The expected number of pivots needed to solve parametric linear programs and the efficiency of the Self-Dual Simplex method

by

Ilan Adler

Department of Industrial Engineering
and Operations Research

University of California, Berkeley

Berkeley, California 94720

Abstract

Using a simple model to generate random linear programs we show that the expected number of pivots needed to solve a parametric linear program with d variables and n inequalities (where the parameterization is applied simultaneously to the objective and the right hand side) is bounded by $\min(d+1, n-d+1)$. These results are extended to single and multiple parametric linear programs in which either the objective function or the right hand side (or both) are parameterized. We also discuss the significance of our main result with respect to the average number of pivots needed to solve a linear program by the Self-Dual method.

Introduction

The riddle of the gap between the time proven actual efficiency of Dantzig's Simplex Method for linear programming and its apparent theoretical inefficiency (for specially constructed "bad" problems) attracted a great deal of research and interest in recent years. For the purpose of this discussion we shall consider linear programs having d variables and n inequalities (and define $m=n-d$); alternatively one can view such linear programs as having n non-negative variables satisfying m equations (and define $d=n-m$). Few variants of the Simplex method has been shown to require exponential number of steps (pivots) in m and d for specially constructed "bad" problems (see Klee and Minty [11], Jeroslow [10] Goldfarb and Sit [9], Avis and Chvatal [4]). On the other hand it has been observed and reported that, typically, the number of pivots for Phase two of the Simplex method is proportional to m (somewhere between $2m$ to $3m$, for example see Dantzig [8], page 160). Moreover, some authors generated randomly large numbers of linear programs and recorded the number of pivots taken to solve these problems by several variants of the Simplex method (see Kuhn and Quandt [12], Charnes et. al [7], Ravindran [14] and Avis and Chvatal [4]). These results verify that in randomly generated linear programs the number of pivots is proportional to m. The striking observation about these studies is that regardless of the method used to generate the linear programs and for all variants tried (including random-pivoting rule and even specially designed "bad" variants) the number of pivots in <u>all</u> cases was proportional to m (where the coefficients of proportionality vary from 2 for "good" variants to about 10 for "bad" variants).

More recently, some attempts were made to calculate the expected number of pivots taken by some variants of the Simplex method for randomly generated linear programs. Most notable in these attempts are the works of Smale [15] and Borgwardt [5] which analyse parametric versions of the Simplex method. In both models it has been shown that the number of pivots is bounded above by a number that is mainly affected by min(m,d), though the bound itself is still much larger than the observed average number of steps either in practice or in randomly generated linear programs. Our results in this paper were motivated primarily by two developments:

(i) The model for generating random linear programs which was presented first by Adler and Berenguer [1,2,3] provides a simple, natural model in which calculations of probabilities and expected values are easily obtained by simple combinatorial arguments.

(ii) The papers of Smale [15] and Borgwardt [5] demonstrate the attractiveness of analysing seldom mentioned parametric versions of the Simplex method (in particular the Self-Dual Simplex method of Dantzig [8], chapter 11). The point is that it is quite easy to check whether or not a particular basic sequence is part of a Simplex path <u>without</u> tracing the whole path whenever parametric variants are used. Thus, one can analyse the average length of a Simplex path by calculating the probability that a given basic sequence is in the path.

Unfortunatly we can not at this point analyse directly the Self-Dual Simplex method by the Adler-Berenguer model because the selection of starting right hand side and objective function <u>after</u> the problem has been generated violates the symmetry assumption of the model. Instead, our main result is a geometrical one that we believe is closely related to the average number of pivots taken by the Self-Dual Simplex method. Specifically we show that if a linear program of the form

$$P(\theta): \max\ (c + \theta c')x$$
$$\text{s.t.}\quad Ax \le b + \theta b' \qquad (\text{where } A \in R^{m\times d},\ b \in R^{m},\ c \in R^{d})$$
$$x \ge 0$$

is generated randomly, then the expected number of basic sequences which are optimal for at least one θ (given that the expected number is not zero) is less than or equal to $\min(m+1, d+1)$.

The relation between this result and the Self-Dual Simplex method is apparent since in this method (after selecting appropriate b', c' while solving for b and c) a basic sequence is part of the Simplex path only if it is optimal for $P(\theta)$ for some θ .

It is also interesting to note that there exists "bad" linear programs of the form of $P(\theta)$ in which the number of basic sequences which are optimal for at least one θ is exponential in m and d . (See Murty [13]).

So if nothing else, our analysis demonstrates that despite the existence of a polyhedral set with a long (exponential) path of basic sequence optimal for $P(\theta)$ from $\theta = -\infty$ to $\theta = \infty$, the average length of such a path is on the order of $\min(m+1, d+1)$. So similar gaps in the behavior of variants of the Simplex method should not be surprising.

We devote section 1 to the description of the model for generating random linear programs. In section 2 we present some preliminary results related to the model presented in section 1. In section 3 we present and prove our main result as described above. In sections 4 and 5 we show that our results can be straightfowardly applied to single and multiple parametric linear programs in which either the objective function or the right hand side or both are parametrized. In section 6 we summarize our findings in light of the particular model which is used, and make some general observations about this model and offer suggestions for further research.

Notation

Given a matrix $A \varepsilon R^{pxq}$ and sequences $I \subset \{1,\ldots,p\}$ and $J \subset \{1,\ldots,q\}$, we denote by $A_{I\cdot}$ the submatrix of A associated with the rows in I ; by $A_{\cdot J}$ the submatrix of A associated with the columns in J ; by A_{IJ} the submatrix of A associated with the rows in I and the columns in J . We also denote by $A_{i\cdot}$ the i^{th} row of A and by $A_{\cdot j}$ the j^{th} column of A .

We denote by Pr[E] the probability of the occurrence of event E .

We will also use the symbol ⊠ at the end of a proof of a theorem, lemma or corollary.

1. Generating Random Linear Programs.

In this section we shall present the basic model for generating linear programs which was first introduced by Adler-Berenguer [1,2].

Definition:

A matrix $A \in R^{m \times d}$ is said to be <u>full</u> if every $\ell \times \ell$ submatrix of A is of full rank (where $\ell = \min(m,d)$).

The model:

Select $\hat{A} \in R^{m \times d}$, $\hat{b} \in R^m$, $\hat{c} \in R^d$ such that $(\hat{A}, I, \hat{b})$ and $(\hat{A}^T, I, \hat{c})$ are full with probability 1. (These conditions assure that every $d(m)$ supporting hyperplanes of the constraints of the generated primal (dual) linear program will intersect at a point and that no more than $d(m)$ of the supporting hyperplanes will intersect at a point).

For every i $(i=1,\ldots,m)$ let

$$(\tilde{A}_{i\cdot}, b_i) = \begin{cases} (\hat{A}_{i\cdot}, \hat{b}_i) & \text{with probability } 1/2 \\ -(\hat{A}_{i\cdot}, \hat{b}_i) & \text{with probability } 1/2 \end{cases}$$

and for every j $(j=1,\ldots,d)$ let

$$(A_{\cdot j}, c_j) = \begin{cases} (\tilde{A}_{\cdot j}, \hat{c}_j) & \text{with probability } 1/2 \\ -(\tilde{A}_{\cdot j}, \hat{c}_j) & \text{with probability } 1/2. \end{cases}$$

Then generate the linear program:

$$P: \quad \max c^T x$$
$$\text{s.t. } Ax \leq b$$
$$x \geq 0$$

Let $n=m+d$, then there are 2^n different linear programs that can be generated from a given $\hat{A},\hat{b},\hat{c}$. We shall refer to each one of these programs as an <u>occurrence</u> of P. Note that one can generate equivalent 2^n linear programs by keeping $\hat{A},\hat{b},\hat{c}$ and flipping the n inequalities between

$\leq$ and $\geq$ with probability of 1/2. For details of this model and its relationship to other models of generating random linear programs and to other forms of linear programs we refer the reader to Adler-Berenguer [1,2,3].

An important aspect of this model is that we shall consider only results (that is, probabilities and expected values) which are <u>invariant</u> with respect to the choice of $\hat{A},\hat{b},\hat{c}$ and are functions of m and d only.

Another key property of the model is that the dual of P ,

$$\begin{aligned} D:\ & \min\ b^T y \\ & \text{s.t. } A^T y \geq c \\ & \qquad y \geq 0 \end{aligned}$$

is generated implicitly in the same way as P , thus all results with respect to P can be applied to D by replacing d with m .

Given this model one can, by simple combinatorial arguments, calculate probabilities for a variety of events (e.g. P is feasible or D has an optimal solution) and expected values of some random variables (e.g. the number of feasible basic solutions for P given it is feasible).

We shall quote few of these results in the next section; for a detailed discussion and analysis see Adler-Berenguer [1,2,3].

2. Preliminary Results

In this section we present two theorems that will play a major role in the development of our main results in section 3.

Theorem 2.1 (Adler-Berenguer [2,3])

Consider the following pair of primal-dual linear programs

P: max $c^T x$

s.t. $x \varepsilon \underline{\overline{X}}(b) = \{x | Ax \leq b, x > 0\}$

D: min $b^T y$

s.t. $y \varepsilon \underline{\overline{Y}}(c) = \{y | A^T y \geq c, y \geq 0\}$

Let us assume that A is full.

Then,

exactly one of the following occurs

(a) $\underline{\overline{X}}(0) \neq \{0\}$; $\underline{\overline{Y}}(0) = \{0\}$; and $\underline{\overline{X}}(b) \neq \emptyset$ for all $b \varepsilon R^m$

or:

(b) $\underline{\overline{X}}(0) = \{0\}$; $\underline{\overline{Y}}(0) \neq \{0\}$; and $\underline{\overline{Y}}(c) \neq \emptyset$ for all $c \varepsilon R^d$.

Theorem 2.1 is a stronger version of the strong duality theorem of linear programming for the special case in which (A,I) is full. Obviously this theorem applies to the model presented in section 1 since we assume that $(\hat{A}, I)$ is full.

Theorem 2.2 (Adler-Berenguer [1,2,3])

Under the assumption for the model presented in section 1

(2.1) Pr[P is feasible (i.e. $\underline{\overline{X}}(b) \neq \emptyset$)] $= \dfrac{\sum_{i=0}^{d} \binom{n}{i}}{2^n}$

Corollary 1 (Adler-Berenguer [2,3])

For the same model

$$\Pr[D \text{ is feasible (i.e. } Y(c) \neq \emptyset)] = \frac{\sum_{i=0}^{m} \binom{n}{i}}{2^n}$$

Theorem 2.2 (and other results of randomly generated linear programs) does not depend on the form of the linear program (for detailed discussion see

Adler-Berenguer [2,3]). In particular, we will need in later sections the following corollary to theorem 2.2.

Corollary 2 (Adler-Berenguer [2,3])

Consider the linear program

$$P': \max \ c^T x \quad \text{s.t.}\ Ax \le b$$

Suppose P' is generated analogously to P in section 1 from $\hat{A} \varepsilon R^{n \times d}$, $\hat{b} \varepsilon R^n$, $\hat{c} \varepsilon R^d$ where $(\hat{A}, \hat{b})$ is full, then

(2.2) $$\Pr[P' \text{ is feasible}] = \frac{\sum_{i=0}^{d} \binom{n}{i}}{2^n} .$$

3. Parametric Objective-R·H·S Linear Program

In this section we develop and present our main results about the expected number of extreme points which are optimal for a parametric objective-r·h·s linear program.

Let us generate randomly the following parametric (with respect to the objective function and the right hand side) linear program.

$$P(\theta):\quad \max\ (c+\theta c')x$$
$$\text{s.t. } Ax \leq b+\theta b'$$
$$x \geq 0$$

$P(\theta)$ is generated from $\hat{A}\varepsilon R^{mxd}$, $\hat{b},\hat{b}'\varepsilon R^{m}$, $\hat{c},\hat{c}'\varepsilon R^{d}$ in a procedure analogous to the one presented in section 1. (Here we assume that $(\hat{A},I,\hat{b},\hat{b}')$ and $(\hat{A}^T,I,\hat{c},\hat{c}')$ are full matrices).

Let us define as a basic sequence any sequence of indices $s = (J,K)$ composed of $J\subset\{1,\ldots,m\}$, $K\subset\{1,\ldots,d\}$ and such that $|s| = m$. We shall denote by $\bar{s} = (\bar{K},\bar{J})$ the sequence of indices composed of $\bar{J} = \{1,\ldots,m\} - J$ and $\bar{K} = \{1,\ldots,d\} - K$ (note that $|\bar{s}| = d$).

Given a basic sequence $s = (J,K)$ we denote

(3.1) $$\bar{b} = (A_{\bullet K}, I_{\bullet J})^{-1}b \;;\quad \bar{b}' = (A_{\bullet K}, I_{\bullet J})^{-1}b'$$
$$\bar{c} = (A^T_{\bullet\bar{J}}, -I_{\bullet\bar{K}})^{-1}c \;;\quad \bar{c}' = (A^T_{\bullet\bar{J}}, -I_{\bullet\bar{K}})^{-1}c'$$

We shall say that s is a primal feasible for $P(\theta)$ if $\bar{b}+\theta\bar{b}' \geq 0$ and that it is dual feasible for $P(\theta)$ if $\bar{c}+\theta\bar{c}' \geq 0$. (Note that s is primal (dual) feasible for $P(\theta)$ iff $\bar{s}$ is dual (primal) feasible for $D(\theta)$). Our main objective in this section is to investigate the expected cardinality of the set S defined by:

$S = \{$all basic sequences s which are optimal for $P(\theta)$ for at least one $\theta\}$.

Note that $s \varepsilon S$ iff s is primal and dual feasible for $P(\theta)$ for at least one θ.

Lemma 3.1

Given $P(\theta)$ which is generated randomly as above and a basic sequence s we have:

$$Pr[s \varepsilon S] = \frac{n+1}{2^n}$$

Proof:

Let $s = (J,K)$ and $\bar{b}$, $\bar{b}'$, $\bar{c}$, and $\bar{c}'$ be as defined above in (3.1). As we already observed, s is optimal for at least one θ iff there exists θ such that

$$(3.2) \quad \begin{pmatrix} \bar{b} \\ \bar{c} \end{pmatrix} + \theta \begin{pmatrix} \bar{b}' \\ \bar{c}' \end{pmatrix} \geq 0$$

Thus

$$(3.3) \quad Pr[s\varepsilon S] = \frac{\text{no of occurrences for which exists a } \theta \text{ satisfiying (3.2)}}{\text{no of occurrences}}$$

Note that (3.2) can be considered as a one-dimensional system of n inequalities.

$$(3.4) \quad -\begin{pmatrix} \bar{b}' \\ \bar{c}' \end{pmatrix} \theta \leq \begin{pmatrix} \bar{b} \\ \bar{c} \end{pmatrix}$$

Moreover, the matrix $\begin{pmatrix} -\bar{b}', & b \\ -\bar{c}' & c \end{pmatrix}$ is full and $(-\bar{b}_i', \bar{b}_i)$ (or $(-\bar{c}_j', \bar{c}_j)$) is as likely as $-(-\bar{b}_i', \bar{b}_i)$ (or $-(-\bar{c}_j', \bar{c}_j)$), $(i=1,\dots,m;\ j=1,\dots,d)$.

Hence (3.4) can be viewed as a randomly generated system of n inequalities in R^1 which is generated by a model similar to the one presented in section 1. So by corollary 2 to theorem 2.2, the system (3.4) has a solution with probability $\frac{\sum_{i=0}^{1} \binom{n}{i}}{2^n}$. Therefore

$$Pr[s\varepsilon S] = \frac{\sum_{i=0}^{1} \binom{n}{i}}{2^n} = \frac{n+1}{2^n}$$

☒

Theorem 3.1

$$E(|S|) = \binom{n}{d} \frac{n+1}{2^n}$$

Proof:

Let $I_s = \begin{cases} 1 & \text{if } s \varepsilon S \\ 0 & \text{if } s \notin S \end{cases}$

Then (by using Lemma 3.1)

$$E(|S|) = \sum_{s \varepsilon S} E(I_s) = \sum_{s \varepsilon S} \Pr[s \varepsilon S] = \binom{n}{d} \frac{n+1}{2^n}$$

⊠

Obviously $E(|S|)$ is not a good measure for the expected number of basic sequences in S since it includes occurrences in which $S = \emptyset$. Thus we shall be interested in $E(|S|/|S|>0)$.

Since $E(|S|) = E(|S|/|S| > 0) \Pr[|S| > 0] + E(|S|/|S| = 0) \Pr[|S| = 0]$ we get that $E(|S|/|S|>0) = E(|S|)/\Pr[|S| > 0]$. So we have to find $\Pr[|S| > 0]$ which is given in the next lemma.

Lemma 3.2

$$\Pr[|S| > 0] = \frac{\binom{n+2}{d+1} - \binom{n}{d}}{2^n}$$

Proof:

$\Pr[|S| > 0] = \Pr[P(\theta)$ and $D(\theta)$ are feasible for at least one θ] =

$$(3.5) \quad \frac{\text{no of occurrences for which } P(\theta) \text{ and } D(\theta) \text{ are feasible for at least one } \theta \text{ .}}{\text{no. of occurrences}}$$

But $P(\theta)$ is feasible for at least one θ if the system of n inequalities in R^{d+1}:

$$Ax - \theta b' \leq b$$

$$x \geq 0$$

has a solution.

Since $(A, -b', I, b)$ is full, $A_{\cdot j}$ is as likely as $-A_{\cdot j}$ and $(A_{i\cdot}, -b'_i, b_i)$ is as likely as $-(A_{i\cdot}, -b'_i, b_i)$, we get by theorem 2.2 that

(3.6) no. of occurrences for which $P(\theta)$ is feasible for at least one θ is $\sum_{i=0}^{d+1} \binom{n}{i}$

Similarly $D(\theta)$ is feasible for at least one θ if the system of n inequalities in R^{m+1}

$$A^T y - \theta c' \geq c$$

$$y \geq 0$$

has a solution. So,

(3.7) no. of occurrences for which $D(\theta)$ is feasible for at least one θ is $\sum_{i=0}^{m+1} \binom{n}{i}$

Let us define

N_P = no. of occurrences for which $P(\theta)$ is feasible for some θ and $D(\theta)$ is infeasible for all θ.

N_D = no. of occurrences for which $D(\theta)$ is feasible for some θ and $P(\theta)$ is infeasible for all θ.

N_{PD} = no. of occurrences for which $P(\theta)$ is feasible for some θ and $D(\theta)$ is feasible for some θ.

Now,

from (3.6) we get $N_P + N_{PD} = \sum_{i=0}^{d+1} \binom{n}{i}$

from (3.7) we get $N_D + N_{PD} = \sum_{i=0}^{m+1} \binom{n}{i}$

Thus

(3.8) $N_P + N_D + 2N_{PD} = \sum_{i=0}^{d+1} \binom{n}{i} + \sum_{i=0}^{m+1} \binom{n}{i}$

Since there are 2^n occurrences and by theorem 2.1 we get that

(3.9) $N_P + N_D + N_{PD} = 2^n$

But, by theorem 2.2 for every occurrence either $P(\theta)$ or $D(\theta)$ is feasible for <u>all</u> θ, thus N_{PD} = no. of occurrences for which $P(\theta)$ and $D(\theta)$ are <u>simultaneously</u> feasible for some θ.

Subtracting (3.9) from (3.8) we get

$$N_{PD} = \sum_{i=0}^{d+1} \binom{n}{i} + \sum_{i=0}^{m+1} \binom{n}{i} - 2^n = \binom{n}{d-1} + \binom{n}{d} + \binom{n}{d+1} \quad \text{(since } d=n-m\text{)}$$

$$= \binom{n+2}{d+1} - \binom{n}{d} \quad \text{(by using the identity } \binom{k}{\ell} + \binom{k}{\ell-1} = \binom{k+1}{\ell}\text{)} .$$

Hence from (3.5) we have

$$\Pr[|S| > 0] = \frac{\binom{n+2}{d+1} - \binom{n}{d}}{2^n}$$

☒

Combining theorem 3.1 and lemma 3.2 we get our main result:

<u>Theorem 3.2</u>

$$E(|S|/|S| > 0) = \frac{(n+1)(d+1)(m+1)}{(n+2)(n+1) - (d+1)(m+1)}$$

<u>Proof</u>:

$$E(|S|/|S| > 0) = \frac{E(|S|)}{\Pr[|S| > 0]} = \binom{n}{d} \frac{n+1}{2^n} \Big/ \frac{\binom{n+2}{d+1} - \binom{n}{d}}{2^n} =$$

$$= \frac{\binom{n}{d}(n+1)}{\binom{n}{d}\left(\frac{(n+2)(n+1)}{(d+1)(m+1)} - 1\right)} = \frac{(n+1)(d+1)(m+1)}{(n+2)(n+1) - (d+1)(m+1)}$$

☒

<u>Corollary 1</u>

$$\frac{2}{3} \min(d+1, m+1) \le E(|S|/|S| > 0) \le \min(d+1, m+1)$$

Corollary 2

$$\lim_{\substack{n\to\infty \\ m \text{ is fixed}}} E(|S|/|S|>0) = m+1$$

$$\lim_{\substack{h\to\infty \\ d \text{ is fixed}}} E(|S|/|S|>0) = d+1$$

Corollary 3

Let $\alpha = \frac{\min(m,d)}{n}$ So $(1-\alpha) = \frac{\max(m,d)}{n}$ then for large n

$$E(|S|/|S|>0) \approx \frac{(1-\alpha)\alpha}{1-\alpha(1-\alpha)} n = \frac{1-\alpha}{1-\alpha(1-\alpha)} \min(m,d)$$

Theorem 3.2 and its corollaries indicate that if the parametric objective-r•h•s linear program $P(\theta)$ is generated randomly then the expected number of basic sequences which are optimal for $P(\theta)$ for at least one θ (provided that there exists at least one such basic sequence) is quite small, especially whenever d (or m) is fixed and n goes to ∞ .

The list of all the optimal basic sequence for $P(\theta)$ starting at $\theta = -\infty$ and going through $\theta = 0$ and then to $\theta = \infty$ forms a path of adjacent basic solutions (i.e. each basic sequence can be obtained from its predecessor by one pivot).

Thus the results of theorem 3.2 and its corollaries can be viewed as the expected number of pivots in a parametric path for $P(\theta)$ as described above (given the path is not empty).

Hence we established that if one generates $P(\theta)$ randomly (according to our model) and a parametric (cost and r•h•s) analysis is performed, the expected number of pivots will be in the order of min(m+1,d+1) (given that there exists at least one Θ for which $P(\Theta)$ has an optimal solution).
We will show in the next section that similar results hold for the parametric r•h•s or parametric objective function linear program. More intriguing

is the relevance of theorem 3.2 to the efficiency of the Simplex method. We will offer here two interpretations to this result that may throw some light over the observed efficiency of the Simplex method.

(a) Consider the linear program

$$P: \max c^T x$$
$$\text{s.t. } Ax \leq b \quad (\text{where } A \in R^{m \times d},\ b \in R^m,\ c \in R^d)$$
$$x \geq 0$$

The Self-Dual method of Dantzig ([8], Chapter 11) proceeds by defining $b' = e_m$, $c' = e_d$ (where e_k is a vector of k 1's) and then solving $P(\theta)$ (as defined in the beginning of this section) starting at $\theta = \infty$ and moving towards $\theta = 0$. Obviously the number of pivots taken by this algorithm is bounded above by the number of basic sequences which are optimal for at least one θ.

Note, that if we want to find the expected number of pivots taken by the Self-Dual Simplex method we have to generate randomly A,b,c while fixing b', c'. This procedure however, differs from our procedure in which b', c' are generated in the same symmetrical way as A, b and c. Thus, Theorem 3.2 cannot be directly used to obtain the expected number of pivots taken by the Self-Dual Simplex method. However, we believe that the similarity of the two models strongly indicates that the expected number of pivots taken by the Self-Dual Simplex method is bounded by a linear or polynomial function of $\min(m,d)$.

(b) Any "bad" problem with respect to the Self-Dual Simplex method (that is, a problem that takes exponential number of pivots to solve) is also an occurrence in which the number of basic sequences which are optimal for at least one θ is exponential. Thus, theorem 3.2 demonsrates that despite the existence of "bad" case for problem $P(\theta)$ (see Murty [13]), the average is quite small, so it should not be surprising if the same relationship holds between a "bad" and an average problem solved by the Self-Dual Simplex method.

4. Parametric Objective Function Linear Programs and Parametric R.H.S Linear Programs

In this section we shall show how the approach which was developed in section 3 can be applied to two types of parametric linear programs (i) parametric objective function and (ii) parametric r.h.s.

Since the applications of the method presented in section 3 are straight forward we shall present only short outlines of the proofs.

Consider the following parametric objective function linear program

$$P(\theta): \max(c + \theta c')x$$
$$\text{s.t. } Ax \leq b$$
$$x \geq 0$$

where $P(\theta)$ is generated from $\hat{A} \varepsilon R^{mxd}$, $\hat{b} \varepsilon R^m$, $\hat{c}, \hat{c}' \varepsilon R^d$ similarly to $P(\theta)$ in section 3.

As in section 3 we define

S = {set of all basic sequences which are optimal for $P(\theta)$ for at least one θ }

Lemma 4.1

Given $P(\theta)$ which is generated randomly as above and a basic sequence s we have $Pr[s \varepsilon S] = \frac{d+1}{2^n}$.

Proof:

The proof is very similar to the proof to Lemma 3.1 except that we look for θ such that $\bar{c} + \theta \bar{c}' \geq 0$. Since $\bar{c}$ and $\bar{c}'$ are d-vectors we get $d+1$ instead of $n+1$ in the expression for $Pr[s \varepsilon S]$.

☒

Theorem 4.1

$$E(|S|) = \binom{n}{d} \frac{d+1}{2^n}$$

Proof:

See the proof to theorem 3.1.

☒

Lemma 4.2

$$\Pr[|S| > 0] = \frac{\binom{n+1}{d}}{2^n}$$

Proof:

Let us follow the proof of Lemma 3.2. In our case:

$$N_P + N_{PD} = \sum_{i=0}^{d} \binom{n}{i} \quad \text{(probability of } \begin{matrix} Ax \leq b \\ x \geq 0 \end{matrix} \text{ being feasible)}$$

$$N_D + N_{PD} = \sum_{i=0}^{m+1} \binom{n}{i} \quad \text{(probability of } \left\{ \begin{matrix} A^T y \geq c + \theta c' \\ y \geq 0 \end{matrix} \right\} \text{ being feasible)}$$

Thus:

$$N_{PD} = \sum_{i=0}^{d} \binom{n}{i} + \sum_{i=0}^{m+1} \binom{n}{i} - 2^n = \binom{n}{d-1} + \binom{n}{d} = \binom{n+1}{d}$$

So:

$$\Pr[|S| > 0] = \frac{\binom{n+1}{d}}{2^n}$$

☒

Theorem 4.2

$$(4.1) \quad E(|S|/|S| > 0) = \frac{(d+1)(m+1)}{n+1}$$

Proof:

Following the same lines as in the proof of theorem 3.2

$$E(|S|/|S| > 0) = \frac{E(|S|)}{\Pr[|S| > 0]} = \frac{\binom{n}{d}(d+1)}{\binom{n+1}{d}} = \frac{(d+1)(m+1)}{n+1}$$

☒

The results of theorem 4.2 obviously establish an upper bound to the expected number of pivot steps (or ranges of optimality) which are encountered by performing parametric analysis on a linear program with objective function c and parametric objective function c' (where both are selected randomly and when we condition on having an optimal solution for at least one θ).

Since the expression (4.1) is symmetric in d and m , it is obvious that if we consider the parametric r•h•s linear program,

$$P(\theta): \quad \max\ c^T x$$
$$\text{s.t. } Ax \leq b + \theta b'$$
$$x \geq 0$$

and define S as in the beginning of the section we get that all the results in this section are applied also to $P(\theta)$ above.

In particular it shows that the average number of steps of a parametric r•h•s linear program is the same as that of a parametric objective function linear program.

Corollary 1

$$\frac{1}{2}\min(d+1,m+1) \leq E(|S|/|S|>0) \leq \min(d+1,m+1)$$

Corollary 2

(i) $\lim\limits_{\substack{n\to\infty \\ m \text{ fixed}}} E(|S|/|S|>0) = m+1$

(ii) $\lim\limits_{\substack{n\to\infty \\ d \text{ fixed}}} E(|S|/|S|>0) = d+1$

Corollary 3

Let $\alpha = \dfrac{\min(m,d)}{n}$ (so $(1-\alpha) = \dfrac{\max(m,d)}{n}$) then for large n

$$E(|S|/|S|>0) \approx (1-\alpha)\,\alpha n = (1-\alpha)\min(m,d)$$

A comparison between the results in this section and the previous one (in particular theorems 3.2 and 4.2) shows that the average number of steps of simultaneous parametization of the objective and r•h•s is almost the same as the parametization of only the objective function (or r•h•s) and exactly the same whenever $n \to \infty$ (while d or m are fixed).

This observation is not surprising considering the following intuitive argument. Let

$$P(\theta,\delta): \quad \max\ (c + \theta c')^T x$$

$$\text{s.t. } Ax \leq b + \delta b' \quad (\text{where } A \in R^{m \times d},\ b \in R^m,\ c \in R^d)$$

Consider the partitioning of the θ,δ space to rectangles which represent the regions of θ,δ over which a basic sequence is optimal for $P(\theta,\delta)$ (or regions in which either $P(\theta,\delta)$ or its dual are infeasible).

Thus, in theorem 4.2 we get that the expected number of such rectangles (given that there exists at least one optimal basic sequence) that the θ axis goes through is equal to $\min(m+1,d+1)$. Similarly if we apply theorem 4.2 to the r•h•s parametric case we get that the expected number of rectangles that the δ axis goes through is also equal to $\min(m+1,d+1)$.

On the other hand, theorem 3.2 gives the expected number of rectangles that the line $\theta = \delta$ goes through; but since going through either axis gives the same expected number of rectangles it can be argued that going along any other line through the (θ,δ) space (including $\theta = \delta$), gives about the same expected number of rectangles. Note that the comparison between theorems 3.2 and 4.2 (and the above argument) suggests the superiority of using the Self-Dual Simplex method over a two-phase parametric algorithm in which a primal feasible solution is obtained by parameterizing the r•h•s and then an optimal solution is obtained by parameterizing the objective function.

5. Multiparametric Linear Programs.

It is a relatively easy task to extend the results of sections 3 and 4 to the following three multiparametric linear programs:

$$P_{cb}(\theta_1,\ldots,\theta_k): \quad \max(c + \sum_{\ell=1}^{k} \theta_\ell c^\ell)^T x$$

$$\text{s.t. } Ax \leq b + \sum_{\ell=1}^{k} \theta_\ell b^\ell \qquad (k \leq \min(m,d))$$

$$x \geq 0$$

$$P_c(\theta_1,\ldots,\theta_k): \quad \max(c + \sum_{\ell=1}^{k} \theta_\ell c^\ell)^T x \qquad (k \leq d)$$

$$\text{s.t. } Ax \leq b$$

$$x \geq 0$$

$$P_b(\theta_1,\ldots,\theta_k): \quad \max \; c^T x \qquad (k \leq m)$$

$$\text{s.t. } Ax \leq b + \sum_{\ell=1}^{k} \theta_\ell b^\ell$$

$$x \geq 0$$

All three models are generated from $\hat{A} \varepsilon R^{mxd}$, $\hat{b}, \hat{b}^1, \ldots, \hat{b}^k \varepsilon R^m$, $\hat{c}, \hat{c}^1, \ldots, \hat{c}^k \varepsilon R^d$ analogously to $P(\theta)$ as presented in sections 3 and 4.

For all cases we define S as before by

S_{cb} = {set of all basic sequences for which there exists an optimal solution for $P_{cb}(\theta)$ for at least one $(\theta_1,\ldots,\theta_k)$}.

S_c = {set of all basic sequences for which there exists an optimal solution for $P_c(\theta)$ for at least one $(\theta_1,\ldots,\theta_k)$} .

S_b = {set of all basic sequences for which there exists an optimal solution for $P_b(\theta)$ for at least one $(\theta_1,\ldots,\theta_k)$} .

Again we can use the techniques of sections 3 and 4 (which are presented in detail in section 3) except that we have to modify the dimensionality of the spaces in which we work.

Lemma 5.1

(i) $$\Pr[s \varepsilon S_{cb}] = \frac{\sum_{i=0}^{k} \binom{n}{i}}{2^n}$$

(ii) $$\Pr[s \varepsilon S_{c}] = \frac{\sum_{i=0}^{k} \binom{d}{i}}{2^n}$$

(iii) $$\Pr[s \varepsilon S_{b}] = \frac{\sum_{i=0}^{k} \binom{m}{i}}{2^n}$$

Theorem 5.1

(i) $$E(|S_{cb}|) = \binom{n}{d} \frac{\sum_{i=0}^{k} \binom{n}{i}}{2^n}$$

(ii) $$E(|S_{c}|) = \binom{n}{d} \frac{\sum_{i=0}^{k} \binom{d}{i}}{2^n}$$

(iii) $$E(|S_{b}|) = \binom{n}{d} \frac{\sum_{i=0}^{k} \binom{m}{i}}{2^n}$$

Lemma 5.2

(i) $$\Pr[|S_{cb}| > 0] = \frac{\sum_{i=d-k}^{d+k} \binom{n}{i}}{2^n}$$

(ii) $$\Pr[|S_{c}| > 0] = \frac{\sum_{i=d-k}^{d} \binom{n}{i}}{2^n}$$

(iii) $$\Pr[|S_{b}| > 0] = \frac{\sum_{i=m-k}^{m} \binom{n}{i}}{2^n}$$

Theorem 5.2

(i) $$E(|S_{cb}| / |S_{cb}| > 0) = \frac{\binom{n}{d} \sum_{i=1}^{k} \binom{n}{i}}{\sum_{i=d-k}^{d+k} \binom{n}{i}}$$

$$\text{(ii)}\quad E(|S_c|/|S_c| > 0) = \frac{\binom{n}{d}\sum_{i=0}^{k}\binom{d}{i}}{\sum_{i=d-k}^{d}\binom{n}{i}}$$

$$\text{(iii)}\quad E(|S_b|/|S_b| > 0) = \frac{\binom{n}{d}\sum_{i=0}^{k}\binom{m}{i}}{\sum_{i=m-k}^{m}\binom{n}{i}}$$

6. Concluding Remarks

(a) As was mentioned in the introduction and in the discussion following theorem 3.2, we have so far failed to get direct results for the expected number of steps of the Self-Dual Simplex method. Trying to overcome this difficulty seems to be a rewarding endeavor.

(b) Any result related to the expected number of steps of a variant of the Simplex method requires a careful analysis of the model used to generate the random linear programs.

In our own case, it is useful to quote the following theorem from Adler and Berenguer [2,3].

Theorem 6.1

Let P be a linear program generated randomly as described in section 1, then

(i) $$\Pr[P \text{ is infeasible}] = \frac{\sum_{i=0}^{m-1} \binom{n}{i}}{2^n}$$

(ii) $$\Pr[P \text{ is unbounded}] = \frac{\sum_{i=0}^{d-1} \binom{n}{i}}{2^n}$$

(iii) $$\Pr[P \text{ is optimal}] = \frac{\binom{n}{d}}{2^n}$$

Thus in our model, the probability for a linear program to be optimal (i.e. to have an optimal solution) is very small for large n .

In particular, if d (or m) are fixed and $n \to \infty$ then the linear program is practically infeasible (unbounded). So, a result about the average speed of the Simplex method is actually a result about the average speed to find that a linear program is infeasible (unbounded). Since finding that a linear program is unbounded (or infeasible) is easier than finding an optimal solution (because there are many more terminating basic sequences), an effort should be made to consider models which generate only linear programs that have optimal solutions.

(c) Even restricting our model to linear programs with optimal solution, it is still far from representing linear programs which are encountered in practical applications. The assumptions on general position and non-degeneracy can be justified by observing that a small perturbation of the constraint of any linear program would lead to their satisfaction. However these assumptions leave out a lot of important classes of problems (such as transportation, assignment etc.). More disturbing is the symmetry assumption which does not seem to be relevant to practical problems. The challenge here is to consider models for randomization closer to reality but which are simple enough to allow for theoretical analysis.

(d) Getting a good result for one variant of the Simplex method, leaves still unanswered the question of the small average number of steps of virtually any variant of the Simplex method. As was noted in the introduction, parametric versions of the Simplex method are easier to analyse. The challenge then, is to extend the analysis to non-parametric versions (in particular to the oldest of all variants, the one that pivots on the column of the most negative reduced cost).

(e) As was discussed in Adler and Berenguer [1,3], it is very difficult to get the variances for the random variables whose expected values are so easily obtained. Nonetheless, it is a problem to be considered, given the very small observed variance of the number of steps for all known variants of the Simplex method.

Some progress with respect to the remarks above has already been made and we hope to report on it in the near future.

References

[1] I. Adler and S. E. Berenguer, "Random linear programs", Operations Research Center Report No. 81-4, University of California, Berkeley (1981).

[2] I. Adler and S. E. Berenguer, "Duality theory and the random generation of linear programs", Manuscript, Department of Industrial Engineering and Operations Research, University of California, Berkeley (1981).

[3] I. Adler and S. E. Berenguer, "Generating random linear programs", Revised manuscript, Department of Industrial Engineering and Operations Research, Berkeley (1983); submitted to Mathematical Programming.

[4] D. Avis and V. Chvâtal, "Notes on Bland's pivoting rule", Mathematical Programming Study 8 (1978) 24-34.

[5] K. Borgwardt, "Some distribution-independent results about the asymptotic order of the average number of steps of the Simplex method", Mathematics of Operations Research 7 (1982) 441-462.

[6] R. C. Buck, "Partition of Space", The American Mathematical Monthly 50 (1943) 541-544.

[7] A. Charnes, W. Raike, J. D. Stutz and A. S. Walters, "On generation of test problems for linear programming codes", Communications of the ACM 17 (1974) 583-586.

[8] G. B. Dantzig, Linear Programming and Extensions, (Princeton University Press, Princeton, New Jersey 1963).

[9] D. Goldfarb and W. Y. Sit, "Worst case behavior of the steepest edge Simplex method", Discrete Applied Math 1 (1979) 277-285.

[10] R. G. Jeroslow, "The Simplex algorithm with the pivot rule of maximizing improvement criterion", Discrete Mathematics 4 (1973) 367-377.

[11] V. Klee and G. J. Minty, "How good is the Simplex algorithm?", in: O. Shisha, ed., Proceedings of the Third Symposium on Inequalities (Academic Press, New York, 1972) pp. 159-175.

[12] H. W. Kuhn and R. E. Quandt, "An experimental study of the Simplex method", in: N.C. Metropolis, A. H. Taub, J. Todd and C. B. Thompkins, eds, Experimental arithmetic, high speed computing and mathematics, Proceedings of the 15th Symposium in Applied Mathematics of the AMS (American Mathematical Society, Providence, RI, 1963) pp. 107-124.

[13] K. G. Murty, "Computational complexity of parametric linear programming", Mathematical Programming 19 (1980) 213-219.

[14] A. Ravindran, "A comparison of the primal-Simplex and complementary pivot methods for linear programming", Naval Research Logistic Quarterly 20 (1973) 95-100.

[15] S. Smale, "On the average speed of the Simplex method of linear programming", Manuscript, Department of Mathematics, University of California, Berkeley (1982).