SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 836

February 1989
Revised April 1990

PROBABILISTIC MODELS FOR
LINEAR PROGRAMMING

by

Michael J. Todd

This work was partially supported by NSF Grant ECS-8602534 and ONR Contract N00014-87-K-0212.

Probabilistic Models for Linear Programming

Michael J. Todd*

Abstract. We propose and investigate new probabilistic models for linear programming. In contrast to previous models, ours guarantee the existence of optimal solutions and are symmetric under duality. While in some respects our distributions are very special, there is sufficient flexibility to permit an arbitrary degree of primal and/or dual degeneracy, either just at the optimal solution or throughout the feasible region using null variables. Moreover, the precision of the distributions allows us to compute the probability that the feasible region is bounded as well as the distribution of the distance to a constraint hyperplane and that of the components of a vertex. Interest in these measures stems from Karmarkar's algorithm, and we also introduce a model for generating random linear programming problems on a simplex.

*Research supported in part by NSF Grant ECS-8602534 and ONR Contract N00014-87-K-0212.

1. Introduction.

There have been many probabilistic models suggested for generating random polyhedra or random linear programming problems. The concern in geometric probability (see Efron [12], Raynaud [22], Renyi and Sulanke [23] and W. Schmidt [26]) has generally been in the asymptotic number of extreme points and facets of the convex hull of a collection of randomly generated points. By taking polars, these results concern the asymptotic number of facets and vertices of polyhedra defined by a collection of randomly generated inequalities.

In the mathematical programming literature, the first paper appears to have been by Motzkin [20], who computed the probability of feasibility of a system of inequalities each of which could have either sense with equal probability independently. This was the first appearance of the sign-invariant model, which was later studied by Prekopa [21] and then extensively in the eighties by Adler and Berenguer [2,3] and May and Smith [18]. This research culminated in results on the expected length of paths of a parametric algorithm by Haimovich [13] and Adler [1] and of a complete 2-phase algorithm by Adler, Karp and Shamir [4], Adler and Megiddo [5], and Todd [30]. The sign invariant models comes in two flavors: unconditional, as studied by Motzkin, Prekopa, and the last three papers [4,5,30]; and conditional on the nonemptiness of the feasible region or of the efficient path in [1,2,3,13,18]. It is unfortunate that the strongest results, with a quadratic bound on the expected number of steps for a complete 2-phase algorithm, require a distribution that generates infeasible or unbounded problems with probabilities approaching one as the dimension increases. Only in the case where the number of inequalities is twice the dimension can one easily remedy this defect, at the cost of increasing the experiment in the bound to $5/2$.

Another sign-invariant model was that considered in Smale's work [27]. This model was later considerably generalized by Smale [28] and Blair [6].

A very interesting model for random linear programs was proposed by Vershik and Sporyshev [33, 34]. This also satisfies the sign-invariance property, but is based on the Grassmannian approach introduced by Vershik in which a linear programming instance is associated with a subspace of appropriate dimension in an appropriately-dimensioned space, and its intersection with a certain affine cone. Vershik and Sporyshev's distribution is the unique distribution on such subspaces that is invariant under orthogonal transformations, and this is very similar to our approach in section 2. These models, however, all retain the drawback that almost all problems were infeasible or unbounded.

A very different model was proposed and studied by Borgwardt [8,9,10]. Here the origin was always feasible, and invariance under sign changes was replaced by rotational invariance of the normals to the constraints and the objective function gradient. Borgwardt's problems were always feasible and usually (with sufficiently many inequalities) bounded. A major drawback is that the algorithm studied uses the known feasible point (the origin) as a starting point. Borgwardt has since proposed an extension that uses parametric ideas to obtain an initial feasible solution if one exists [11]. B. Schmidt and Mattheiss [25] considers a generalization of Borgwardt's model where the normals to the constraints are only required to be invariant under sign switches rather than arbitrary orthogonal transformations.

While the sign-invariant models (without conditioning) have duals that are similarly distributed, this is not the case for the models of Borgwardt and Schmidt and Mattheis. Moreover, none of these models permits the introduction of primal and dual degeneracy, which is a feature of most real-world problems. In this paper we propose new probabilistic models that avoid these drawbacks. We do not have results on the expected behavior of algorithms applied to problems generated according to our models. However, we have obtained distributions and/or expectations for several measures of their feasible regions.

In choosing the measures to study, we have been influenced by the recent interior–point methods for linear programming arising from Karmarkar’s algorithm [15]. Hence we have been interested in metrical rather than just combinatorial properties of the polyhedra generated, and this has required the choice of specific distributions rather than those with, say, just sign–invariance. Usually an interior feasible point is known (although we do not intend the algorithms we apply to know such a point), and then we are interested in the distance to a (nearest?) constraint hyperplane (related to the radius of the inscribed ball) and that to a (farthest?) vertex (related to the radius of the circumscribed ball). We are also interested in the angle between constraint hyperplanes. (In this regard, see Van Dam, Frenk, and Telgen [32], who also warn of the dangers of choosing a probabilistic model carelessly.)

We now roughly describe our model. We use n throughout for the number of inequalities and d for the dimension (barring degeneracy) of the feasible region. We denote by $m := n - d$ their difference. Thus our constraints may be given by m equations in n nonnegative variables or n inequalities in d unrestricted variables. Our basic model is given in standard form. We generate $(\hat{x}, \hat{s}) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$ from some "permutation symmetric" distribution; $\hat{x}$ will be a feasible solution to the primal problem, and $\hat{s}$ the objective function gradient (and also a feasible solution to a standard form dual expressed in terms of the slacks in the conventional dual). The feasible region is the intersection of the nonnegative orthant and an affine space that is the translation by $\hat{x}$ of a random d –dimensional subspace in $\mathbb{R}^n$ drawn from the unique distribution that is invariant under orthogonal transformations. The right hand side and objective function are then chosen to ensure feasibility of the given primal and dual solutions, similar to the method of Rosen and Suzuki [23] in nonlinear programming.

Section 2 discusses a distribution on random subspaces that is invariant under orthogonal transformations and uses this to define our basic model for random linear

programming problems precisely and shows how they can be generated. We also demonstrate symmetry under duality for this model. In section 3 we investigate several aspects of nondegenerate problems generated from our model. We give equivalent inequality forms of the problem, which shows that combinatorially (but not metrically) our feasible regions are equivalent to a special case of those of Borgwardt [8,9,10]. We calculate the probability that the feasible region is bounded, which follows from results of Schmidt and Mattheiss [25]. Next we compute the distribution of the distance from $\hat{x} = e$ (e denotes the vector of ones) to a constraint hyperplane (which may not contain a facet) and of the cosine of the angle between the normals to two such. We then obtain the distribution of the basic components of a vertex of the feasible region when $\hat{x} = e$; this is apparently the first time that such a distribution has been calculated for any probabilistic model. We show that the expected distance from $\hat{x}$ to a vertex is infinite. We also obtain tight bounds on the expected number of vertices. In the case $m = d$, for instance, this number grows like $(2.44)^d$, compared to a maximum number of vertices growing like $(2.60)^d$ and an expected number growing like 2^d for the conditioned or unconditioned sign-invariant model. This measure relates to a conjecture of Liebling [17] that the good behavior of the simplex method may be related to the small expected number of vertices in linear programming problems.

In section 4 we show how to generate $\hat{x}$ and $\hat{s}$ in our basic model to produce any degree of primal and dual degeneracy at the optimal solution. We also compute the probability that the feasible region is full-dimensional and show that all vertices other than $\hat{x}$ are nondegenerate with probability one. Section 5 introduces another probabilistic model with much more severe degeneracy using null and unbounded (dual null) variables: Here we must control the coefficient matrices as well as $\hat{x}$ and $\hat{s}$. Some of the properties of this model can be deduced from the fact that, after eliminating the null variables and projecting away the unbounded variables, the result is distributed like our basic model.

Finally, section 6 introduces a model for generating random linear programming problems on the simplex, as in Karmarkar's canonical form. This model is not symmetric under duality. We obtain results parallel to those of section 3, and argue, based on an analysis of G. Samorodnitsky, that the ratio of radii of circumscribing and inscribed balls may be expected to be $O(\sqrt{n})$ typically.

To conclude this section we give three well-known examples of feasible regions of linear programming problems and compute the measures we are interested in for these. Here and throughout the paper, we use e to denote a vector of ones of appropriate but varying dimension, even in a single expression; no confusion should result.

Example 1.1: the cube. Let $m = d$ and let I denote the identity matrix of order m . Consider

$$X = \{x \in \mathbb{R}^n: [I, I]x = e, x \geq 0\}.$$

Then X is the unit d -cube embedded in $\mathbb{R}^n$ centered at $\hat{x} = \frac{1}{2}e$. The distance from $\hat{x}$ to any constraint hyperplane is $\frac{\sqrt{d}}{2}$, that to any vertex $\frac{\sqrt{n}}{2}$ of circumscribed and inscribed balls is thus $\sqrt{d}$. The normals to the constraint hyperplanes (in the null space of the constraint matrix) are of the form $\pm(e_j - e_{d+j})$, $j = 1, \dots, d$, where e_j is the j th unit vector. The cosine of the angle between two of these is either 0 or -1 , with an average of $-1/(2d-1)$. The number of vertices is 2^d .

Example 1.2: the simplex. Let $m = 1$ and

$$X = \{x \in \mathbb{R}^n: e^T x = n, x \geq 0\},$$

the regular simplex of side $n\sqrt{2} = (d+1)\sqrt{2}$ centered at $\hat{x} = e$. The distance from $\hat{x}$ to any constraint hyperplane is $\sqrt{n/(n-1)} = \sqrt{(d+1)/d}$ and that to any vertex $\sqrt{n(n-1)} = \sqrt{(d+1)d}$; thus the ratio of radii of circumscribed and inscribed balls is d . The outward

normals to the constraint hyperplanes are of the form $e - ne_j$, $j = 1, \dots, n$, and the cosine of the angle between two such is $-1/(n-1) = -1/d$. The number of vertices is $n = d+1$.

As an intermediate example, we choose

Example 1.3: the assignment polytope. Here we identify $\mathbb{R}^n$ with $\mathbb{R}^{r \times r}$, the set of $r \times r$ matrices, where $n = r^2$, and let

$$X = \{x \in \mathbb{R}^n: xe = e, e^T x = e^T, x \geq 0\}$$

The equality constraints are dependent; any one can be omitted, giving $m = 2r-1$ and $d = (r-1)^2$. The center of X is $\hat{x} = (1/r)ee^T$ (here $e \in \mathbb{R}^r$, so that ee^T is the $r \times r$ matrix of ones). The closest point to $\hat{x}$ in X on the hyperplane $x_{11} = 0$ is $\hat{x} - \frac{1}{r(r-1)^2} (re_1 - e)(re_1 - e)^T$, so the distance to any constraint hyperplane is $1/(r-1) = 1/\sqrt{d}$. A vertex of X corresponds to a permutation, e.g. $x = I$ (the $r \times r$ identity) so that the distance from $\hat{x}$ to any vertex is $\sqrt{r-1} = d^{1/4}$. Hence the ratio of radii of circumscribed and inscribed balls is $d^{3/4}$, a substantial decrease from d with only $n = d + O(\sqrt{d})$ inequalities. The cosine of the angle between normals to constraint hyperplanes $x_{ij} \geq 0$ and $x_{ik} \geq 0$ or $x_{hj} \geq 0$ is $-1/(r-1) = -1/\sqrt{d}$; that for $x_{ij} \geq 0$ and $x_{hk} \geq 0$ ($h \neq i$, $j \neq k$) is $-1/(r-1)^2 = -1/d$, for an average of $-1/(r^2-1) \sim -1/d$. The number of vertices is $r!$; for $r \geq 4$ this is less than $(r-1)^{r-1} = (\sqrt{d})^{\sqrt{d}} = (\exp(\ln d/2\sqrt{d}))^d$, which grows very slowly. This is mostly due to the very high degree of degeneracy; each vertex corresponds to a huge number of feasible bases.

2. Random subspaces and linear programming problems.

In this section we define a distribution on d -dimensional subspaces of $\mathbb{R}^n$ and hence on standard form linear programming problems.

We start with some well-known properties of normal random variables and random orthogonal matrices. A random vector or matrix is called standard Gaussian if each entry is independently distributed as a standard normal random variable. If $\mathbf{x}$ is a standard Gaussian random vector, so is $Q\mathbf{x}$ for any orthogonal matrix Q , and $\mathbf{x}/\|\mathbf{x}\|$ is uniformly distributed on the unit sphere in $\mathbb{R}^n$, denoted S^{n-1} . Conversely, if $\mathbf{x} \in \mathbb{R}^n$ is distributed in $\mathbb{R}^n$ according to any rotationally-symmetric distribution (i.e., one that is invariant under orthogonal transformations), then $\mathbf{x}/\|\mathbf{x}\|$ is uniformly distributed in S^{n-1} and thus $\mathbf{x}$ is standard Gaussian up to a scale factor.

If $A \in \mathbb{R}^{n \times d}$ is a standard Gaussian random matrix, clearly every column $(\in \mathbb{R}^n)$ and row $(\in \mathbb{R}^d)$ of A is a standard Gaussian random vector, and if U and V are orthogonal matrices of order n and d respectively, UAV is also a standard Gaussian random matrix. Let $\mathcal{O}^{n \times d}$ denote the set of $n \times d$ matrices with orthonormal columns; we call its elements orthonormal matrices. Then $\mathcal{O}^{n \times n}$ is the set of orthogonal matrices of order n . There is a unique distribution (the Haar measure) on $\mathcal{O}^{n \times n}$ that is invariant under left orthogonal multiplication; i.e., the measure of $Q\mathcal{X} := \{QX : X \in \mathcal{X}\}$ for any measurable $\mathcal{X} \subseteq \mathcal{O}^{n \times n}$ and $Q \in \mathcal{O}^{n \times n}$, is equal to that of $\mathcal{X}$. See Halmos [14]. This measure is also invariant under right orthogonal multiplication – the measure of $\mathcal{X}Q := \{XQ : X \in \mathcal{X}\}$ equals that of $\mathcal{X}$. We call this the orthogonally invariant (OI) distribution. If Q is a random orthogonal matrix with the OI distribution, the properties above imply that $Q^{-1} = Q^T$ is too.

Such a random orthogonal matrix can be generated as follows. Let A be a standard Gaussian random $n \times n$ matrix, and factor it as QR , where $Q \in \mathcal{O}^{n \times n}$ and $R \in \mathbb{R}^{n \times n}$ is upper triangular with nonnegative diagonal entries. With probability one, A is nonsingular, R has positive diagonal entries and this factorization is unique. Then Q is

a random orthogonal matrix with the OI distribution. Stewart [29] shows this, and also gives an efficient way to generate Q as a product of Householder matrices. Since the argument is simple and instructive, we give it here. The distribution of A is invariant under left multiplication by orthogonal matrices. Thus $A \stackrel{d}{=} UA$ and hence $AR^{-1} \stackrel{d}{=} UAR^{-1}$ for any $U \in \mathcal{O}^{n \times n}$. But this means that $Q \stackrel{d}{=} UQ$, so that by the uniqueness of the Haar measure, Q has the OI distribution.

Now we choose $d \leq n$ and consider orthonormal $n \times d$ matrices.

Lemma 2.1. There is a unique distribution on $\mathcal{O}^{n \times d}$ that is invariant under left orthogonal multiplication. This distribution is also invariant under right orthogonal multiplication, and is the distribution of the $n \times d$ matrix consisting of the first d columns of a random OI-distributed orthogonal matrix of order n .

(Note that while there is a unique distribution invariant under left orthogonal multiplication, several are invariant under right orthogonal multiplication. For example, we could take $Y = \begin{bmatrix} V \\ 0 \end{bmatrix}$ where V is a random orthogonal $d \times d$ matrix with the OI distribution.)

Proof. Let $U \in \mathcal{O}^{n \times n}$, and let $Q \in \mathcal{O}^{n \times d}$ have the OI distribution. Write $Q = [Y, Z]$, with $Y \in \mathcal{O}^{n \times d}$. Then $UQ = [UY, UZ]$. Since Q and UQ have the same distribution, so do Y and UY . Hence the distribution of Y is invariant under left orthogonal multiplication. Moreover, if $V \in \mathcal{O}^{d \times d}$, then $W := \begin{pmatrix} V & 0 \\ 0 & I \end{pmatrix} \in \mathcal{O}^{n \times n}$ and $QW = [YV, Z]$. Since Q and QW have the same distribution, so do Y and YV . Hence the distribution of Y is also invariant under right orthogonal multiplication.

To show uniqueness, suppose $Y \in \mathcal{O}^{n \times d}$ has a distribution invariant under left orthogonal multiplication. Let $B \in \mathbb{R}^{n \times m}$, where $m := n - d$, be a standard Gaussian random matrix independent of Y , and let $A = [Y, B]$. Then A is nonsingular with

probability one and the distribution of A is invariant under left orthogonal multiplication, so that if $A = QR$ with Q orthogonal and R upper triangular with nonnegative diagonal, Q has the OI distribution. But it is easy to see that Y consists of exactly the first d columns of Q .

We call the distribution in the lemma the OI distribution on $\mathcal{O}^{n \times d}$.

Theorem 2.2 There is a unique distribution on d -dimensional subspaces of $\mathbb{R}^n$ that is invariant under orthogonal transformations. A subspace from this distribution can be generated as:

- a) The range of a random orthonormal $n \times d$ matrix with the OI distribution;
- b) The orthogonal complement of the range of a random orthonormal $n \times m$ matrix, with the OI distribution, where $m := n - d$;
- c) The range of a standard Gaussian random $n \times d$ matrix; or
- d) The null space of a standard Gaussian random $m \times n$ matrix.

Proof. From lemma 2.1, the range of a random orthonormal matrix in $\mathbb{R}^{n \times d}$ with the OI distribution is a random d -dimensional subspace that is invariant under orthogonal transformations. Conversely, given such a subspace L , let $b_1, \dots, b_d$ be an orthonormal basis for L , let $B = [b_1, \dots, b_d]$ and let $Y = BV$, where V is a random orthogonal matrix in $\mathcal{O}^{d \times d}$ with the OI distribution independent of L . If $B' = [b'_1, \dots, b'_d]$ corresponds to another basis, then $B' = BW$ for some $W \in \mathcal{O}^{d \times d}$ so $Y = B'(W^{-1}V)$, and $W^{-1}V$ also has the OI distribution. Hence Y is well-defined as a function of L , and since QB corresponds to a basis for $QL := \{Q\ell : \ell \in L\}$, it is easy to see that Y has the OI distribution on $\mathcal{O}^{n \times d}$. Since L is the range of Y , the random d -subspace L arises as in (a). From lemma 2.1, the distribution is unique.

Let Z be a random orthonormal $n \times m$ matrix with the OI distribution, so that $Q = [Z, Y]$ has the OI distribution on $\mathcal{O}^{n \times n}$ for some Y . Since permutation matrices are orthogonal, $Q' = [Y, Z]$ is also OI-distributed so that Y is a random orthonormal $n \times d$ matrix with the OI distribution. But the orthogonal complement of the range of Z is the range of Y , so that (b) generates the same distribution as (a).

Let A be a standard Gaussian random $n \times d$ matrix, and let $A = YR$ with $Y \in \mathcal{O}^{n \times d}$ and R $d \times d$ upper triangular with nonnegative diagonal. Then as in the square case Y is invariant under left orthogonal multiplication, so is OI-distributed. Since the range of A is with probability one that of Y , (c) generates the same distribution as (a).

Finally, if A is a standard Gaussian random $m \times n$ matrix, then so is A^T . Let $A^T = ZR$ as above; then Z has the OI distribution on $\mathcal{O}^{n \times m}$. With probability one, the null space of A is the orthogonal complement of the range of Z , so that (d) generates the same distribution as (b).

We say a random subspace generated as in the theorem has the OI distribution. From (a) and (b) we deduce

Corollary 2.3 The orthogonal complement of a random d -subspace with the OI distribution is a random m -subspace with the OI distribution, where $d+m = n$.

We are now ready to define our first model for random linear programming problems.

Model 1. Let $\{\mu_{md} : m+d = n, m > 0, d > 0\}$ be a family of distributions on $\mathbb{R}_+^n \times \mathbb{R}_+^n$ that is "permutation symmetric" in the sense that, if $(\hat{x}, \hat{s})$ has distribution μ_{md} , then $(\hat{s}, \hat{x})$ has distribution μ_{dm} . Let $(\hat{x}, \hat{s})$ have distribution μ_{md} , and let A

be a standard Gaussian random matrix independent of $(\hat{x}, \hat{s})$. Then we say that the random standard-form $m \times n$ linear programming problem

$$\begin{aligned} (P_1) \quad & \min \hat{s}^T x \\ & Ax = b \\ & x \geq 0, \end{aligned}$$

where $b := A\hat{x}$, is distributed according to model 1.

Note that $\hat{x}$ is always feasible to (P_1) . Also, the equality constraints can be written as $x - \hat{x} \in L$, where L is the null space of A . From theorem 2.2 we see that the standard form problem

$$\begin{aligned} \min \quad & \hat{s}^T x \\ & x - \hat{x} \in L \\ & x \geq 0 \end{aligned}$$

is also distributed according to model 1 if $(\hat{x}, \hat{s})$ has distribution μ_{md} and L has the OI distribution on d -subspaces of $\mathbb{R}^n$ and is independent of $(\hat{x}, \hat{s})$.

The dual to (P) is normally written as

$$\begin{aligned} \max \quad & (A\hat{x})^T y \\ & A^T y + s = \hat{s} \\ & s \geq 0 \end{aligned}$$

with $s \in \mathbb{R}^n$ a vector of slack variables. However, a more symmetric form results from eliminating the unconstrained variables y . If the rows of matrix B span the null space of A , then the dual can be written in terms of s alone (see [31]) as

$$\begin{aligned} (D_1) \quad & \min \hat{x}^\top s \\ & Bs = c \\ & s \geq 0 \end{aligned}$$

where $c := B\hat{s}$. The equality constraints of (D) can be written as $s - \hat{s} \in L^\perp$, where $L^\perp$ is the orthogonal complement of the null space of A . From corollary 2.3 and the requirement that $\{\mu_{\text{md}}\}$ be permutation symmetric, we deduce

Theorem 2.4 The dual of a random $m \times n$ linear programming problem distributed according to model 1 is a random $d \times n$ linear programming problem that is also distributed according to model 1.

Since $\hat{x}$ is always feasible in (P) and $\hat{s}$ in (D), we have

Proposition 2.5 A random linear programming problem distributed according to model 1 (and hence its dual) always has an optimal solution.

To conclude this section we describe two ways to choose a permutation symmetric $\{\mu_{\text{md}}\}$. First, we can choose it to concentrate all its mass on the point (e, e) , so that $\hat{x} = \hat{s} = e$. (Recall that e is the vector of ones.) In this case, (P_1) and (D_1) have feasible regions with interiors (relative to the appropriate affine subspaces), which are useful for considering interior point algorithms or measures of polyhedra that require known interior

points. Second, we can choose $\{\mu_{md}\}$ so that $\hat{x}$ and $\hat{s}$ are optimal solutions to (P_1) and (D_1) respectively. By complementary slackness it is necessary and sufficient that $\hat{x}^T \hat{s} = 0$. Choose an m -subset of $\{1, 2, \dots, n\}$ at random, with all m -subsets equally likely, and choose a vector $q \in \mathbb{R}_+^n$ from some distribution (independent of A). Let the components of $\hat{x}$ be those of q for coordinates in the chosen m -subset and zero otherwise, and let the components of $\hat{s}$ be those of q for components not in the chosen m -subset and zero otherwise. This allows us to specify the optimal solutions, and also specify any desired degree of primal and dual degeneracy at the optimal solutions; we investigate the consequences in section 4.

3. Properties of random linear programming problems.

In this section (P_1) denotes a random $m \times n$ standard form linear programming problem distributed according to model 1. We concentrate on the case where all components of $\hat{x}$ and $\hat{s}$ are positive with probability one, so that both primal and dual feasible regions have nonempty interiors (relative to the affine subspace constraints). For the most part we assume that $\hat{x} = \hat{s} = e$ with probability one, but some results are valid for any distribution on $(\hat{x}, \hat{s})$, and others for any distribution with compact support.

3.1 Relationship to inequality-constrained problems

Problem (P_1) can be written as

$$\begin{aligned} \min \quad & \hat{s}^T x \\ & x - \hat{x} \in L \\ & x \geq 0 \end{aligned}$$

where L is a random d -subspace in $\mathbb{R}^n$ with the OI distribution. By theorem 2.2, L can

be written as $\{Qu: u \in \mathbb{R}^d\}$ where Q has the OI distribution on $\mathcal{O}^{n \times d}$. By substituting $x = \hat{x} + Qu$, we obtain:

Proposition 3.1 (P_1) is equivalent to the random linear programming problem

$$(P_{1u}) \quad \begin{aligned} & \min (Q^T \hat{s})^T u \\ & Qu \geq -\hat{x}, \end{aligned}$$

where $(\hat{x}, \hat{s})$ has distribution μ_{md} and Q is a random $n \times d$ orthonormal matrix with the OI distribution, independent of $(\hat{x}, \hat{s})$. Moreover, the mapping $u \rightarrow \hat{x} + Qu$ is an isometry between the feasible region of (P_{1u}) and that of (P_1) .

Since we have an isometry, measures such as distances to hyperplanes or vertices, or angles between hyperplanes, are the same in (P_1) and in (P_{1u}) . This is not the case for the next transformation.

From theorem 2.2, L can also be written as $\{-Ev: v \in \mathbb{R}^d\}$, where E is a standard Gaussian random $n \times d$ matrix. (In fact, we can take $E = B^T$, where B is the constraint matrix of (D_1) .) By substituting $x = \hat{x} - Ev$ we obtain

Proposition 3.2. (P_1) is equivalent to the random linear programming problem

$$(P_{1v}) \quad \begin{aligned} & \max (E^T \hat{s})^T v \\ & Ev \leq \hat{x} \end{aligned}$$

where $(\hat{x}, \hat{s})$ has distribution μ_{md} and E is a standard Gaussian random $n \times d$ matrix independent of $(\hat{x}, \hat{s})$.

Here the mapping $v \rightarrow \hat{x} - Ev$ is not an isometry, but at least preserves the combinatorial structure of the feasible region of (P_{1v}) , so that measures such as probability of boundedness and expected number of vertices are the same in (P_{1v}) and (P_1) .

If $\hat{x} = \hat{s} = e$ with probability one, (P_{1v}) is almost a special case of the random linear programming problem studied extensively by Borgwardt; see [8,9,10]. The only difference is that Borgwardt uses an objective function $c^T v$ where c is a standard Gaussian random vector independent of E . Our objective $(E^T e)^T v$ assures the existence of an optimal solution; $E^T e$ is standard Gaussian, but is clearly not independent of E .

If $\hat{x}$ is positive with probability one, then $u = 0$ and $v = 0$ lie in the interiors of the feasible regions of (P_{1u}) and (P_{1v}) respectively with probability one. Borgwardt uses this to start his algorithm at the feasible point $v = 0$; even if $\hat{x} = e$ with probability one, our intention is that algorithms for (P_1) not be aware of this fact, so that a genuine phase I (which is guaranteed success) is required.

3.2. Probability of boundedness.

Here we prove

Theorem 3.3. The probability that the feasible region of (P_1) is bounded is

$$p_{\text{nd}} := \frac{\binom{n-1}{0} + \dots + \binom{n-1}{m-1}}{2^{n-1}}.$$

Proof. We compute the probability q_{nd} that the feasible region of (P_1) is unbounded. This is the probability that $Ax = 0$, $x \geq 0$, $x \neq 0$, has a solution, where A is a standard Gaussian random $m \times n$ matrix. Now by Theorem 2.2, $L = \{x: Ax = 0\}$ can be written as $\{Ev: v \in \mathbb{R}^d\}$, where E is a standard Gaussian random $n \times d$ matrix. Hence

q_{nd} is the probability that $Ev \geq 0$, $Ev \neq 0$ has a feasible solution. But this has been computed by Schmidt and Matheiss [25], and yields the desired p_{nd} .

By theorem 2.4, the corresponding probability for (D) is p_{nm} obtained by substituting m for d (and hence d for m). From the theorem above, this is $1-p_{nd}$. Now it is well-known (see, e.g., [16; pp. 688–9]) that at least one of any pair of dual linear programming problems has an unbounded feasible region. Because $p_{nd} + p_{nm} = 1$, it follows that, with probability one, exactly one of (P_1) and (D_1) has a bounded feasible region. A similar result for a different model appears in Adler and Berenguer [3].

3.3 Distance to constraint hyperplanes.

Here we compute the distribution of the distance from $\hat{x} = e$ to a constraint hyperplane $x_j = 0$ of (P_1) in the affine space $Ax = b$. Note that this hyperplane may or may not correspond to a facet of the feasible region of (P_1) . It does not appear to be easy to compute the probability that such a hyperplane does not correspond to a facet, i.e., that a particular constraint is redundant, whereas this is simple in the sign-invariant model of Adler–Berenguer [2,3] and May–Smith [18]. However, they do not compute the distribution of the distance, as we do.

Theorem 3.4. Assume that $\hat{x} = e$ with probability one. Then the distance from $\hat{x}$ to the hyperplane $x_j = 0$ in the affine space $Ax = b$ is distributed like $\beta^{-\frac{1}{2}}$, where β is a beta random variable with parameters $d/2$ and $m/2$. It has expected value

$$\frac{\Gamma(n/2) \Gamma((d-1)/2)}{\Gamma((n-1)/2) \Gamma(d/2)}$$

which is asymptotic to $(\frac{n}{d})^{\frac{1}{2}}$ as $d \rightarrow \infty$ (and hence $n \rightarrow \infty$).

Proof. By proposition 3.1 we can equivalently compute the distribution of the distance from $u = 0$ to the hyperplane $q_j^\top u = -1$, where $q_j^\top$ is the j th row of the random orthonormal OI-distributed $n \times d$ matrix Q . But this distance is $\|q_j\|^{-1}$. Now q_j consists of the first d components of a column of a random orthogonal matrix with the OI distribution, which has rotational symmetry. Hence this column can be generated as $(\lambda_1/\|\lambda\|, \dots, \lambda_n/\|\lambda\|)$, where λ is a standard Gaussian random vector, and so $\|q_j\|^{-1}$ is distributed like $\beta^{-\frac{1}{2}}$, where $\beta = (\lambda_1^2 + \dots + \lambda_d^2)/(\lambda_1^2 + \dots + \lambda_n^2)$. Since $\lambda_1^2 + \dots + \lambda_d^2$ has the gamma distribution with parameter $d/2$ (or the chi-square distribution with d degrees of freedom) and $\lambda_{d+1}^2 + \dots + \lambda_n^2$ is independent with gamma distribution with parameter $m/2$, β has the beta distribution with parameters $d/2$ and $m/2$. The expected value of $\beta^{-\frac{1}{2}}$ is as given – see, for instance, Wilks [34,p.174]. The approximation results from Stirling's formula.

Since β lies in $(0,1)$ with probability one, the distance to a hyperplane lies in $(1,\infty)$; this is also clear since $\|q_j\| < 1$ with probability one. Thus a ball of radius one around $\hat{x}$ is always contained in the feasible region. In fact, if n/d is bounded, the radius of the largest inscribed ball around $\hat{x}$ is of order one. Indeed, the distance to any hyperplane is concentrated around $[(n-2)/(d-2)]^{\frac{1}{2}}$. To see this, we note that β^{-1} has expected value $(n-2)/(d-2)$ and variance $2m(n-2)/(d-2)^2(d-4)$. If $m = d$, these are of order of magnitude 2 and $4/d$ respectively, and Chebychev's inequality then demonstrates the concentration.

3.4 Angles between constraint hyperplane normals.

In this subsection we obtain the distribution of the angle between the outward normals to two constraint hyperplanes; again, we note that these may not correspond to

facets of the feasible region. Let γ be a random variable denoting the cosine of the angle between the outward normals to the first two constraints of (P_1) . We will prove

Theorem 3.5. γ is distributed like

$$\frac{\epsilon}{(1 + (d-1)\frac{F}{\beta})^{\frac{1}{2}}}$$

where ϵ , F and β are independent, ϵ is ± 1 each with probability $\frac{1}{2}$, F has the Snedecor (F —) distribution with $d-1$ and 1 degrees of freedom and β the beta distribution with parameters $m/2$ and $d/2$.

Proof. Let q_1^T and q_2^T be the first two rows of a random orthonormal $n \times d$ matrix Q with the OI distribution. Then, by proposition 3.1, γ is distributed like $q_1^T q_2 / \|q_1\| \|q_2\|$. We generate q_1 and q_2 as follows. Let A be a standard Gaussian random matrix and factor $A = UR$ with $U \in \mathcal{O}^{n \times n}$ and R upper triangular with nonnegative diagonal. Finally, let q_1 and q_2 be the first d components of u_1 and u_2 , the first two columns of U .

Now the first two columns of A are $(\lambda_1, \dots, \lambda_n)$ and $(\mu_1, \dots, \mu_n)$ with the λ_i 's and μ_i 's independent standard normal random variables. Because of orthogonal invariance, we can perform rotations in the first d and last m coordinates so that $(\lambda_1, \dots, \lambda_n)$ becomes $(\xi, 0, \dots, 0, \eta)$, where ξ is $N(0, d)$ and $\eta \sim N(0, m)$. Moreover, the second column of A is still of the form $(\mu_1, \dots, \mu_n)$ with the μ_i 's as above. Then

$$\begin{aligned}
q_1 &= (\xi/(\xi^2+\eta^2)^{\frac{1}{2}}, 0, \dots, 0) \\
q_2 &= \theta(\mu_1 - \frac{\xi\mu_1 + \eta\mu_n}{\xi^2 + \eta^2} \xi, \mu_2, \dots, \mu_d) \\
&= \theta(\frac{\eta}{(\xi^2 + \eta^2)^{\frac{1}{2}}} \frac{\eta\mu_1 - \xi\mu_n}{(\xi^2 + \eta^2)^{\frac{1}{2}}}, \mu_2, \dots, \mu_d)
\end{aligned}$$

for a suitable normalizing factor θ . Now for any ξ, η , $(\eta\mu_1 - \xi\mu_n)/(\xi^2 + \eta^2)^{\frac{1}{2}}$ is distributed like a standard normal random variable ν_1 independent of ξ, η and $\mu_2, \dots, \mu_d$. Thus γ is distributed like $(q_1/\|q_1\|)^T q_2/\|q_2\|$, and simple algebraic manipulations reduce this to the form in the theorem, where $\epsilon = \text{sign}(\xi\eta\nu_1)$ is ± 1 each with probability $\frac{1}{2}$,

$F = \frac{(\mu_2^2 + \dots + \mu_d^2)/(d-1)}{\nu_1^2}$ has a Snedecor distribution with $d-1$ and 1 degrees of freedom,

and $\beta = \eta^2/(\xi^2 + \eta^2)$ has a beta distribution with parameters $m/2$ and $d/2$. Moreover, these are independent.

The theorem implies that the expected value of γ is 0, since its distribution is symmetric about 0; the same is true for the angle between the two constraint hyperplanes. The variance of γ is

$$E\gamma^2 = E\left[\frac{1}{1+(d-1)F/\beta}\right] = E\left[1 - \frac{d-1}{d-1+\beta/F}\right].$$

Now Jensen's inequality shows that $E\gamma^2 \leq 1 - (d-1)/(d-1+E(\beta/F))$, which can be written as $1 - (d-1)/(d-1+E\beta.EF')$, where F' has a Snedecor distribution with 1 and $d-1$ degrees of freedom. If $m \leq d^2 - 2d$, the latter expression is at most $1/d$, since $E\beta = m/n$ and $EF' = (d-1)/(d-2)$. Thus constraint hyperplanes tend to be more orthogonal than random OI-distributed hyperplanes in a d -dimensional subspace. (Note that if u and v

are standard Gaussian random vectors in $\mathbb{R}^d$, the normals to two random hyperplanes, then without loss of generality $u = (1, 0, \dots, 0)^T$ and $v = (\lambda_1, \dots, \lambda_d)^T$ with each λ_i independently distributed as a standard normal random variable. Then the square of the cosine of the angle between u and v is $\lambda_1^2 / (\lambda_1^2 + \dots + \lambda_d^2)$, which has a beta distribution with parameters $1/2$ and $(d-1)/2$ and mean $1/d$.)

3.5 Vertices.

Here we are concerned with the distribution of the basic components of a vertex of the feasible region of the random linear programming problem (P_1) . The next subsection deals with the expected number of vertices. To our knowledge this is the first time that the actual distribution of the components of a vertex has been computed for any probabilistic model of a polyhedron.

The development below relies on the choice $\hat{x} = e$. A similar analysis is possible for any fixed choice of $\hat{x}$, for example $\hat{x}_j = 1$ for $1 \leq j \leq m$, $\hat{x}_j = 0$ for $m < j \leq n$. It is convenient henceforth to use δ for $d^{\frac{1}{2}}$.

Theorem 3.6. Suppose $\hat{x} = e$ with probability one in (P_1) . Let $\lambda_0, \lambda_1, \dots, \lambda_m$ be independent $N(0,1)$ random variables, and condition on the event that $\lambda_i \leq |\lambda_0|/\delta$ for $i = 1, \dots, m$. Then a vertex of (P_1) has basic variables distributed like $1 - \delta\lambda_i/|\lambda_0|$ for $i = 1, \dots, m$ (and nonbasic variables 0).

Proof. Any vertex corresponds to a basis matrix. Since $\hat{x} = e$ and the distribution of A is invariant under permutations of the columns, we can consider without loss of generality the basis matrix A_β consisting of the first m columns of A . Let A_ν consist of

the remaining columns. If $\mathbf{x} =: \begin{bmatrix} \bar{\mathbf{x}}_\beta \\ \bar{\mathbf{x}}_\nu \end{bmatrix}$ is the corresponding basic (not necessarily feasible) solution, then $\bar{\mathbf{x}}_\nu = 0$ and

$$\begin{aligned} \bar{\mathbf{x}}_\beta &= \mathbf{A}_\beta^{-1} \mathbf{b} = \mathbf{A}_\beta^{-1} (\mathbf{A}_\beta \mathbf{e} + \mathbf{A}_\nu \mathbf{e}) \\ &= \mathbf{e} - \delta \mathbf{A}_\beta^{-1} (-\delta^{-1} \mathbf{A}_\nu \mathbf{e}) \end{aligned}$$

where we have used $\mathbf{e}$ for vectors of ones of appropriate dimensions. Let

$\mathbf{u} := \mathbf{A}_\beta^{-1} (-\delta^{-1} \mathbf{A}_\nu \mathbf{e})$ and $\mathbf{a}_0 := \delta^{-1} \mathbf{A}_\nu \mathbf{e}$, which is a standard Gaussian random vector independent of the columns $\mathbf{a}_1, \dots, \mathbf{a}_m$ of $\mathbf{A}_\beta$. Thus $\bar{\mathbf{x}}$ is feasible and hence a vertex if $\mathbf{u} \leq \delta^{-1} \mathbf{e}$.

Now

$$(\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_m) \begin{pmatrix} 1 \\ \mathbf{u} \end{pmatrix} = 0,$$

and $\bar{\mathbf{A}} = (\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_m)$ is a standard Gaussian random $m \times (m+1)$ matrix. By theorem 2.2, its null space is a random 1-subspace of $\mathbb{R}^{m+1}$ with the OI distribution, and hence the line generated by a standard Gaussian random vector $(\lambda_0, \lambda_1, \dots, \lambda_m)$. Without loss of generality, we can assume λ_0 is positive, and then $u_i = \lambda_i / \lambda_0 = \lambda_i / |\lambda_0|$ for $i = 1, \dots, m$.

Hence $\bar{\mathbf{x}}$ is a vertex iff $\bar{x}_i = 1 - \delta u_i \geq 0$ for $i = 1, \dots, m$, which holds iff $\lambda_i \leq |\lambda_0| / \delta$ for $i = 1, \dots, m$, and then $\bar{x}_i = 1 - \delta u_i = 1 - \delta \lambda_i / |\lambda_0|$ for $i = 1, \dots, m$, $\bar{x}_j = 0$ for $j = m+1, \dots, n$. This completes the proof.

Corollary 3.7. If $\hat{\mathbf{x}} = \mathbf{e}$ with probability one, then every vertex of (P_1) is nondegenerate with probability one.

Corollary 3.8. The expected distance of a vertex $\bar{\mathbf{x}}$ of (P_1) from $\hat{\mathbf{x}} = \mathbf{e}$ is infinite.

Proof. Let $\bar{x} = \begin{bmatrix} \bar{x} \\ \bar{x}_\nu \end{bmatrix} = \begin{bmatrix} e^{-\delta u} \\ 0 \end{bmatrix}$ as in the proof above. Then $\|\bar{x} - e\| = (d + d\|u\|^2)^{\frac{1}{2}} > \delta\|u\|$. Now certainly $\lambda_i \leq |\lambda_0|/\delta$ for all i if $\lambda_i \leq 0$ for all i , and this has probability 2^{-m} . Hence

$$\begin{aligned} E(\|\bar{x} - e\|) &> \delta E(\|\lambda\|/|\lambda_0| \mid \lambda_i \leq |\lambda_0|/\delta \text{ for all } i) \\ &\geq \delta 2^{-m} E(\|\lambda\|/|\lambda_0| \mid \lambda_i \leq 0 \text{ for all } i). \end{aligned}$$

Now the distribution of $\|\lambda\|$ is the same when conditioned on $\lambda_i \leq 0$ for all i or when unconditioned. Also, $\|\lambda\|/|\lambda_0| = ((\lambda_1^2 + \dots + \lambda_m^2)/\lambda_0^2)^{\frac{1}{2}} = (mF)^{\frac{1}{2}}$, where F has a Snedecor distribution with parameters m and 1. But, in such a case, $F^{\frac{1}{2}}$ has infinite mean, yielding the result.

3.6. Expected number of vertices

Let V denote the number of vertices of the feasible region of (P_1) with $\hat{x} = e$. There are $\binom{n}{m} = \binom{n}{d}$ basic solutions, and the probability of each being feasible can be deduced from theorem 3.6. By conditioning on $|\lambda_0|$, we find

$$EV = \binom{n}{d} \int_0^\infty [\Phi(\kappa/\delta)]^m (2\varphi(\kappa)) d\kappa, \quad (1)$$

where Φ is the cumulative distribution function and φ the density function of a standard normal random variable (so that 2φ is the density of $|\lambda_0|$). Now making the change of variable $\theta = \kappa/\delta$, and substituting $\varphi(\kappa) = (2\pi)^{-\frac{1}{2}} \exp(-\kappa^2/2)$, we find

$$EV = (2/\pi)^{\frac{1}{2}} \binom{n}{d} \delta \int_0^\infty [\Phi(\theta)]^m [\exp(-\theta^2/2)]^d d\theta. \quad (2)$$

(A very closely related expression, obtained by very different methods, is given by Raynaud [22, eqn. (1.13)]. The slight discrepancy is caused by the fact that some of Raynaud's facets (those separating the origin from the polytope) do not correspond via polarity to vertices of the polar.)

Theorem 3.8. Suppose $m = \alpha d$ and let $\tau = \tau(\alpha)$ be the maximum value of $[\Phi(\theta)]^\alpha \exp(-\theta^2/2)$ and $\sigma = \tau \cdot (1+\alpha)^{1+\alpha} \cdot \alpha^{-\alpha}$. Then, for any $\epsilon > 0$, there are positive constants $c_1 = c_1(\alpha, \epsilon)$ and $c_2 = c_2(\alpha)$ such that

$$c_1(\sigma - \epsilon)^d \leq \text{EV} \leq c_2 \sigma^d. \quad (3)$$

Proof. (The argument below is very similar to Laplace's method in the asymptotic expansion of integrals, see Bleisten and Handelsman [7], which can be used to obtain more precise asymptotic estimates. Our result is valid for all d .) First note that, if $f_\alpha(\theta) := [\Phi(\theta)]^\alpha \exp(-\theta^2/2)$, then $f'_\alpha(\theta) = [\Phi(\theta)]^{\alpha-1} \exp(-\theta^2/2)(\alpha\varphi(\theta) - \theta\Phi(\theta))$, which is positive for $\theta \leq 0$, strictly decreasing for $\theta \geq 0$, and negative for sufficiently large θ , so that f_α has a unique maximum $\tau = \tau(\alpha)$ at some positive θ . Next note that

$$\begin{aligned} \binom{n}{d} &= \frac{[(1+\alpha)d]!}{d! \alpha d!} \approx \frac{(1+\alpha)^{(1+\alpha)d} \sqrt{2\pi(1+\alpha)d}}{\sqrt{2\pi d} \alpha^{\alpha d} \sqrt{2\pi \alpha d}} \\ &= (2\pi \alpha d / (1+\alpha))^{-\frac{1}{2}} ((1+\alpha)^{1+\alpha} \alpha^{-\alpha})^d \end{aligned}$$

by Stirling's formula as $d \rightarrow \infty$, so that

$$(2/\pi)^{\frac{1}{2}} \binom{n}{d} \delta \approx c_3 ((1+\alpha)^{1+\alpha} \alpha^{-\alpha})^d \quad (4)$$

for some constant $c_3 = c_3(\alpha)$. Moreover, we can obtain two-sided bounds similar to (4) and valid for all d . For the upper bound, first choose some constant c_4 so that

$[\exp(-\theta^2/2)]^{\frac{d-1}{d}} \leq \tau$ for $\theta \geq c_4$. Then

$$\begin{aligned} & \int_0^\infty [\Phi(\theta)]^m [\exp(-\theta^2/2)]^d d\theta \\ & \leq \int_0^{c_4} \{[\Phi(\theta)]^\alpha \exp(-\theta^2/2)\}^d d\theta + \sqrt{2\pi} \int_{c_4}^\infty 1 \{[\exp(-\theta^2/2)]^{\frac{d-1}{d}}\}^d \varphi(\theta) d\theta \\ & \leq (c_4 + \sqrt{2\pi}) \tau^d, \end{aligned} \tag{5}$$

which with (4) yields the right hand inequality of (3). For the lower bound, choose constants $c_5 = c_5(\alpha, \epsilon)$ and $c_6 = c_6(\alpha, \epsilon) > c_5$ so that

$f_\alpha(\theta) \geq \tau - \epsilon \cdot \alpha^\alpha (1 + \alpha)^{-1 - \alpha}$ for $c_5 \leq \theta \leq c_6$. Then

$$\begin{aligned} \int_0^\infty [\Phi(\theta)]^m [\exp(-\theta^2/2)]^d d\theta & \geq \int_{c_5}^{c_6} \{[\Phi(\theta)]^\alpha \exp(-\theta^2/2)\}^d d\theta \\ & \geq (c_6 - c_5) (\tau - \epsilon \alpha^\alpha (1 + \alpha)^{-1 - \alpha})^d, \end{aligned}$$

which with (4) yields the left hand inequality of (3).

We now evaluate the base of the exponent in (3) for three values of α and compare the growth of EV with its growth in other models.

If $\alpha = 1$ so that $m = d$, then f_α is maximized for some θ between .5060 and .5061, and $.61023 < \tau < .61027$. Thus $\sigma \approx 2.44$, and EV grows like $(2.44)^d$. (The probability that a basic solution is feasible shrinks like τ^d or $.61^d$.) By comparison, the expected number of vertices in the sign-invariant model of Adler and Berenguer [2,3] and May and Smith [17] (who condition on the nonemptiness of the feasible region) and Prekopa [21] (who doesn't) grows like 2^d . The maximum number of vertices for a d -polytope with exactly n facets (in the probabilistic models, the number of facets is at most n) is $\binom{n - \lceil \frac{d+1}{2} \rceil}{n-d} + \binom{n - \lceil \frac{d+2}{2} \rceil}{n-d}$ (McMullen [19]); using Stirling's formula, this grows like $(3\sqrt{3}/2)^d$ or $(2.598)^d$.

If $\alpha = 2$ so that $m = 2d$ and $n = 3d$, then f_α is maximized between .7652 and 7653, and $.45157 < \tau < .45161$. Thus $\sigma \approx 3.048$. Again, the sign-invariant model yields $\approx 2^d$ as the expected number of vertices, while the maximum number grows like $(3.494)^d$. (In the sign-invariant model, the expected number of facets is less than $2d = (2/3)n$; since $3.048 > 2.598$, it appears likely that the expected number of facets in our model is greater than $2d$, since there are "too many vertices".)

Finally, if $\alpha = \frac{1}{2}$ so that $m = \frac{1}{2}d$ and $n = 3d/2$, then f_α is maximized for some θ between .3067 and .3068, with $.75150 < \tau < .75153$. Thus $\sigma \approx 1.952$. Here the sign-invariant model gives growth like $(1.837)^d$, while the maximum number grows like 2^d .

Finally, we mention that Schmidt and Mattheiss [25] used linear regression to fit a model for $\log(\text{EV})$. Their third model is similar to our model for generating random polyhedra (in the form of the inequality problem (P_{1v}) of section 3.1). Taking only the leading terms of their fitted model yields a prediction of asymptotic growth of $(1.43)^d$ for $\alpha = 1$, $(1.75)^d$ for $\alpha = 2$, and $(1.23)^d$ for $\alpha = \frac{1}{2}$. The results are noticeably lower than

our exact asymptotic results because of the extra terms in their model and the fact that their data was limited to very small values of d and n .

4. Random linear programming problems with degenerate optima.

In this section we assume $(\hat{x}, \hat{s})$ is chosen so that $\hat{x}$ and $\hat{s}$ are optimal solutions to (P_1) and (D_1) as at the end of Section 2. Moreover, we can prescribe any degree of primal and dual degeneracy. Because of the invariance of the distribution of A under permutations of columns, we can assume that the first m components of $\hat{s}$ and the last d of $\hat{x}$ are zero with probability one. (Of course, an algorithm cannot use the fact that the first m indices yield an optimal basic index set.) To be specific, we choose a fixed k , $0 < k \leq m$, and r , $0 < r \leq d$, and assume that the first k (last r) components of $\hat{x}$ ($\hat{s}$) are positive, with the rest zero, with probability one.

In this model, the inequality forms as given in propositions 3.1 and 3.2 remain equivalent, and theorem 3.3 giving the probability that the feasible region is bounded is still valid, as is theorem 3.5 on the angles between constraint hyperplanes. Moreover, if the components of $\hat{x}$ are bounded with probability one, then, by considering a basis with one of the first k indices nonbasic, one can show as in corollary 3.8 that the expected distance to a vertex other than $\hat{x}$ is infinite.

Similarly, following the proof of theorem 3.6 one obtains

Proposition 4.1. In the model above, for any vertex $\bar{x}$ other than $\hat{x}$, the probability that $\bar{x}$ is degenerate is zero.

Since $\hat{x}$ lies in the boundary of $\mathbb{R}_+^n$, it is quite possible that the feasible region of (P) contains no strictly positive points, or that it reduces to the singleton $\{\hat{x}\}$. However, we have

Proposition 4.2. In the model above, let $F(P_1)$ denote the feasible region of (P_1) and $\mathbb{R}_{++}^n$ the strictly positive orthant. Then

$$P(F(P_1) \cap \mathbb{R}_{++}^n \neq \emptyset) = P(F(P_1) \neq \{\hat{x}\}) = p := \frac{\binom{n-k-1}{0} + \dots + \binom{n-k-1}{d-1}}{2^{n-k-1}}$$

Proof. Let L be the random OI-distributed d -subspace $\{x: Ax = 0\}$, so that $F(P_1) = (L + \{\hat{x}\}) \cap \mathbb{R}_{++}^n$. Now the $n-k$ hyperplanes $x_j = 0$, $j = k+1, \dots, n$, go through $\hat{x}$, and their restrictions to L are in general position with probability one. In this case, they divide L into $2[\binom{n-k-1}{0} + \dots + \binom{n-k-1}{d-1}]$ d -dimensional cones (see, e.g. Winder [36]). Since L has the OI distribution, it is invariant under sign reversals of the components, so the probability that any of the 2^{n-k} "quadrants" obtained by choosing signs for $x_{k+1}, \dots, x_n$ corresponds to one of these cones is the same, and hence equal to p . Finally, this is equal to both $P(L + \{\hat{x}\}) \cap \mathbb{R}_{++}^n \neq \emptyset$ and $P((L + \{\hat{x}\}) \cap (\mathbb{R}_+^n \setminus \{\hat{x}\}) \neq \emptyset) = P(F(P) \neq \{\hat{x}\})$ since the hyperplanes are in general position.

For $k = m$, $p = 1$ as we expect, since then $\hat{x}$ is a nondegenerate basic feasible solution, and hence increasing all nonbasic components a sufficiently small amount and adjusting the basic components correspondingly yields a strictly positive feasible point. Also, if $k \geq m-d+1$, then $d-1 \geq \frac{1}{2}(n-k-1)$ so that $p \geq \frac{1}{2}$; this holds even if $d > m$ and $k = 0$, in which case $\hat{x} = 0$ and $F(P_1)$ is a cone, d -dimensional with probability p .

5. Generating problems with null and unbounded variables.

Proposition 4.1 shows that the model of the previous section generates problems with a very limited degree of degeneracy, only appearing at the optimal solution. Such problems may be appropriate for testing interior-point algorithms, which proceed through

the (relative) interior of the feasible region to an optimal point, usually only approaching one vertex closely, but are not realistic for edge-following algorithms since typical real-world problems have the majority of their basic feasible solutions degenerate. This section considers another model for generating degenerate problems.

We concentrate on the incorporation of null variables, those which are zero in all feasible solutions. Null variables in the dual correspond to unbounded variables in the primal, those that can be unbounded in an optimal solution. Suppose the last r primal variables and the first k dual (slack) variables are null. We enforce this by requiring that the sum of these nonnegative variables be zero. Our problems then take the form:

$$\begin{aligned}
 \min \quad & \hat{s}_2^T x_2 + \hat{s}_3^T x_3 \\
 & e^T x_3 = 0 \\
 (P_2) \quad & A_1 x_1 + A_2 x_2 + A_3 x_3 = b := A_1 \hat{x}_1 + A_2 \hat{x}_2 \\
 & x_1 \geq 0, \ x_2 \geq 0, \ x_3 \geq 0
 \end{aligned}$$

and

$$\begin{aligned}
 \min \quad & \hat{x}_1^T s_1 + \hat{x}_2^T s_2 \\
 & e^T s_1 = 0 \\
 (D_2) \quad & B_1 s_1 + B_2 s_2 + B_3 s_3 = c := B_2 \hat{s}_2 + B_3 \hat{s}_3 \\
 & s_1 \geq 0, \ s_2 \geq 0, \ s_3 \geq 0.
 \end{aligned}$$

Here the partitions are into the first k , middle $n-k-r$, and last r components. We assume that $k < m$ and $r < d$, where $m-1$ is the number of rows of A_1 , A_2 and A_3 and $d = n-m$. We abuse notation in using e for a vector of ones of appropriate dimension, since e_j is typically reserved for the j th unit vector. Since $\hat{x}$ is feasible, $\hat{x}_3 = 0$, and similarly $\hat{s}_1 = 0$. (As we shall see below, it is wise to ensure $\hat{x}_1 \neq e$ and similarly $\hat{s}_3 \neq e$.)

In order that these are dual problems, we require that the rows of $A = (A_1, A_2, A_3)$ be orthogonal to $(e^\top, 0, 0)^\top$ and to the rows of $B = (B_1, B_2, B_3)$, which are in turn orthogonal to $(0, 0, e^\top)^\top$. Without loss of generality, we assume that the rows of A are also orthogonal to $(0, 0, e^\top)^\top$ and those of B to $(e^\top, 0, 0)^\top$. Thus we generate $m-1$ vectors in the $(n-2)$ -dimensional subspace orthogonal to $(0, 0, e^\top)^\top$ and $(e^\top, 0, 0)^\top$ as the rows of A . The null space of A and $(0, 0, e^\top)^\top$ is then generically d -dimensional, contains the vector $(e^\top, 0, 0)^\top$, and is orthogonal to $(0, 0, e^\top)^\top$. It is not hard to see that there is a unique distribution (which we will call the OI distribution) on such subspaces that is invariant under orthogonal transformations that leave $(e^\top, 0, 0)^\top$ and $(0, 0, e^\top)^\top$ unchanged. (See also section 6.) In order to generate such a subspace, we have

Proposition 5.1. Suppose each row $a^\top = (a_1^\top, a_2^\top, a_3^\top)$ of A is generated independently as follows. Let $(\bar{a}_1^\top, \bar{a}_2^\top, \bar{a}_3^\top)^\top$ be a standard Gaussian random vector, and let $a_i = \bar{a}_i - (\bar{a}_i^\top e) e / e^\top e$ for $i = 1$ and 3 (here $e \in \mathbb{R}^k$ so that $e^\top e = k$ for $i = 1$, and $e \in \mathbb{R}^r$ so that $e^\top e = r$ for $i = 3$). Let $Q \in \mathcal{O}^{n \times n}$ (and hence also $Q^\top$) satisfy

$$Q \begin{bmatrix} e \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} e \\ 0 \\ 0 \end{bmatrix}, \quad Q \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix},$$

Then AQ has the same distribution as A , and if L is the null space of

$\begin{bmatrix} 0 & 0 & e^\top \\ A_1 & A_2 & A_3 \end{bmatrix}$, L has the OI distribution on d -subspaces containing $(e^\top, 0, 0)^\top$ and orthogonal to $(0, 0, e^\top)^\top$.

Proof. Let $(\hat{\bar{a}}_1^\top, \hat{a}_2^\top, \hat{\bar{a}}_3^\top) = (\bar{a}_1^\top, a_2^\top, \bar{a}_3^\top)Q$; then this vector is also a standard Gaussian random n -vector. Moreover, it is not hard to see that, if $(\hat{a}_1^\top, \hat{a}_2^\top, \hat{a}_3^\top)$ is obtained from this as is $(a_1^\top, a_2^\top, a_3^\top)$ from $(\bar{a}_1^\top, a_2^\top, \bar{a}_3^\top)$, then $(\hat{a}_1^\top, \hat{a}_2^\top, \hat{a}_3^\top) = (a_1^\top, a_2^\top, \bar{a}_3^\top)Q$. Hence each row of AQ is generated in the same way as a row of A from a standard Gaussian random vector, and these are all independent. Thus $AQ \stackrel{d}{=} A$. The last part follows easily, since, for an orthogonal matrix as in the lemma, QL is the null space of $\begin{bmatrix} 0 & 0 & e^\top \\ A_1 & A_2 & A_3 \end{bmatrix} Q^\top$, which has the same distribution as $\begin{bmatrix} 0 & 0 & e^\top \\ A_1 & A_2 & A_3 \end{bmatrix}$.

Model 2. Let ν_{mdkr} , $m+d=n$, $0 \leq k < m$, $0 \leq r < d$, be a family of distributions on $\mathbb{R}_+^n \times \mathbb{R}_+^n$ such that, if $(\hat{x}, \hat{s})$ is drawn from this distribution, then $\hat{x}_3 = 0$ and $\hat{s}_1 = 0$ with probability one, and $((\hat{s}_3, \hat{s}_2, 0), (0, \hat{x}_2, \hat{x}_1))$ has the distribution ν_{dmrk} , and the induced distribution $\mu_{n-k, d-r}$ on $(\hat{x}_2, \hat{s}_2)$ is permutation symmetric in the sense of model 1. Let $((\hat{x}_1, \hat{x}_2, 0), (0, \hat{s}_2, \hat{s}_3))$ be drawn from distribution ν_{mdkr} and let A be generated as in proposition 5.1. Then we say that (P_2) is distributed according to model 2.

It follows from the above and proposition 5.1 that, if the first and third sets of variables are interchanged, the dual (D_2) of a random linear programming problem distributed according to model 2 is also distributed according to model 2, with d and m and k and r interchanged.

In the rest of this section, (P_2) is a random linear programming problem distributed according to model 2. It is clear that x_3 must be zero in any feasible solution to (P_2) . Moreover, since $A_1 e = 0$ (the rows of A are orthogonal to $(e^\top, 0, 0)^\top$), if

$(\bar{x}_1^\top, \bar{x}_2^\top, \bar{x}_3^\top)^\top$ is feasible or optimal in (P_1) , so is $(\bar{x}_1^\top + \theta e, \bar{x}_2^\top, \bar{x}_3^\top)^\top$ for any $\theta \geq 0$. This justifies the nomenclature of unbounded variables for those in x_1 . In addition, it shows that an optimal solution to (P_2) can be obtained from one to

$$\begin{aligned}
 (P'_2) \quad & \min \quad \hat{s}_2^\top x_2 + \hat{s}_3^\top x_3 \\
 & e^\top x_3 = 0 \\
 & A_1 x_1 + A_2 x_2 + A_3 x_3 = b \\
 & x_2 \geq 0, x_3 \geq 0
 \end{aligned}$$

where x_1 is unrestricted in sign, since a suitable multiple of $(e^\top, 0, 0)^\top$ can be added to any optimal solution to (P'_2) to make it nonnegative.

Proposition 5.2. By deleting the x_3 -variables and eliminating (contracting) the x_1 -variables, (P'_2) is converted into a random $(m-k) \times (n-k-r)$ linear programming problem distributed according to model 1.

Proof. Let Π denote orthogonal projection onto the orthogonal complement of the column space of A_1 (which has dimension $k-1$ with probability one). Then $A_1 x_1 + A_2 x_2 = b := A_1 \hat{x}_1 + A_2 \hat{x}_2$ for some x_1 iff $\Pi A_2 x_2 = \Pi b = \Pi A_2 \hat{x}_2$. Thus (P'_2) is converted into

$$\begin{aligned}
 (P'_2') \quad & \min \quad \hat{s}_2^\top x_2 \\
 & \Pi A_2 x_2 = \Pi A_2 \hat{x}_2 \\
 & x_2 \geq 0.
 \end{aligned}$$

For any $Q' \in \mathcal{O}^{(n-k-r) \times (n-k-r)}$,

$$Q = \begin{bmatrix} I & 0 & 0 \\ 0 & Q' & 0 \\ 0 & 0 & I \end{bmatrix} \in \mathcal{O}^{n \times n}$$

leaves $(e^T, 0, 0)^T$ and $(0, 0, e^T)^T$ invariant. Hence by proposition 5.1, $\Pi A_2 Q'$ has the same distribution as ΠA_2 . Moreover the rows of A_2 are independent and the rank of ΠA_2 $m-k$ with probability one. Thus the null space of ΠA_2 is a random $(n-m-r)$ -subspace with the OI distribution, and hence (P_1') is a random $(m-k) \times (n-k-r)$ linear programming problem.

Thus some of the properties of the random linear programming problem (P_2) of this section can be deduced from those of the random linear programming problem (P_2') . Of course (P_2) is always unbounded, but some of its vertices correspond one-to-one with those of (P_2') (although there are other vertices with more than one of the components of x_1 zero). Because $(e^T, 0, 0)^T$ is in the null space of A , the choice $\hat{x}_1 = e$ implies that $(0, \hat{x}_2^T, 0)^T$ is also feasible; this accidental vanishing of all the components of x_1 shows that such a choice of $\hat{x}_1$ is perhaps unwise. A better choice is to have each component of $\hat{x}_1$ uniform in $[0, 1]$, or a positive normal random variable. In any case, the expected number of vertices of (P_2) exceeds that for (P_2') , which we calculated in section 3.

6. Random linear programming problems on the simplex.

In this final section, we consider generating and investigating random linear programming problems whose feasible region is a subset of the simplex of nonnegative vectors whose components sum to n . Our motivation comes from Karmarkar's algorithm [15]. However, we hasten to add that, even if the initial problem is generated from our

model, subsequent problems arising in the algorithm from a projective transformation of the original will not have the same distribution. Nevertheless we hope the results are of some interest.

As in Karmarkar's method, we assume e is feasible, so we take $\hat{x} = e$. Then our problem is

$$\begin{aligned} \min \quad & \hat{s}^T x \\ \text{subject to} \quad & Ax = Ae \\ & e^T x = n \\ & x \geq 0 \end{aligned}$$

for some $(m-1) \times n$ matrix A . The equality constraints require that $x - e$ lie in a subspace L , the null space of A and e^T , which lies in the subspace $\langle e \rangle^\perp$, the orthogonal complement of the linear span $\langle e \rangle$ of e .

Let the columns of $W \in \mathbb{R}^{n \times (n-1)}$ form an orthonormal basis for $\langle e \rangle^\perp$. Since $(e/\sqrt{n}, W)$ is orthogonal, $ee^T/n + WW^T = I$. Moreover, any orthogonal transformation of $\mathbb{R}^n$ that leaves $\langle e \rangle$ and $\langle e \rangle^\perp$ invariant is of the form $\epsilon ee^T/n + WZW^T$, where $\epsilon \in \mathcal{O}^{1 \times 1}$ ($\epsilon = \pm 1$) and $Z \in \mathcal{O}^{(n-1) \times (n-1)}$. It takes a vector $Wu \in \langle e \rangle^\perp$ into WZu . Thus a random d -subspace of $\langle e \rangle^\perp$ is invariant under such transformations iff it is of the form WM , where M is a random OI-distributed d -subspace of $\mathbb{R}^{n-1}$. We say a random d -subspace of $\langle e \rangle^\perp$ has the OI distribution if it is of this form.

By theorem 2.2, we can generate such a subspace as the range of WQ , where Q is a random $(n-1) \times d$ orthonormal matrix, or of WB , where B is a standard Gaussian random $(n-1) \times d$ matrix. In fact, it is not necessary to compute a suitable W . If $\bar{A}$ is a standard Gaussian random $n \times d$ matrix, then so is $V^T \bar{A}$, where $V = (e/\sqrt{n}, W)$ is orthogonal, and so $B = W^T \bar{A}$ is a standard Gaussian random $(n-1) \times d$ matrix. But WB

is then $WW^T\bar{A} = (I - ee^T/n)\bar{A}$, so that a random d -subspace L of $\langle e \rangle^\perp$ can be generated in this way from a standard Gaussian random matrix $\bar{A}$.

Alternatively, we can generate L as WM , where M is the null space of a standard Gaussian random $(m-1) \times (n-1)$ matrix B . Again, we can generate B as $\bar{A}W$, where $\bar{A}$ is a standard Gaussian random $(m-1) \times n$ matrix. Then

$$\begin{aligned} L &= \{Wv: \bar{A}Wv = 0\} \\ &= \{x: e^T x = 0, \bar{A}WW^T x = 0\} \\ &= \{x: e^T x = 0, \bar{A}(I - ee^T/n)x = 0\}. \end{aligned} \quad (6)$$

We have chosen to write L in terms of $A := \bar{A}(I - ee^T/n)$ instead of $\bar{A}$ because the rows of A (in contrast to those of $\bar{A}$) lie in $\langle e \rangle^\perp$ and span the orthogonal complement of L in $\langle e \rangle^\perp$. In addition, $Ae = 0$ which gives the canonical form for Karmarkar's algorithm.

(In some contexts it is more convenient to use $\bar{A}$ than A . Hence we record here that a random OI-distributed d -subspace L of $\langle e \rangle^\perp$ can be written as

$$L = \{x: e^T x = 0, \bar{A}x = 0\} \quad (7)$$

where $\bar{A}$ is a standard Gaussian random $(m-1) \times n$ matrix. This shows that the intersection of $\langle e \rangle^\perp$ with a random OI-distributed $(d+1)$ -subspace is a random OI-distributed d -subspace of $\langle e \rangle^\perp$.)

Model 3. We say that a random linear programming problem on the simplex is distributed according to model 3 if it is of the form

$$\begin{array}{ll}
(P_3) & \min \hat{s}^T x \\
& Ax = 0 \\
& e^T x = n \\
& x \geq 0
\end{array}$$

where $\hat{s}$ is drawn from some distribution on $\mathbb{R}_+^n$ and $A = \bar{A}(I - ee^T/n)$ for a standard Gaussian random $(m-1) \times n$ matrix $\bar{A}$ independent of $\hat{s}$. (Note the similarity in constructing A with the problems of section 5.)

The dual of (P_3) is

$$\begin{array}{ll}
(D_3) & \min e^T s \\
& Bs = B\hat{s} \\
& s \geq 0,
\end{array}$$

where the rows of B span the orthogonal complement of the row space of A and e^T . Hence the rows of B span a random OI-distributed d -subspace of $\langle e \rangle^\perp$, and B can be generated as $\bar{B}(I - ee^T/n)$ from a standard Gaussian random $d \times n$ matrix $\bar{B}$.

We no longer have symmetry between (P_3) and (D_3) . Moreover, while both problems are clearly always feasible, the feasible region of (P_3) is always bounded, while that of (D_3) is always unbounded, containing in particular the ray $\{\hat{s} + \theta e: \theta \geq 0\}$. This ray corresponds, in the traditional dual

$$\begin{array}{ll}
(D'_3) & \max \quad nz \\
& A^T y + ez \leq \hat{s},
\end{array}$$

to $z \rightarrow -\infty$ with y fixed.

Now we develop results parallel to those of section 3 for the random linear programming problem (P_3) distributed according to model 3. First, using the representations above, we see that (P_3) is equivalent to the random linear programming problem

$$(P_{3u}) \quad \begin{aligned} & \min (Q^T W^T \hat{s})^T u \\ & WQ u \geq -e, \end{aligned}$$

where Q is a random $(n-1) \times d$ orthonormal matrix with the OI distribution. Moreover, the mapping $u \rightarrow e + WQu$ is an isometry between the feasible region of (P_{3u}) and that of (P_3) . Similarly, (P_3) is equivalent to

$$(P_{3v}) \quad \begin{aligned} & \max (E^T \hat{s})^T v \\ & E v \leq e \end{aligned}$$

where $E = (I - ee^T/n)E$ for a standard Gaussian random $n \times d$ matrix E . Indeed, we can take $E = B^T$ with B as in (D_3) , and then (P_{3v}) is the conventional dual of (D_3) .

Next we compute the distance from $\hat{x} = e$ to a constraint hyperplane. This is the distance from $\hat{u} = 0$ to a constraint hyperplane in (P_{3u}) . Since each row of W has length $((n-1)/n)^{\frac{1}{2}}$, we can follow the proof of theorem 3.4 and conclude that this distance is distributed like $(n/(n-1))^{\frac{1}{2}} \beta^{-\frac{1}{2}}$, where β is a beta random variable with parameters $d/2$ and $(m-1)/2$. The distance is always at least $(n/(n-1))^{\frac{1}{2}}$ and has expectation $\approx \left[\frac{n(n-3)}{(n-1)(d-2)} \right]^{\frac{1}{2}} \approx \left[\frac{n}{d} \right]^{\frac{1}{2}}$ as $d \rightarrow \infty$ (and hence $n \rightarrow \infty$). Hence, if $m = O(d)$, the radius of the inscribed sphere around e to the feasible region of (P_2) is likely to be $O(1)$; this quantity is of interest in the analysis of Karmarkar's algorithm.

The distribution of the cosine γ of the angle between outward normals to two constraint hyperplanes can be calculated as in section 3.4, using $(W^T q_1)^T$ and $(W^T q_2)^T$ in place of q_1^T and q_2^T . The result is complicated and adds little insight. We merely note that $E\gamma < 0$; forcing boundedness of the feasible region appears to bias the angles between constraint hyperplanes.

Now we turn to the distribution of the components of a vertex $\bar{x}$ of the feasible region of (P_3) .

Theorem 6.1. Let $\mu_1, \dots, \mu_m$ be normal random variables with covariance matrix $I + ee^T/d$, and condition on the event that $\mu_i \geq 0$ for $i = 1, \dots, m$. Then a vertex of the feasible region of (P_3) has basic variables distributed like $n\mu_i/e^T \mu$, $i = 1, \dots, m$ (and nonbasic variables zero).

Proof. Here it is convenient to represent the feasible region using (7) as those x satisfying

$$\bar{A}x = \bar{A}e$$

$$e^T x = n$$

$$x \geq 0$$

where $\bar{A}$ is a standard Gaussian random $(m-1) \times n$ matrix. Again we use δ for $d^{\frac{1}{2}}$. A basic solution $\bar{x}$ corresponds to a basis consisting, say, of the first m columns of $[\bar{A}]$. Let $\bar{A} = [\bar{A}_\beta, \bar{A}_\nu]$ and $\bar{x}^T = [\bar{x}_\beta^T, 0]$ be partitions into the first m and last d columns. Then

$$(\delta^{-1} \bar{A}_\nu e)(-\delta) + \bar{A}_\beta (\bar{x}_\beta - e) = 0$$

$$\delta(-\delta) + e^T (\bar{x}_\beta - e) = 0.$$

Now $\bar{a}_0 := \delta^{-1} \bar{A}_\nu e$ is a standard Gaussian random vector independent of $\bar{a}_1, \dots, \bar{a}_m$, the columns of $\bar{A}_\beta$. Let $u := \bar{x}_\beta e$. Then $\begin{pmatrix} -\delta \\ u \end{pmatrix}$ lies in the null space of the standard Gaussian random $(m-1) \times (m+1)$ matrix $[\bar{a}_0, \bar{A}_\beta]$ and also in $\langle f \rangle^\perp$, where $f^\top := (\delta, e^\top)$. But the intersection of these two is a random OI-distributed 1-subspace of $\langle f \rangle^\perp$, and as in the case where $\langle f \rangle^\perp$ is replaced by $\langle e \rangle^\perp$, can be generated as $(I - ff^\top / f^\top f) \begin{pmatrix} \lambda_0 \\ \lambda \end{pmatrix}$, where $\begin{pmatrix} \lambda_0 \\ \lambda \end{pmatrix}$ is a standard Gaussian random vector in $\mathbb{R}^{1+m}$. Hence $\begin{pmatrix} -\delta \\ u \end{pmatrix}$ is proportional to

$$\begin{bmatrix} \lambda_0 - \frac{\delta \lambda_0 + e^\top \lambda}{n} \delta \\ \lambda - \frac{\delta \lambda_0 + e^\top \lambda}{n} e \end{bmatrix}.$$

Thus u_i is distributed like

$$-\delta \frac{n\lambda_i - \delta\lambda_0 - e^\top \lambda}{n\lambda_0 - \delta^2\lambda_0 - \delta e^\top \lambda} = \frac{d \delta^{-1}\lambda_0 + e^\top \lambda - n\lambda_i}{m\delta^{-1}\lambda_0 - e^\top \lambda}$$

and so $\bar{x}_i = 1 + u_i$ is distributed like

$$n \cdot \frac{\delta^{-1}\lambda_0 - \lambda_i}{m\delta^{-1}\lambda_0 - e^\top \lambda}.$$

Note that $e^\top \bar{x}_\beta = n$ as expected. To complete the proof, let $\mu_i = \delta^{-1}\lambda_0 - \lambda_i$, so that

$$\mu = (\delta^{-1}e, -I) \begin{bmatrix} \lambda_0 \\ \lambda \end{bmatrix}$$

and $\bar{x}_1$ is distributed like $n\mu_1/e^\top \mu$. Also, since $\begin{bmatrix} \lambda_0 \\ \lambda \end{bmatrix} \sim N(0, I)$,
 $\mu \sim N(0, (\delta^{-1}e, -I)(\delta^{-1}e, -I)^\top) = N(0, I + ee^\top/d)$. Finally, $\bar{x}$ is feasible iff $\mu \geq 0$ or $\mu \leq 0$.
 But, because of the symmetry of μ , we can as well condition on $\mu \geq 0$.

If $\bar{x}$ is as given in the theorem, then

$$\|\bar{x} - e\| = \left[\frac{n^2}{m} \cdot \frac{e^\top \mu^2 / m}{(e^\top \mu / m)^2} - m \right]^{\frac{1}{2}},$$

where μ^2 denotes $(\mu_1^2, \dots, \mu_m^2)^\top$. G. Samorodnitsky has proved (private communication) that this quantity is $O(\sqrt{n})$ (if $n = O(d)$ and $n = O(m)$) in probability, but the proof is rather technical because of the dependence among the μ_i 's and the conditioning, and we therefore omit it. This gives some indication of the good behavior of Karmarkar's algorithm in practice, since the worst-case bounds are based on the ratio of radii of circumscribing and inscribed balls. The above expressions suggests that a circumscribing ball of radius $O(\sqrt{n})$ may exist most of the time.

Finally, the proof of the theorem allows us to obtain an expression for the expectation of V_3 , the number of vertices of the feasible region of (P_3) . Since a basic solution is feasible if $\mu \geq 0$ or if $\mu \leq 0$, both of which have the same probability, we can condition on λ_0 to obtain

$$EV_3 = 2 \binom{n}{d} \int_{-\infty}^{\infty} [\Phi(\kappa/\delta)]^m \varphi(\kappa) d\kappa \quad (8)$$

(cf. (1) in section 3.6). Now the analysis of theorem 3.8 remains valid (noting that the integral from $-\infty$ to 0 in (8) is less than $[\Phi(0)]^m = (f_\alpha(0))^d$), so that EV_3 satisfies bounds as in (3). In particular, EV_3 grows like $(2.44)^d$, $(3.048)^d$, and $(1.952)^d$ in the case that $m = d$, $m = 2d$, and $m = \frac{1}{2}d$ respectively.

Acknowledgement. I am grateful to the referees for their helpful comments and especially to Gennady Samorodnitsky for his insights and his analysis of the distance to a vertex described in section 6.

References

1. I. Adler, "The expected number of pivots needed to solve parametric linear programs and the efficiency of the self-dual simplex method", manuscript, Department of Industrial Engineering and Operations Research, University of California, Berkeley, California (May 1983).
2. I. Adler and S.E. Berenguer, "Random linear programs", Operations Research Center Report No. 81-4, University of California, Berkeley, California (1981).
3. I. Adler and S.E. Berenguer, "Generating random linear programs," unpublished manuscript, Department of Industrial Engineering and Operations Research, University of California, Berkeley, California (1983).
4. I. Adler, R.M. Karp and R. Shamir, "A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps," Report UCB CSD 83/158, Computer Science Division, University of California, Berkeley, California (1983).
5. I. Adler and N. Megiddo, "A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension," Journal of the ACM 32 (1985), 871-895.
6. C. Blair, "Random linear programs with many variables and few constraints," Mathematical Programming 34 (1986), 62-71.
7. N. Bleistein and R.A. Handelsman, Asymptotic Expansions of Integrals, (Holt, Rinehart and Winston, New York, 1975).
8. K.-H. Borgwardt, "Some distribution-independent results about the asymptotic order of the average number of pivot steps of the simplex method," Mathematics of Operations Research 7 (1982), 441-462.
9. K.-H. Borgwardt, "The average number of pivot steps required by the simplex-method is polynomial", Zeitschrift fur Operations Research 26 (1982) 157-177.
10. K.-H. Borgwardt, The Simplex Method - A Probabilistic Approach, Springer-Verlag, Berlin-Heidelberg-New York, 1987.

11. K.-H. Borgwardt, "Probabilistic Analysis of the simplex method," in: Operations Research Proceedings 1987, 16th DGOR Meeting, pp. 564–576.
12. B. Efron, "The convex hull of a random set of points," Biometrika 52 (1965), 331–343.
13. M. Haimovich, "The simplex algorithm is very good! — On the expected number of pivot steps and related properties of random linear programs", manuscript, Columbia University, New York, New York (April 1983).
14. P.R. Halmos, Measure Theory, Van Nostrand, Princeton, New Jersey, 1950.
15. N. Karmarkar, "A new polynomial time algorithm for linear programming," Combinatorica 4 (1984), 373–395.
16. C.E. Lemke, "Bimatrix equilibrium points and mathematical programming," Management Science 11 (1965), 681–689.
17. T.M. Liebling, "On the number of iterations of the simplex method," Methods of Operations Research XVII (1972) 248–264.
18. J.H. May and R.L. Smith, "Random polytopes: their definition, generation, and aggregate properties", Mathematical Programming 24 (1982), 39–54.
19. P. McMullen, "The maximum number of faces of a convex polytope," Mathematika 17 (1970), Part 2, 179–184.
20. T.S. Motzkin, "The probability of solvability of linear inequalities", in: H.A. Antosiewicz, ed., Proceedings of the second symposium in linear programming (National Bureau of Standards and Directorate of Management Analysis, USAF, 1955) pp. 607–611.
21. A. Prekopa, "On the number of vertices of random convex polyhedra", Periodica Mathematica Hungarica 2 (1972), 259–282.
22. H. Raynaud, "Sur l'enveloppe convexe des nuages de points aleatoires dans R^n : I", Journal of Applied Probability 7 (1970), 35–48.
23. A Rényi and R. Sulanke, "Über die konvexe hülle von n zufällig gewählten punkten" I and II, Zeitschrift für Wahrscheinlichkeitstheorie 2 (1963), 75–84, and 3 (1964), 138–147.
24. J.B. Rosen and S. Suzuki, "Construction of nonlinear programming test problems," Communications of the ACM 8 (1965), 113.
25. B.K. Schmidt and T.H. Mattheiss, "The probability that a random polytope is bounded," Mathematics of Operations Research 2 (1977), 292–296.

26. W. Schmidt, "Some results in probabilistic geometry," Zeitschrift für Wahrscheinlichkeitstheorie 9 (1968), 158–162.
27. S. Smale, "On the average number of steps of the simplex method of linear programming", Mathematical Programming 27 (1983), 241–262.
28. S. Smale, "The problem of the average speed of the simplex method", in: A. Bachem, M. Grötschel and B. Korte, eds., Mathematical programming: the state of the art (Springer–Verlag, Berlin–Heidelberg–New York–Tokyo, 1983) pp. 530–539.
29. G.W. Stewart, "The efficient generation of random orthogonal matrices with an application to condition estimators," SIAM Journal on Numerical Analysis 17 (1980), 403–409.
30. M.J. Todd, "Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems," Mathematical Programming 35 (1986), 173–192.
31. M.J. Todd and Y. Ye, "A centered projective algorithm for linear programming," Technical Report 763, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1988.
32. W.B. Van Dam, J.B.G. Frenk and J. Telgen, "Randomly generated polytopes for testing mathematical programming algorithms," Mathematical Programming 26 (1983), 172–181.
33. A.M. Vershik and P.V. Sporyshev, "An estimate of the average number of steps in the simplex method, and problems in asymptotic integral geometry," Soviet Math. Dokl. 28 (1983), 195–199.
34. A.M. Vershik and P.V. Sporyshev, "An asymptotic estimate of the average number of steps of the parametric simplex method," USSR Comput. Maths. Math. Phys. 26 (1986), 104–113.
35. S.S. Wilks, Mathematical Statistics, John Wiley, New York, 1962.
36. R. Winder, "Partitions of n -space by hyperplanes," Journal of SIAM on Applied Mathematics 14 (1966), 811–818.