore, we obtain

$$\|u_\beta\| = \| - B^{-1}Nu_v\| \leqslant \|B^{-1}N\|\|u_v\| \leqslant \frac{4n}{3\xi}\|B^{-1}N\|x^T s.$$

From (2), we also have $v_\beta = -B^T w$ and $v_v = -N^T w$. Hence, $v_v = N^T B^{-T} v_\beta$. Therefore,

$$\|v_v\| \leqslant \|N^T B^{-T} v_\beta\| \leqslant \|B^{-1}N\|\|v_\beta\|.$$

By virtue of Lemma 3.2 we obtain the desired result for $\|v_v\|$. □

Lemma 3.4. *Under the hypothesis of Lemma* 2.2 *we have*

$$(x^+)^T s^+ \leqslant (x^T s)^2 \cdot \Gamma,$$

where

$$\Gamma = \Gamma(x, s) \equiv \frac{64n^3\sqrt{n-m}\max_{j\in v}\|B^{-1}a_j\|}{9\xi^2}\left(1 + \frac{16n^2\sqrt{n-m}\max_{j\in v}\|B^{-1}a_j\|}{9\xi^2}x^T s\right)$$

and a_j denotes the jth column of the matrix A.

Proof. At each predictor step, from (2), we have

$$\frac{u_j}{x_j} + \frac{v_j}{s_j} = -1, \quad j = 1, \ldots, n. \tag{19}$$

From Lemmas 3.1 and 3.3, we have

$$\left|\frac{u_j}{x_j}\right| \leqslant \frac{16n^2\|B^{-1}N\|}{9\xi^2}x^T s = 16\frac{n^3\|B^{-1}N\|}{9\xi^2}\mu, \quad j \in \beta. \tag{20}$$

Combining (19) with (20), we obtain

$$\left|\frac{v_j}{s_j}\right| \leqslant 1 + \left|\frac{u_j}{x_j}\right| \leqslant 1 + \frac{16n^3\|B^{-1}N\|}{9\xi^2}\mu, \quad j \in \beta. \tag{21}$$

Similarly, we have

$$\left|\frac{v_j}{s_j}\right| \leqslant \frac{16n^3\|B^{-1}N\|}{9\xi^2}\mu, \quad \left|\frac{u_j}{x_j}\right| \leqslant 1 + \frac{16n^3\|B^{-1}N\|}{9\xi^2}\mu, \quad j \in v. \tag{22}$$

Hence,

$$\left|\frac{u_j v_j}{x_j s_j}\right| \leqslant \frac{16n^3\|B^{-1}N\|}{9\xi^2}\left(1 + \frac{16n^3\|B^{-1}N\|}{9\xi^2}\mu\right)\mu, \quad j \in \{1, \ldots, n\}.$$

Recall that $\delta = Uv/\mu$, so that

$$\begin{aligned}\|\delta\|^2 &\leqslant \sum_{j=1}^{n}\left[\frac{16n^3\|B^{-1}N\|}{9\xi^2}\left(1+\frac{16n^3\|B^{-1}N\|}{9\xi^2}\mu\right)x_j s_j\right]^2\\ &= \left[\frac{16n^3\|B^{-1}N\|}{9\xi^2}\left(1+\frac{16n^3\|B^{-1}N\|}{9\xi^2}\mu\right)\right]^2\sum_{j=1}^{n}(x_j s_j)^2\\ &\leqslant \left[\frac{16n^3\|B^{-1}N\|}{9\xi^2}\left(1+\frac{16n^3\|B^{-1}N\|}{9\xi^2}\mu\right)\right]^2(x^T s)^2.\end{aligned}$$

Also

$$\begin{aligned}\|B^{-1}N\|^2 &= \lambda_{\max}((B^{-1}N)^T(B^{-1}N))\\ &\leqslant \operatorname{trace}((B^{-1}N)^T(B^{-1}N))\\ &\leqslant \sum_{j\in v}((B^{-1}a_j)^T(B^{-1}a_j))\\ &\leqslant (n-m)\max_{j\in v}\|B^{-1}a_j\|^2.\end{aligned}$$

By combining the above inequalities we obtain the desired result. □

Lemma 3.4 essentially proves the Q-quadratic convergence of the primal-dual gap for a non-degenerate problem. As we mentioned before, the problem RLP is non-degenerate with probability one. The quadratic convergence of the primal-dual gap can be proved for degenerate problems as well, but the proof becomes substantially more complicated (see [23]). In the next section, we will use the upper bound given in Lemma 3.4 to analyze the probabilistic complexity of finding an ε-optimal solution for our model.

4. The average number of iterations

Let us first define $\Gamma_k = \Gamma(x^k, s^k)$, where Γ is as in Lemma 3.4. In order to estimate the number of iterations required to obtain an ε-optimal solution, let us consider two phases. First, we count the number of iterations required to reduce the duality gap such that

$$x^T s \leqslant \frac{9\xi^2}{256n^3\sqrt{n-m}\max_{j\in v}\|B^{-1}a_j\|}. \tag{23}$$

In the worst-case scenario (17) implies

$$(x^k)^T s^k \leqslant \left(1-\frac{C}{\sqrt{n}}\right)^k (x^0)^T s^0,$$

where $C = 8^{-.25}$. Note that $(x^0, s^0) = (e, e)$ is a feasible primal-dual pair. From the well-known inequality $\ln(1+x) \leqslant x$, it follows that (23) is satisfied after at most k_1 iterations, where

$$k_1 \equiv \left\lceil \frac{\sqrt{n}}{C}\ln\frac{4n}{\tau}\right\rceil, \quad \tau \equiv \frac{9\xi^2}{68n^3\sqrt{n-m}\max_{j\in v}\|B^{-1}a_j\|}, \tag{24}$$

and $\lceil\eta\rceil$ is the smallest integer greater than or equal to η. Since the duality gap is reduced at each step, $\Gamma_k \leqslant \Gamma_{k_1}$ for $k \geqslant k_1$. From Lemma 3.4 and (23), it follows that:

$$\begin{aligned}\Gamma_k &\leqslant \frac{64n^3\sqrt{n-m}\max_{j\in v}\|B^{-1}a_j\|}{9\xi^2}(1+1/(16n))\\ &\leqslant \frac{68n^3\sqrt{n-m}\max_{j\in v}\|B^{-1}a_j\|}{9\xi^2} = \tau^{-1} \quad \text{for any } k \geqslant k_1. \end{aligned} \tag{25}$$

From Lemma 3.4 and (25), we get

$$(x^{k+1})^T s^{k+1} \leqslant \tau^{-1}((x^k)^T s^k)^2 \quad \text{for any } k \geqslant k_1, \tag{26}$$

where τ is defined in (24). Multiplying both sides of (26) by τ^{-1} and setting $\omega^k \equiv \tau^{-1}(x^k)^T s^k$, we obtain that $\omega^{k+1} \leqslant (\omega^k)^2$ for $k \geqslant k_1$. It is easily seen from (23), (24) that $\omega^{k_1} \leqslant \frac{1}{3}$. If $\tau \leqslant \varepsilon$, then it follows from (23) and (24) that $(x^{k_1})^T s^{k_1} \leqslant \varepsilon$, which indicates an ε-optimal solution. Consequently we may assume that $\tau > \varepsilon$. Since ε represents the accuracy of the approximate solution it is reasonable to assume that $\varepsilon \leqslant \frac{1}{3}$. In fact we are interested in the case where ε is much smaller. From the k_1th iterate on we always have

$$\omega^k \leqslant (\omega^{k-1})^2 \leqslant \cdots \leqslant (\omega^{k_1})^{2^{k-k_1}} \leqslant (\tfrac{1}{3})^{2^{k-k_1}} \quad \text{for any } k \geqslant k_1.$$

Clearly, $(\frac{1}{3})^{2^{k-k_1}} \leqslant \tau^{-1}\varepsilon$ implies $\omega^k \leqslant \tau^{-1}\varepsilon$ which in turn implies that $(x^k)^T s^k \leqslant \varepsilon$. Hence, an ε-optimal solution is produced in at most

$$K = \lceil \log_2(\ln\tau - \ln\varepsilon) - \log_2 \ln 3\rceil + k_1, \tag{27}$$

iterations. From the definition of k_1 and (27), we have

$$\begin{aligned} K &= \log_2(\ln\tau - \ln\varepsilon) - \frac{\sqrt{n}}{C}\ln\tau + O(\sqrt{n}\ln n)\\ &= \log_2(|\ln\varepsilon|) + \log_2\left(1 + \frac{\ln\tau}{|\ln\varepsilon|}\right) - \frac{\sqrt{n}}{C}\ln\tau + O(\sqrt{n}\ln n)\\ &\leqslant \log_2(|\ln\varepsilon|) + \left(\frac{\sqrt{n}}{C} - \frac{1}{(\ln 2)|\ln\varepsilon|}\right)\ln(\tau^{-1}) + O(\sqrt{n}\ln n), \end{aligned} \tag{28}$$

where the last inequality uses $\log_2(1+t) = \ln(1+t)/\ln 2 \leqslant t/\ln 2$. The expected value $E[K]$ for K is bounded by

$$E[K] \leqslant \log_2(|\ln\varepsilon|) + \left(\frac{\sqrt{n}}{C} - \frac{1}{(\ln 2)|\ln\varepsilon|}\right) E[\ln(\tau)^{-1}] + O(\sqrt{n}\ln n). \tag{29}$$

Eq. (24) indicates that

$$\begin{aligned}\ln(\tau)^{-1} &= \ln\left(\frac{68n^3\sqrt{n-m}\max_{j\in v}\|B^{-1}a_j\|}{9\xi^2}\right)\\ &= O(\ln n) + \ln\left(\max_{j\in v}\|B^{-1}a_j\|\right) - 2\ln\xi. \end{aligned}$$

For the random model considered in this paper, any m components of $\{1, 2, \ldots, n\}$ form a basis (which may not be feasible). Thus, we choose m elements from n elements for $n \geqslant m$ and have

C_n^m basic bases, one of which is the optimal basis. For any fixed basis $\bar{B}$, let $\bar{\beta}$ be its index set and $\bar{\nu} = \{1, 2, \ldots, n\} \setminus \bar{\beta}$. As in the proof of [3, Lemma 5.4], we may write

$$\|\bar{B}^{-1}a_j\|^2 = \frac{\eta_j}{\zeta_j^2} \quad \text{for } j \in \bar{\nu},$$

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

$$\max_{j\in\nu} \|B^{-1}a_j\| \leqslant \max_{\bar{\beta}} \max_{j\in\bar{\nu}} \frac{\sqrt{\eta_j}}{\zeta_j} \leqslant \frac{\sqrt{\hat{\eta}}}{\hat{\zeta}}, \tag{30}$$

where

$$\hat{\eta} = \max_{\bar{\beta}} \max_{j\in\bar{\nu}} \eta_j \quad \text{and} \quad \hat{\zeta} = \min_{\bar{\beta}} \min_{j\in\bar{\nu}} \zeta_j$$

are the maximum and minimum of $(n-m)C_n^m$ random variables, respectively. Thus, it is seen from (30) that

$$\ln\left(\max_{j\in\nu} \|B^{-1}a_j\|\right) \leqslant \frac{\ln\hat{\eta}}{2} - \ln\hat{\zeta}.$$

However, in Lemmas A.2 and A.3 of the Appendix of [3], it is shown that $E[\ln\hat{\eta}] \leqslant \ln((n-m)C_n^m) + \ln(m)$, and $E[-\ln\hat{\zeta}] \leqslant O\left(\ln((n-m)C_n^m)\right)$, and therefore

$$E\left[\ln\left(\max_{j\in\nu} \|B^{-1}a_j\|\right)\right] \leqslant O\left(\ln(n)\right) + O\left(\ln(C_n^m)\right).$$

Also, from Proposition A.1 of the Appendix (compare with [10, Theroem3.1]) it follows that:

$$E[\ln\xi] \geqslant -2\ln(n) - 2\ln\left(C_n^m\right) - 1.$$

Hence, we have

$$\begin{aligned} E[\ln(\tau)^{-1}] &\leqslant O(\ln(n)) + O\left(\ln\left(C_n^m\right)\right) \\ &\leqslant O\left(\ln(n)\right) + O(\min\{n, m\ln(n)\}), \end{aligned} \tag{31}$$

where the last inequality follows from $C_n^m \leqslant 2^n$ and $C_n^m \leqslant n^m$. The above inequality together with

$$\frac{1}{\ln 2|\ln\varepsilon|} \leqslant \frac{\sqrt{n}}{C} \quad \text{for } \varepsilon \leqslant \frac{1}{3} \quad \text{and} \quad C = 8^{-.25}$$

implies

$$\begin{aligned} E[K] &\leqslant \left(\frac{\sqrt{n}}{C} - \frac{1}{\ln 2|\ln\varepsilon|}\right)[O\left(\ln(n)\right) + O\left(\min\{n, m\ln(n)\}\right)] \\ &\quad + O(\sqrt{n}\ln n) + \log_2(|\ln\varepsilon|) \\ &= \left(\frac{\sqrt{n}}{C} - \frac{1}{\ln 2|\ln\varepsilon|}\right) O(\min\{n, m\ln(n)\}) + O(\sqrt{n}\ln n) + \log_2(|\ln\varepsilon|). \end{aligned}$$

Finally, by using the fact that $0 < 1/(\ln 2|\ln \varepsilon|) < 1.5$ for $\varepsilon \leqslant \frac{1}{3}$ we obtain the main result of our paper.

Theorem 4.1. *The expected number of iterations of the (P–C) algorithm required to obtain an ε-optimal solution for random problems RLP and RLD is bounded above by*

$$O(\min\{n^{1.5}, m\sqrt{n}\ln(n)\}) + \log_2(|\ln \varepsilon|).$$

We remark that the quantity $\log_2(|\ln \varepsilon|)$ is very small. For example, with a common stopping tolerance $\varepsilon = 10^{-8}$, we have $\log_2(|\ln \varepsilon|) \approx 4.2$. Therefore, the expected number of iterations is dominated by $O(\min\{n^{1.5}, m\sqrt{n}\ln(n)\})$. For any small $\varepsilon > 0$, our result shows that the complexity bound in the preceding theorem is $O(m\sqrt{n}\ln(n)) + \log_2(|\ln \varepsilon|)$ whenever $m = O(n/\ln(n))$. In establishing $E[\ln \xi] \geqslant -2\ln(n) - 2n - 1$ [10, Theorem 3.1], Huang employed the fact that $C_n^m \leqslant 2^n$. A tighter bound for C_n^m, i.e., $C_n^m \leqslant \min\{2^n, n^m\}$, is used in our analysis. With the help of this tighter bound for C_n^m, it is easily seen that $E[\ln \xi] \geqslant -2\ln(n) - 2\min\{n, m\ln(n)\} - 1$. Therefore, Huang's main result in [10] can also be further enhanced.

Acknowledgment

We are very grateful to the referees for their careful reading of the paper and many comments which improved the presentation of our results.

Appendix

For the sake of completeness of our paper we prove a result of Huang [10] regarding the lower bound for the expected value of $\ln \xi$.

For Todd's non-degenerate random LP model considered in Section 2, any index set with m elements has the same chance to be the optimal index set of the optimal basis since the variable x_i and s_i, $i = 1, 2, \ldots, n$, are symmetric. The fact that a unique basic optimal solution exists with probability one, together with the fact that there are C_n^m basic bases, implies that the probability of the set $\{1, 2, \ldots, m\}$ being chosen as a candidate of optimal index set β is one out of C_n^m, i.e.,

$$P(\beta = \{1, 2, \ldots, m\}) = \frac{1}{C_n^m}.$$

Let $[B|N]$ be the partition of A where $B \in R^{m\times m}$ and $N \in R^{m\times(n-m)}$. Let $\left(x_B^T, 0\right)^T$ be a basic (not necessarily feasible) solution of $Ax = Be + Ne$, where e is a vector of ones of proper dimension. First we know that B is non-singular with probability one. Thus x_B can be expressed as

$$x_B = B^{-1}(Be + Ne) = e - \sqrt{n-m}B^{-1}t,$$

where $t = -Ne/\sqrt{n-m}$ and t is independent of B since N is independent of B. It is easily seen from Theorem 1 of [9, p. 168] that $t \sim N(0, 1)$. Let $\lambda_i \sim N(0, 1)$, $i = 0, 1, 2, \ldots, m$, be mutually independent. Corollary 1 of [7] implies that each element of $B^{-1}t$ is distributed like λ_i/λ_0 (Cauchy distribution). Therefore, x_i is distributed like $1 - \sqrt{n-m}\lambda_i/\lambda_0$ for each $i \in B$.

Let $\left(0, s_N^T\right)^T$ be the solution of $Fs = Fe$. Similarly, we can show that s_j is distributed like $1 - \sqrt{m}\eta_j/\eta_0$ for each $j \in N$, where $\eta_0, \eta_j \sim N(0, 1)$, $j \in N$ and all $\eta_0, \eta_j \in N$ are mutually independent.

Since the random variable $Y_1 = \eta_j/\eta_0$ has a Cauchy distribution with p.d.f. $f_{Y_1}(y_1) = 1/(\pi(1+y_1^2))$, $-\infty < y_1 < \infty$, (see [9, p. 142]), we have the distribution function of the random variable $s_j = 1 - \sqrt{m}\eta_j/\eta_0$,

$$\begin{aligned} F(u) = P(s_j \leqslant u) &= P\left(\eta_i/\eta_0 \geqslant (1-u)/\sqrt{m}\right) \\ &= \int_{\frac{1-u}{\sqrt{m}}}^{\infty} \frac{1}{\pi(1+y_1^2)}\, dy_1 = \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{1-u}{\sqrt{m}}\right). \end{aligned}$$

Thus, the p.d.f. of s_j is

$$f_{s_j}(u) = F'(u) = \frac{1}{\pi\sqrt{m}(1+m^{-1}(1-u)^2)}.$$

For simplicity, we still use $F(u)$ and $f_{s_j}(u)$ for its distribution and p.d.f. of the s_j under $s_j \geqslant 0$. Obviously, $F(u) = P(s_j \leqslant u | s_j \geqslant 0) = 0$ if $u < 0$ and for $u \geqslant 0$, we have

$$F(u) = P(s_j \leqslant u | s_j \geqslant 0) = \frac{1}{P(s_j \geqslant 0)} \int_0^u \frac{1}{\pi\sqrt{m}(1+m^{-1}(1-u)^2)}\, du.$$

Also, it is easily seen that

$$P(s_j \geqslant 0) = \int_0^{\infty} \frac{1}{\pi\sqrt{m}(1+m^{-1}(1-u)^2)}\, du = \frac{1}{2} + \frac{\arctan(1/\sqrt{m})}{\pi} \geqslant \frac{1}{2}.$$

Therefore, the conditional p.d.f. $f_{s_j}(u)$ of s_j under $s_j \geqslant 0$ satisfies

$$f_{s_j}(u) = F'(u) \leqslant \frac{2}{\pi\sqrt{m}(1+m^{-1}(1-u)^2)} \quad \text{for } u > 0. \tag{32}$$

In the same manner, we can prove that the conditional p.d.f. $f_{x_i}(u)$ of x_i under $x_i \geqslant 0$ satisfies

$$f_{x_i}(u) \leqslant \frac{2}{\pi\sqrt{n-m}(1+(n-m)^{-1}(1-u)^2)} \quad \text{for } u > 0. \tag{33}$$

Finally, in order for the set $\{1, 2, \ldots, m\}$ to become the optimal index set, we need to impose the non-negativity of the basic solutions, i.e., the basic solutions $\left(x_B^T, 0\right)^T$ and $\left(0, s_N^T\right)^T$ are optimal to (RLP) and (RLD), if and only if $x_i \geqslant 0$, $i = 1, 2, \ldots, m$ and $s_j \geqslant 0$, $j = m+1, m+2, \ldots, n$. Hence, for the random model, under the condition that $\beta = \{1, 2, \ldots, m\}$, $\xi = \min\{\xi_p, \xi_d\} = \min\{x_1, \ldots, x_m, s_{m+1}, \ldots, s_n\}$, where each s_j and x_i are random variables with their probability density functions satisfying (32) and (33), respectively. It is easily seen from Proposition A.1 of [3] that the p.d.f. $f_\xi(u)$ of ξ satisfies

$$\begin{aligned} f_\xi(u) &\leqslant \frac{1}{P(\beta = \{1, 2, \ldots, m\})} \left\{ \sum_{i=1}^{m} f_{x_i}(u) + \sum_{j=m+1}^{n} f_{s_j}(u) \right\} \\ &\leqslant C_n^m \left\{ \frac{2m}{\pi\sqrt{n-m}(1+(n-m)^{-1}(u-1)^2)} + \frac{2(n-m)}{\pi\sqrt{m}(1+m^{-1}(u-1)^2)} \right\} \\ &\leqslant nC_n^m. \end{aligned}$$

Choosing $c = nC_n^m$, we have $f_\xi(u) \leqslant c$ and therefore,

$$\begin{aligned} E(\ln \xi) &= \int_0^\infty \ln u \; f_\xi(u)\, du = \int_0^{\frac{1}{c}} \ln u \; f_\xi(u)\, du + \int_{\frac{1}{c}}^\infty \ln u \; f_\xi(u)\, du \\ &\geqslant \int_0^{\frac{1}{c}} \ln u \; f_\xi(u)\, du - \ln c \int_{\frac{1}{c}}^\infty f_\xi(u)\, du \\ &\geqslant \int_0^{\frac{1}{c}} \ln u \; f_\xi(u)\, du - \ln c. \end{aligned}$$

Moreover,

$$\left| \int_0^{\frac{1}{c}} \ln u \; f_\xi(u)\, du \right| \leqslant \int_0^{\frac{1}{c}} |\ln u| \; f_\xi(u)\, du \leqslant c \int_0^{\frac{1}{c}} |\ln u|\, du = 1 + \ln c.$$

Thus,

$$E(\ln \xi) \geqslant -2 \ln c - 1 = -2 \ln(nC_n^m) - 1 \geqslant -2 \ln n - 2 \ln(C_n^m) - 1.$$

We have thus proved the following result.

Proposition A.1 (*Huang [10, Theorem 3.1]*). *For Todd's non-degenerate model* 1 *with* $\hat{x} = \hat{s} = e$ [17], *we have*

$$E(\ln \check{\xi}) \geqslant -2 \ln n - 2 \ln(C_n^m) - 1.$$

References

[1] I. Adler, R. Monteiro, Limiting behavior of the affine scaling continuous trajectories for linear programming problems, Contemporary Mathematics, vol. 114, American Mathematical Society, Providence, Rhode Island, USA, 1990, pp. 189–211.

[2] K.M. Anstreicher, J. Ji, F.A. Potra, Y. Ye, Average performance of a self-dual interior-point algorithm for linear programming, in: P. Pardalos (Ed.), Complexity in Numerical Optimization, World Scientific, New Jersey, 1993, pp. 1–15.

[3] K.M. Anstreicher, J. Ji, F.A. Potra, Y. Ye, Probabilistic analysis of an infeasible-interior-point algorithm for linear programming, Math. Oper. Res. 24 (1999) 176–192.

[4] K.M. Anstreicher, J. Ji, Y. Ye, Average performance of an ellipsoid termination criterion for linear programming interior-point algorithms, Working paper 92-1, Department of Management Sciences, University of Iowa, Iowa City, IA, 1992.

[5] K.H. Borgwardt, The Simplex Method-A Probabilistic Approach, Spring, Berlin, Heidelberg, New York, 1987.

[6] K.H. Borgwardt, A sharp upper bound for the expected number of shadow-vertices in LP-polyhedra under orthogonal projection on two-dimensional planes, Math. Oper. Res. 24 (1999) 925–984.

[7] V.L. Girko, On the distribution of solutions of systems of linear equations with random coefficients, Theor. Probab. Math. Stat. 2 (1974) 41–44.

[8] O. Güler, Y. Ye, Convergence behavior of some interior-point algorithms, Math. Programming 60 (1993) 215–228.

[9] R.V. Hogg, A.T. Craig, Introduction to Mathematical Statistics, Macmillan Publishing Co., Inc., New York, 1978.

[10] S. Huang, Average number of iterations of interior-point algorithms for linear programming, Sci. China (A) 43 (2000) 829–835.

[11] S. Huang, Y. Ye, On the average number of iterations of the polynomial interior-point algorithms for linear programming, Working paper series No. 91-16, Department of Management Sciences, The University of Iowa, Iowa City, Iowa 52242, 1991.

[12] P. Huhn, K.H. Borgwardt, An upper bound for the average number of iterations required in phase II of an interior-point-method, in: Operations Research Proceedings 1997, Springer, Berlin, GOR-Conference, Jena, 1998, pp. 19–24.

[13] P. Huhn, K.H. Borgwardt, Interior-point methods: worst case and average case analysis of a phase-I algorithm and a termination procedure, J. Complexity 18 (2002) 833–910.
[14] J. Ji, F.A. Potra, On the average complexity of finding an ε-optimal solution for linear programming, Reports on Computational Mathematics No. 25, Department of Mathematics, University of Iowa, Iowa City, IA, 1992.
[15] S. Mizuno, M. Todd, Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming, Math. Oper. Res. 18 (1993) 964–981.
[16] S. Smale, On the average number of steps of the simplex method of linear programming, Math. Programming 27 (1983) 241–262.
[17] M.J. Todd, Probabilistic models for linear programming, Math. Oper. Res. 16 (1991) 671–693.
[18] M.J. Todd, L. Tuncel, Y. Ye, Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems, Math. Programming 90 (2001) 59–69.
[19] M.J. Todd, Y. Ye, A centered projective algorithm for linear programming, Math. Oper. Res. 15 (1990) 508–529.
[20] J.F. Traub, H. Wozniakowski, Complexity of linear programming, Oper. Res. Lett. 1 (1982) 59–62.
[21] S.A. Vavasis, Y. Ye, A primal-dual interior-point method whose running time depends only on the constraint matrix, Math. Programming 74 (1996) 79–120.
[22] Y. Ye, Towards probabilistic analysis of interior-point algorithms for linear programming, Math. Oper. Res. 19 (1994) 38–52.
[23] Y. Ye, O. Güler, R. Tapia, Y. Zhang, A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming, Math. Programming 59 (1993) 151–162.