Probabilistic analysis of two complexity measures for linear programming problems *

M. J. Todd † Levent Tunçel‡ Yinyu Ye§

September 28, 1998

Abstract

This note provides a probabilistic analysis of $\bar{\chi}_A$, a condition number used in the linear programming algorithm of Vavasis and Ye [14] whose running time depends only on the constraint matrix $A \in \mathbf{R}^{m \times n}$. We show that if A is a standard Gaussian matrix, then $E(\ln \bar{\chi}_A) = O(\min\{m \ln n, n\})$. Thus, the expected running time of linear programming is bounded by a polynomial in m and n for any right-hand side and objective coefficient vectors when A is randomly generated in this way. We show that the same bound holds for $E(\ln \sigma(A))$, where $\sigma(A)$ is another condition number of A arising in complexity analyses of linear programming problems.

*This work was done while the authors were visiting the Mathematical Sciences Research Institute, UC Berkeley, USA.

†School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu. This author was supported in part by NSF through grant DMS-9805602 and ONR through grant N00014-96-1-0050, and also by the Mathematical Sciences Research Institute, UC Berkeley, USA.

‡Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. E-mail: ltuncel@math.uwaterloo.ca. This author's research was supported in part by Tokyo Institute of Technology and research grants from NSERC of Canada.

§Department of Management Sciences, The University of Iowa, Iowa City, IA 52242. E-mail: yinyu-ye@uiowa.edu. This author was supported in part by NSF Grants DMI-9522507 and DMS-9703490, and also by the Mathematical Sciences Research Institute, UC Berkeley, USA.

Keywords: Linear Programming, Interior-Point Methods, Probabilistic Analysis.

1 Introduction

Consider the linear programming (LP) problems in primal-dual form:

$$\begin{array}{rl} \text{Primal: minimize} & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0, \end{array} \tag{1}$$

and

$$\begin{array}{rl} \text{Dual: maximize} & b^T y \\ \text{subject to} & s = c - A^T y \geq 0, \end{array} \tag{2}$$

where $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ are given vectors, $A \in \mathbb{R}^{m \times n}$ is a given matrix of rank m, and $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are unknown vectors.

Recently, Vavasis and Ye [14] proposed a primal-dual "layered-step" interior point (LIP) algorithm for linear programming with data given by real numbers (also see Megiddo, Mizuno and Tsuchiya [7] for a modification of the LIP algorithm). This algorithm follows the central path, either with short steps or with a new type of step called a "layered least squares" (LLS) step. The algorithm returns an exact optimum after a finite number of steps—in particular, after $O(n^{3.5}c(A))$ Newton iterations, where

$$c(A) = \ln \bar{\chi}_A + \ln n$$

and $\bar{\chi}_A$ is a function of the coefficient matrix A. (We state the definition of $\bar{\chi}_A$ and derive some of its properties in Section 2.) This is in contrast to all previous interior methods, whose complexity depends on b and c as well as A. This is important because there are many classes of problems in which A is "well-behaved" but b and c are arbitrary vectors.

In this short note, we prove that, if A is a standard Gaussian matrix, i.e., each component of A is independently and identically distributed (i.i.d.) from the $N(0,1)$ distribution, then

$$E(\ln \bar{\chi}_A) = O(\min\{m \ln n, n\}).$$

Thus, the expected running time of linear programming is bounded by a polynomial in m and n for any right-hand side b and objective coefficient vector c when A is randomly generated in this way. We provide a similar bound on $E(|\ln \sigma(A)|)$, where $\sigma(A)$ is another function of A relevant to the complexity of linear programming algorithms (see Section 3).

2 What is $\bar{\chi}_A$?

Let us define $\mathcal{D}$ to be the set of all positive definite $n \times n$ diagonal matrices. Then, $\bar{\chi}_A$ is defined by

$$\bar{\chi}_A = \sup\{\|A^T(ADA^T)^{-1}AD\| : D \in \mathcal{D}\}.$$

(We use the 2-norm and the induced operator norm.) This quantity is always finite. It appears that Dikin [3] first proved this result, using the Cauchy-Binet formula. Later independent proofs were given by Ben-Tal and Teboulle [2], Stewart [9], and Todd [10]. Forsgren [4] describes the history and gives some extensions of this result.

Since $(ADA^T)^{-1}ADc$ minimizes $\|D^{1/2}(A^Ty - c)\|$, an equivalent way to define these parameters is in terms of weighted least squares:

$$\bar{\chi}_A = \sup\left\{\frac{\|A^Ty\|}{\|c\|} : y \text{ minimizes } \|D^{1/2}(A^Ty - c)\| \text{ for some } c \in \mathbb{R}^n, D \in \mathcal{D}\right\}. \tag{3}$$

Observe that this parameter is invariant when A is scaled by a constant. More generally, it is invariant if A is replaced by RA for any $m \times m$ nonsingular matrix R. This means that $\bar{\chi}_A$ actually depends on the nullspace of A rather than on A itself.

We first prove the following simple result. This can be obtained from Dikin's result algebraically, but we sketch a simple geometric proof based on the arguments of Todd [10].

Lemma 1 *Let $\mathcal{B}$ denote the set of all bases (nonsingular $m \times m$ submatrices) of A. Then,*

$$\bar{\chi}_A \leq \max\{\|A^TB^{-T}\| : \; B \in \mathcal{B}\}.$$

Proof. We use the characterization (3) and analyze those vectors y that are candidates for the solution of the least-squares problems there. Fix c and consider the polyhedral subdivision of $\mathbb{R}^m$ induced by the hyperplanes $a_j^T y = c_j$, $j = 1, \ldots, n$. A least-squares solution y must lie in one of these polyhedra. If it lies in a bounded polyhedron, it can be written as a convex combination of the vertices of the subdivision. If not, it can be expressed as the sum of a convex combination of vertices of the subdivision and a nonnegative combination of extreme rays of the subdivision. But removing the second summand, if present, moves y closer to at least one of the hyperplanes and no further from any of them, and so decreases the least-squares objective no matter what D is. Thus all least-squares solutions lie in the convex hull of the vertices of the subdivision. Each such vertex has the form $B^{-T} c_B$ for some basis, where c_B is the corresponding subvector of c, and thus

$$\|A^T y\|/\|c\| \leq \max\{\|A^T B^{-T} c_B\|/\|c\|\} \leq \max\{\|A^T B^{-T}\|\}.$$

∎

The following lemma is from Vavasis and Ye [14]:

Lemma 2 *Let B be a basis of A. Then, $\|A^T B^{-T}\| \leq \bar{\chi}_A$.*

PROOF. Let D_ϵ be the matrix with ones in positions corresponding to B and $0 < \epsilon << 1$ (very small) in the remaining diagonal positions. In the limit as $\epsilon \to 0^+$, we have that $AD_\epsilon A^T$ tends to BB^T. Since B is invertible, $(AD_\epsilon A^T)^{-1}$ tends to $(BB^T)^{-1}$. Also, AD_ϵ tends to a matrix with a copy of B, and the remaining columns filled in by zeros. The resulting product in the definition in the limit is precisely $[A^T B^{-T}, 0]$. ∎

Combining the two lemmas, we have:

Theorem 1 *Let $\mathcal{B}$ denote the set of all bases of A. Then*

$$\bar{\chi}_A = \max\{\|B^{-1} A\| : \; B \in \mathcal{B}\}.$$

Let B be a basis of A, and, without loss of generality let $A = [B, \; N] = [B, \; a_{B1}, \; a_{B2}, \; ..., \; a_{B(n-m)}]$. Then,

$$B^{-1} A A^T B^{-T} = I + B^{-1} N N^T B^{-T},$$

where I is the identity matrix in $\mathbb{R}^{m\times m}$. Thus,

$$\begin{aligned}
\|B^{-1}A\|^2 &\le m + \mathrm{trace}(B^{-1}NN^TB^{-T}) \\
&= m + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|^2 \\
&\le \left(m^{1/2} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|\right)^2 .
\end{aligned}$$

Therefore, we have

Corollary 1 *Let $\mathcal{B}$ denote the set of all bases of A. Then,*

$$\bar{\chi}_A \le \sum_{B\in\mathcal{B}} \left(m^{1/2} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|\right).$$

We will use this corollary to bound the expected value of $\ln \bar{\chi}_A$ in Section 4.

First, we relate $\bar{\chi}_A$ to another measure of A, so that our bound gives information about the latter also.

3 Another Condition Measure of A

We now describe another function of A (due to Ye, see for instance [15]) also arising from computational complexity considerations in LP algorithms. Let S denote the range of A^T, so that $S^\perp$ is the nullspace of A. Consider the following pair of feasibility problems (see [13]):

$$(P)\ \left\{\begin{array}{rcl} x &\in& S^\perp, \\ \|x\|_1 &=& 1, \\ x &\ge& 0. \end{array}\right.$$

$$(D)\ \left\{\begin{array}{rcl} s &\in& S, \\ \|s\|_1 &=& 1, \\ s &\ge& 0. \end{array}\right.$$

If we assume that both (P) and (D) have feasible solutions, then there exist a proper partition $[B, N]$ of $\{1, 2, \ldots, n\}$ and a feasible pair (x, s) such that $x \in S^\perp$, $x_N = 0$, $x_B > 0$, $s \in S$, $s_N > 0$, $s_B = 0$. Note that (P) and (D)

cannot both be infeasible. If (P) is infeasible then $N = \{1, 2, \ldots, n\}$ and if (D) is infeasible then $B = \{1, 2, \ldots, n\}$.

Ye's measure for this feasibility pair is defined as follows:

$$\begin{aligned}
\sigma(S^\perp) &:= \min_{j \in B} \max_{x \in S^\perp : \|x\|_1 = 1, x \geq 0} \{x_j\}, \\
\sigma(S) &:= \min_{j \in N} \max_{s \in S : \|s\|_1 = 1, s \geq 0} \{s_j\}, \\
\sigma(A) &:= \min\left\{\sigma(S), \sigma(S^\perp)\right\},
\end{aligned}$$

where we define $\sigma(S)$ (or $\sigma(S^\perp)$) to be 1 if the corresponding problem is infeasible. A pair of feasible solutions or a strictly feasible solution to one (and hence a witness that the other is infeasible) can be found in $O(\sqrt{n}(|\ln \sigma(A)| + \ln n))$ iterations (see [15, 13]), so we are again interested in bounding the expected value of $|\ln \sigma(A)|$.

Let us define for $v \in \mathbb{R}^n$, $J \subseteq \{1, 2, \ldots, n\}$,

$$v_J^+ := \begin{cases} -\infty & \text{if } v_J \leq 0 \\ \max_{i \in J}\{v_i\} & \text{otherwise,} \end{cases}$$

and $v_J^- := -(-v)_J^+$. Next we state a dual characterization of $\sigma(S)$ due to Tunçel [12].

Proposition 1 *Let [B,N] be the strictly complementary partition defined by A. Suppose $S^\perp$ contains no unit vector and $B \neq \emptyset$. Then,*

$$\sigma(S^\perp) := \min_{i \in B} \max_{\substack{x \in S^\perp \\ \|x\|_1 = 1 \\ x \geq 0}} \{x_i\} = \min_{\substack{\gamma \in S \\ 0 < \gamma_B^+ < 1 \\ \gamma_B^+ - \gamma_B^- = 1}} \gamma_B^+.$$

The statement above is slightly different than the one in [12] (there, $[B, N]$ is assumed to be a *proper* partition; however, its proof only uses $B \neq \emptyset$).

In this section, we relate $\bar{\chi}_A$ to the minimum of the σ measure of various matrices closely related to A. It is convenient to work with the inverse of $\bar{\chi}_A$, so we define

$$\rho(S) := \frac{1}{\bar{\chi}_A}.$$

(Recall that $\bar{\chi}_A$ depends only on the nullspace $S^\perp$ of A, so any choice for A yields the same result.) We use the following characterization of $\rho(S)$, due to O'Leary[8]:

Proposition 2 *Let J denote a nonempty subset of the index set $\{1, 2, \ldots, n\}$ and $\bar{J}$ be its complement. Then considering J, γ and ξ as the variables, we have*

$$\begin{array}{rcl} \rho(S) & = & \inf \quad \|\gamma_J\| \\ & & \begin{array}{rcl} \mathrm{sign}(\gamma_i) & = & \mathrm{sign}(\xi_i), \quad i \in \bar{J}, \\ \|\gamma\| = 1, & & \\ \gamma \in S, & & \xi \in S^\perp. \end{array} \end{array}$$

It is not hard to see that the above infimum is attained; after taking limits down a subsequence of a minimizing sequence, we can put into J any components that are zero in γ, and replace the limiting ξ if necessary (if it has a zero component in $\bar{J}$ corresponding to a nonzero component of γ) by a suitable member of the subsequence. Alternatively, in the argument below we can take a near-optimal solution and then take limits.

Let $K \subseteq \{1, 2, \ldots, n\}$. If we multiply by (-1) the columns of A indexed by K, then the resulting matrix (we call it the *signing* of A by K and denote it A_{-K}) has the same $\bar{\chi}$ value as A (this is apparent from Proposition 2). For a subspace U we similarly denote by U_{-K} the subspace consisting of points of U with their components indexed by K multiplied by (-1); note that $(S^\perp)_{-K} = (S_{-K})^\perp$, so we can write $S^\perp_{-K}$ unambiguously. We let $\sigma_K(A)$ denote the σ value of A_{-K}, and similarly $\sigma_K(S)$ and $\sigma_K(S^\perp)$. We write $\underline{\sigma}(A)$ for $\min_{K \subseteq \{1,2,\ldots,n\}} \sigma_K(A)$. We have the following fact. For similar results using different norms, see [12].

Lemma 3

$$\rho(S) \;\leq\; \underline{\sigma}(A).$$

PROOF. Vavasis and Ye [13] proved $\rho(S) \leq \sigma(S)$, $\rho(S^\perp) \leq \sigma(S^\perp)$ (and $\rho(S) \leq \frac{\sigma(A)}{1-\sigma(A)}$). Gonzaga and Lara [5] showed $\rho(S) = \rho(S^\perp)$. Therefore, we have

$$\rho(S) \;\leq\; \sigma(A). \tag{4}$$

Since we established that every signing of A gives the same $\bar{\chi}_A$ value, applying inequality (4) to each signing K concludes the proof. ∎

Theorem 2

$$\frac{1}{\sqrt{n}}\underline{\sigma}(A) \ \le\ \rho(S) \ \le\ \underline{\sigma}(A).$$

Proof. The right-hand inequality is Lemma 3. It remains to prove the left-hand inequality.

First we treat some special cases. If the unit vector e_j lies in S, and S is not spanned by e_j, then it is easy to check that

$$\rho(S\backslash\{j\}) = \rho(S) \text{ and } \underline{\sigma}(A\backslash\{j\}) = \underline{\sigma}(A),$$

where $S\backslash\{j\}$ denotes the subspace of $\mathbb{R}^{n-1}$ obtained by eliminating the jth component of every member of S and $A\backslash\{j\}$ denotes the matrix obtained from A by deleting its jth column. If S is spanned by a unit vector, then it is easy to see that both $\rho(S)$ and $\underline{\sigma}(A)$ are one, and the inequality holds. It also holds whenever $\rho(S)$ is one. So from now on we assume that $\rho(S) < 1$ and that neither S nor $S^\perp$ contains a unit vector.

We consider the optimization problem in Proposition 2 and let γ^*, ξ^*, and J give an optimal solution. Since $\|\gamma^*_J\| < 1 = \|\gamma^*\|$, we have γ^*_J and $\gamma^*_{\bar{J}}$ nonzero. (If γ^*_J were zero, γ^* could not be orthogonal to ξ^*.) We can also assume that J contains all indices of zero components of γ^*, so that γ^* and hence ξ^* are nonzero on $\bar{J}$. We use this solution to produce a feasible solution to the dual characterization in Proposition 1 of $\sigma_K(S^\perp)$ (for some signing of A) with a close objective value.

Let γ^*_j be the largest (in absolute value) of the components indexed by J, and let γ^*_k be the largest (in absolute value) of the components indexed by $\bar{J}$. Then at least one of $|\gamma^*_j|$ and $|\gamma^*_k|$ is at least $1/\sqrt{n}$. Now scale γ^* to get γ with $|\gamma_j| + |\gamma_k| = 1$; the scale factor is no more than $\sqrt{n}$, so we have $|\gamma_j| = \|\gamma_J\|_\infty \le \|\gamma_J\|$ at most $\sqrt{n}\rho(S)$.

By assumption, e_j does not lie in S, so there is some $\hat{\xi}$ in $S^\perp$ with $\hat{\xi}_j \ne 0$. If we had ξ^*_j zero, we could add a small multiple of $\hat{\xi}$ to ξ^* so that its jth component would have the same sign as γ_j and then move j from J to $\bar{J}$, contradicting the optimality of γ^*. In fact, let $\hat{J} \subseteq J$ be the set of indices where γ_J is nonzero. The same argument shows that ξ^* is nonzero on $\hat{J}$. So ξ^* is nonzero on $\hat{J}$, where it has the opposite sign to γ (or we could again move components to $\bar{J}$, decreasing the objective function), and on $\bar{J}$, where it has the same sign as γ.

Now we choose a signing A_{-K} of A so that ξ^* is nonpositive. It is then negative on $\hat{J}$ and $\bar{J}$, so γ is positive on $\hat{J}$ and negative on $\bar{J}$. Moreover, $-\xi^*$ is nonnegative, so for this signing, B contains $\hat{J}$ and $\bar{J}$. Now since γ is negative on $\bar{J}$ and $j \in B$, we have $\gamma_B^+ = \gamma_j$. Similarly, $\gamma_B^- = \gamma_k$. Hence γ is feasible in the dual characterization of $\sigma_K(S^\perp)$ with objective function value at most $\sqrt{n}\rho(S)$ and the proof is complete. ∎

Corollary 2 *For any distribution on A, we have*

$$E(|\ln \underline{\sigma}(A)|) \leq E(\ln \bar{\chi}_A) \leq E(|\ln \underline{\sigma}(A)|) + \frac{1}{2}\ln n.$$

4 Analysis of Random A

In the preceding two sections, we have presented a characterization of $\bar{\chi}_A$ in terms of bases of A and we have related $\bar{\chi}_A$ to $\underline{\sigma}(A)$. Some probabilistic analyses of these related entities have been carried out for A with a Gaussian distribution (see, for examples, Anstreicher et al. [1], Girko [6], Todd [11], and Ye [15]), and they can both be used in our probabilistic analysis presented in this section. For simplicity, we shall use the characterization of $\bar{\chi}_A$ to prove our next main theorem.

We have the following lemma due to Girko [6] (see also Todd [11]).

Lemma 4 *Let $B \in \mathbb{R}^{m\times m}$ be a standard Gaussian matrix. Furthermore, let $a \in \mathbb{R}^m$ be a standard Gaussian vector, independent of B. Then, $B^{-1}a$ is a random vector*

$$B^{-1}a = \frac{1}{\eta_0}\begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_m \end{pmatrix},$$

where the η_i, $i = 0, 1, ..., m$, are i.i.d. from $N(0,1)$.

The next lemma is similar to results of Anstreicher et al. [1].

Lemma 5 *Let $B \in \mathbb{R}^{m\times m}$ be a standard Gaussian matrix. Furthermore, let $a \in \mathbb{R}^m$ be a standard Gaussian vector independent of B. Then,*

$$E(\|B^{-1}a\|^{1/2}) \leq m^{1/4} \cdot \left(1 + \frac{2}{\sqrt{2\pi}}\right).$$

Proof. From the preceding lemma

$$\begin{aligned}
E\left(\|B^{-1}a\|^{1/2}\right) &= E\left(\frac{(\sum_{i=1}^m \eta_i^2)^{1/4}}{|\eta_0|^{1/2}}\right) \\
&= E\left(\left(\sum_{i=1}^m \eta_i^2\right)^{1/4}\right) \cdot E\left(|\eta_0|^{-1/2}\right) \\
&\le \left(E\left(\sum_{i=1}^m \eta_i^2\right)\right)^{1/4} \cdot E\left(|\eta_0|^{-1/2}\right) \\
&= m^{1/4} E\left(|\eta_0|^{-1/2}\right),
\end{aligned}$$

since the η_i, $i = 0, 1, ..., m$, are i.i.d. from $N(0,1)$. Now consider $E\left(|\eta_0|^{-1/2}\right)$. Let $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$, the density function of $N(0,1)$. Then

$$\begin{aligned}
E\left(|\eta_0|^{-1/2}\right) &= 2\int_0^\infty x^{-1/2} f(x)dx \\
&= 2\left(\int_1^\infty x^{-1/2} f(x)dx + \int_0^1 x^{-1/2} f(x)dx\right) \\
&\le 2\left(\int_1^\infty f(x)dx + \int_0^1 x^{-1/2} f(x)dx\right) \\
&= 2\left(\int_0^\infty f(x)dx + \int_0^1 (x^{-1/2} - 1) f(x)dx\right) \\
&= 1 + 2\int_0^1 (x^{-1/2} - 1) f(x)dx \\
&\le 1 + 2\int_0^1 (x^{-1/2} - 1) \frac{1}{\sqrt{2\pi}} dx \\
&= 1 + \frac{2}{\sqrt{2\pi}}.
\end{aligned}$$

Thus, we have the desired result. ∎

Finally,

Theorem 3 *Let $A \in \mathbb{R}^{m \times n}$ be a standard Gaussian matrix with $m \ge 3$. Then, with probability one A has full row rank, and*

$$E(\ln \bar{\chi}_A) \le (2m + 2) \ln n.$$

PROOF. From Corollary 1 and the last lemma,

$$
\begin{aligned}
& E(\ln \bar{\chi}_A) \\
\leq\ & E\left(\ln \sum_{B\in\mathcal{B}} \left(m^{1/2} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|\right)\right) \\
\leq\ & E\left(\ln \left(\sum_{B\in\mathcal{B}} \left(m^{1/4} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|^{1/2}\right)\right)^2\right) \\
=\ & E\left(2\ln \sum_{B\in\mathcal{B}} \left(m^{1/4} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|^{1/2}\right)\right) \\
=\ & 2E\left(\ln \sum_{B\in\mathcal{B}} \left(m^{1/4} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|^{1/2}\right)\right) \\
\leq\ & 2\ln E\left(\sum_{B\in\mathcal{B}} \left(m^{1/4} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|^{1/2}\right)\right) \\
=\ & 2\ln\left(\sum_{B\in\mathcal{B}} \left(m^{1/4} + \sum_{j=1}^{n-m} E\left(\|B^{-1}a_{Bj}\|^{1/2}\right)\right)\right) \\
\leq\ & 2\ln\left(\sum_{B\in\mathcal{B}} \left(m^{1/4} + \sum_{j=1}^{n-m} m^{1/4}\left(1+\frac{2}{\sqrt{2\pi}}\right)\right)\right) \\
\leq\ & 2\ln\left(\sum_{B\in\mathcal{B}} \left((n-m+1)m^{1/4}\left(1+\frac{2}{\sqrt{2\pi}}\right)\right)\right) \\
\leq\ & 2\ln\left(\begin{pmatrix} n \\ m \end{pmatrix}\left((n-m+1)m^{1/4}\left(1+\frac{2}{\sqrt{2\pi}}\right)\right)\right) \\
=\ & 2\ln\begin{pmatrix} n \\ m \end{pmatrix} + 2\ln\left((n-m+1)m^{1/4}\left(1+\frac{2}{\sqrt{2\pi}}\right)\right) \\
\leq\ & 2(m\ln n - \ln m) + 2\ln(n-m+1) + \ln m/2 + 2\ln 2 \\
\leq\ & (2m+2)\ln n.
\end{aligned}
$$

∎

The expectation bound can be further sharpened if we consider the power

$1/(1+\epsilon)$ for any fixed $\epsilon > 0$, rather than $(1/2)$, in Lemma 5. We can show

$$E(\|B^{-1}a\|^{1/(1+\epsilon)}) \le m^{1/2(1+\epsilon)} \cdot \left(1 + \frac{2}{\epsilon\sqrt{2\pi}}\right).$$

Then, we can modify the proof in Theorem 3 to

$$\begin{aligned}
&E(\ln \bar{\chi}_A) \\
&\quad\le\ E\left(\ln \sum_{B\in\mathcal{B}} \left(m^{1/2} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|\right)\right) \\
&\quad\le\ E\left(\ln \left(\sum_{B\in\mathcal{B}} \left(m^{1/2(1+\epsilon)} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|^{1/(1+\epsilon)}\right)\right)^{1+\epsilon}\right) \\
&\quad=\ (1+\epsilon)E\left(\ln \sum_{B\in\mathcal{B}} \left(m^{1/2(1+\epsilon)} + \sum_{j=1}^{n-m} \|B^{-1}a_{Bj}\|^{1/(1+\epsilon)}\right)\right) \\
&\quad\le\ (1+\epsilon)\ln \left(\sum_{B\in\mathcal{B}} \left(m^{1/2(1+\epsilon)} + \sum_{j=1}^{n-m} m^{1/2(1+\epsilon)}\left(1+\frac{2}{\epsilon\sqrt{2\pi}}\right)\right)\right) \\
&\quad\le\ (1+\epsilon)\ln \left(\binom{n}{m}\left((n-m+1)m^{1/2(1+\epsilon)}\left(1+\frac{2}{\epsilon\sqrt{2\pi}}\right)\right)\right) \\
&\quad=\ (1+\epsilon)\ln \binom{n}{m} + (1+\epsilon)\ln\left((n-m+1)m^{1/2(1+\epsilon)}\left(1+\frac{2}{\epsilon\sqrt{2\pi}}\right)\right).
\end{aligned}$$

Suppose $n > m \ge 3$. If we choose $\epsilon = 1/m$, then the bound is

$$E(\ln \bar{\chi}_A) \le (m+3)\ln n$$

(we use $\ln(1+2/(\sqrt{2\pi}m)) \le \ln m + \ln(2/\sqrt{2\pi}+1/3) \approx \ln m + \ln 1.1312$); if we let $\epsilon = 1/n$ and note that

$$\binom{n}{m} \le 2^n,$$

then

$$E(\ln \bar{\chi}_A) \le n\ln 2 + 3\ln n + 1$$

(using $\ln(1+2/(\sqrt{2\pi}n)) \le \ln n + \ln(2/\sqrt{2\pi}+1/4)$ and $\ln 2 + (5/4)\ln(2/\sqrt{2\pi}+1/4) < 1$).

To summarize, we have

Corollary 3 *Let $A \in \mathbb{R}^{m\times n}$ be a standard Gaussian matrix with $n > m \geq 3$. Then, with probability one A has full row rank, and*

$$E(\ln \bar{\chi}_A) \leq \min\{(m+3)\ln n,\ n\ln 2 + 3\ln n + 1\}.$$

From Corollary 2 we deduce

Corollary 4 *Let $A \in \mathbb{R}^{m\times n}$ be a standard Gaussian matrix with $n > m \geq 3$. Then, with probability one A has full row rank, and*

$$E(|\ln \underline{\sigma}(A)|) \leq \min\{(m+3)\ln n,\ n\ln 2 + 3\ln n + 1\}.$$

We also have a corollary for other distributions of A.

Corollary 5 *Let the linear programming constraint matrix $A \in \mathbb{R}^{m\times n}$ be randomly generated. If for any basis B and any column vector a of A not in B, we have*

$$E\left(\|B^{-1}a\|^{\frac{1}{m^\alpha}}\right) \leq O(m^\beta)$$

for some constants α, $\beta \geq 0$, then the expected running time of the LIP algorithm is bounded by a polynomial in m and n for any right-hand side and objective function vectors.

References

[1] K. M. Anstreicher, J. Ji, F. Potra, and Y. Ye, *Probabilistic analysis of an infeasible primal-dual algorithm for linear programming.* Working paper, Dept. of Management Sciences, University of Iowa (Iowa City, IA, 1992), to appear in Mathematics of Operations Research.

[2] A. Ben-Tal and M. Teboulle, *A geometric property of the least squares solution of linear equations.* Linear Algebra and Its Applications 139 (1990) 165–170.

[3] I. I. Dikin, *On the speed of an iterative process.* Upravlaemye Sistemy 12 (1974) 54–60.

[4] A. Forsgren, *On linear least-squares problems with diagonally dominant weight matrices.* SIAM Journal on Matrix Analysis and Applications 17 (1996) 763–788.

[5] C. C. Gonzaga and H. Lara, *A note on properties of condition numbers.* Linear Algebra and Its Applications 261 (1997) 269–273.

[6] V. L. Girko, *On the distribution of solution of systems of linear equations with random coefficients.* Theory of Probability and Mathematical Statistics 2 (1974) 41–44.

[7] N. Megiddo, S. Mizuno and T. Tsuchiya, *A modified layered-step interior-point algorithm for linear programming.* IBM Technical Report, IBM Research Lab, San Jose, CA, 1996.

[8] D. P. O'Leary, *On bounds for scaled projections and pseudoinverses.* Linear Algebra and Its Applications 132 (1990) 115–117.

[9] G. W. Stewart, *On scaled projections and pseudoinverses.* Linear Algebra and Its Applications 112 (1989) 189–193.

[10] M. J. Todd, *A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm.* Operations Research 38 (1990) 1006–1018.

[11] M. J. Todd, *Probabilistic models for linear programming.* Mathematics of Operations Research 16 (1991) 671–693.

[12] L. Tunçel, *On the condition numbers for polyhedra in Karmarkar's form.* Operations Research Letters, to appear.

[13] S. A. Vavasis and Y. Ye, *Condition numbers for polyhedra with real number data.* Operations Research Letters 17 (1995) 209–214.

[14] S. A. Vavasis and Y. Ye, *A primal-dual interior-point method whose running time depends only on the constraint matrix.* Mathematical Programming 74 (1996) 79–120.

[15] Y. Ye, *Toward probabilistic analysis of interior-point algorithms for linear programming.* Mathematics of Operations Research 19 (1994) 38–52.