Annals of Operations Research 62(1996)29–57

An infeasible-start algorithm for linear programming whose complexity depends on the distance from the starting point to the optimal solution*

Robert M. Freund
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

This paper presents an algorithm for solving a linear program LP (to a given tolerance) from a given prespecified starting point. As such, the algorithm does not depend on any "big $\mathcal{M}$" initialization assumption. The complexity of the algorithm is sensitive to and is dependent on the quality of the starting point, as assessed by suitable measures of the extent of infeasibility and the extent of nonoptimality of the starting point. Two new measures of the extent of infeasibility and of nonoptimality of a starting point are developed. We then present an algorithm for solving LP whose complexity depends explicitly and only on how close the starting point is to the set of LP feasible and optimal solutions (using these and other standard measures), and also on n (the number of inequalities). The complexity results using these measures of infeasibility and nonoptimality appear to be consistent with the observed practical sensitivity of interior-point algorithms to certain types of starting points. The starting point can be any pair of primal and dual vectors that may or may not be primal and/or dual feasible, and that satisfies a simple condition that typically arises in practice or is easy to coerce.

Keywords: Linear program, interior-point, barrier function, Newton method, polynomial-time.

1. Introduction

Consider a linear program P and its dual D given by:

P:
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

*Research supported in part by the MIT-NTU Collaboration Agreement.

whose dual is given by:

$$\text{D:} \qquad \begin{array}{ll} \underset{y,s}{\text{maximize}} & b^T y \\ \text{subject to} & A^T y + s = c, \\ & s \geq 0. \end{array}$$

Instead of developing algorithms and analyzing the complexity of algorithms for solving P that depend only on the data triplet (A, b, c), this paper includes a starting pair of primal and dual points $(\hat{x}, \hat{y}, \hat{s})$ as the data for the problem and for an algorithm for solving P. The starting point $(\hat{x}, \hat{y}, \hat{s})$ can be any pair of primal and dual vectors that may or may not be primal and/or dual feasible. This paper considers the question of developing an algorithm for solving P that accepts as input the array $(A, b, c, \hat{x}, \hat{y}, \hat{s})$. The goal of the paper has been to simultaneously develop measures of the quality of the starting point $(\hat{x}, \hat{y}, \hat{s})$ by assessing the extent of infeasibility and nonoptimality of the starting $(\hat{x}, \hat{y}, \hat{s})$ in an appropriate way, and to develop an algorithm for solving P that is sensitive to and is dependent on the quality of the starting point, as assessed by these appropriate measures of the extent of infeasibility and the extent of nonoptimality of the starting point.

Intuitively, if the starting point $(\hat{x}, \hat{y}, \hat{s})$ is closer to the sets of feasible and optimal solutions of the primal and the dual in some appropriate measures, a well-designed algorithm should be able to solve the given LP in fewer iterations. This intuition is made precise with the introduction of two new measures of the extent of infeasibility and of nonoptimality of a starting point. We then present an algorithm for solving P whose complexity depends explicitly and only on how close the starting point is to the set of feasible and optimal solutions of P, and also on n (the number of inequalities). The complexity results using these measures of infeasibility and nonoptimality appear to be consistent with the practical sensitivity of interior-point algorithms to certain types of starting points.

This paper is part of the general research on interior-point methods for linear programming stemming from Karmarkar's seminal work [10]. It also falls into the category of "infeasible-start" combined Phase-I and Phase-II methods for solving a linear program, and also is part of the general research on solving linear programming problems from "warm" or "hot" starts.

The combined Phase-I and Phase-II methods for solving a linear program have been motivated by a desire to remove the explicit use of a "big-$\mathcal{M}$" scalar in the formulation of the linear program, and to take advantage of the geometry and mathematics of interior-point methods to simultaneously improve upon the feasibility and optimality of algorithm iterates, see for example de Gellinck and Vial [8], Anstreicher [1,2], Todd [25,26], and Todd and Wang [27]. However, these algorithms were developed to be initiated at an arbitrary or "cold" start. More recently, Anstreicher [3], Vial [29], and Ye et al. [30] have developed other attractive methods

for solving the combined Phase-I and Phase-II problem, all analyzed from the vantage point of an arbitrary cold start.

The use of "warm" or "hot" starting points into the design and analysis of interior-point algorithms for linear programming is motivated by the desire to explicitly take advantage of the quality of the starting point. Polyak [20], Gill et al. [9], and also [4–7] have used the tools of shifted barrier functions and other parameterization methods to study the use of "warm" start algorithms for solving linear programs, see also Mitchell [15].

Motivated by the success of the OB1 code in solving NETLIB problems by a method that combines Phase-I and Phase-II [12,14], many researchers have studied convergence and theoretically efficient variants of the OB1 interior-point methodological strategy, see for example Kojima et al. [11], Mizuno [16], and Mizuno et al. [17], Potra [21–23], and Zhang [31,32]. The algorithm developed in this paper is motivated by the results of many of these latter papers, but most particularly Mizuno [16].

The mathematical development in the paper begins in section 2, which presents measures of closeness of the starting point $(\hat{x}, \hat{y}, \hat{s})$ to the feasible regions of P and D, and to the optimal regions of P and D. The analysis assumes that the starting point can be any pair of primal and dual vectors $(\hat{x}, \hat{y}, \hat{s})$ that may or may not be primal and/or dual feasible, and that satisfies a simple condition that typically arises in practice or is easy to coerce. In addition to typical measures of infeasibility such as $\|b - A\hat{x}\|$ and $\|c - A^T\hat{y} - \hat{s}\|$ and nonoptimality such as $\hat{x}^T\hat{s}$, we also present one new measure of infeasibility referred to as δ_1, and one new measure of nonoptimality referred to as δ_2. Elementary properties of these measures and limiting properties of these measures as $(\hat{x}, \hat{y}, \hat{s})$ approach a feasible and/or optimal point are analyzed as well.

In section 3, we introduce a parameterized infeasible-start barrier problem for P that is a variation of a standard parameterized infeasible-start barrier problem but instead uses two parameters ε and μ, as shown below:

$$\begin{aligned} \text{BP}(\varepsilon, \mu): \qquad & \underset{x}{\text{minimize}} \quad (c + \varepsilon[A^T\hat{y} + \hat{s} - c])^T x - \varepsilon\mu \sum_{j=1}^{n} \ln(x_j) \\ & \text{subject to} \quad Ax = b + \varepsilon[A\hat{x} - b], \\ & \qquad\qquad x > 0. \end{aligned}$$

This two-parameter parameterization bears a close resemblance in particular with the recent analysis in Mizuno et al. [18], but the choice of the parameters is slightly different and allows for a monotone change in the parameters ε and μ as the underlying problem is solved.

In section 4, we present an algorithm for solving P by tracing the path of (approximate) solutions of BP(ε, μ) as the parameter ε is decreased and as the

parameter μ is increased, in such a way that the value of ε goes to zero and the value of $\varepsilon\mu$ goes to zero, and that bounds the maximum value that the parameter μ can achieve. The complexity of the algorithm is analyzed in section 5. In theorems 5.1, 5.2, and 5.3, respectively, we present bounds on the maximum number of iterations needed to find an ε-approximate feasible solution of P, an ε-approximate feasible solution of D, and an ε-approximate optimal solution of P and D, that depend only on ε, $\|b - A\hat{x}\|$, $\|c - A^T\hat{y} - \hat{s}\|$, $\hat{x}^T\hat{s}$, δ_1, δ_2, and on n (the number of inequalities). These results bound the complexity of the infeasible-start interior-point method in terms of the quality of the starting point. These complexity results specialize to a complexity of $O(nL)$ iterations when the starting point is infeasible in the primal and/or the dual and the measures of closeness δ_1 and δ_2 are bounded by a constant independent of n, and specialize to a complexity of $O(\sqrt{n}L)$ iterations in the case when the starting point is feasible in the primal and in the dual.

(In [23], Potra develops an algorithm and complexity bounds that also depend on the quality of the starting point; however, it is not apparent how the results in [23] are to be compared to the results contained herein.)

Sections 6 and 7 contain proofs of the necessary mathematical properties of the algorithm of section 4. In section 8, we discuss two aspects of the results. One has to do with "cold" start versus "hot" start strategies in the algorithm and in the complexity results. It is argued that the complexity results appear to be consistent with the observed practical sensitivity of interior-point algorithms to certain types of "cold" and "hot" starting points. Finally, we remark on the use of the algorithm for detecting infeasibility in the primal and/or the dual problem.

Notation

$\mathbb{R}^n$ denotes real n-dimensional space. $e = (1, 1, \ldots, 1)^T$ is the vector of ones of appropriate dimension. $\|\cdot\|$ denotes the Euclidean norm and $\|\cdot\|_\infty$ denotes the infinity-norm, i.e., $\|v\|_\infty = \max_j\{|v_j|\}$. Also, let $\|\cdot\|_1$ denote the L_1 norm, i.e., $\|v\|_1 = \sum_{j=1}^n |v_j|$. If x and s are n-vectors, then X and S denote the diagonal matrices whose diagonal components correspond to the components of x and s, respectively. We will adopt the following conventions for the use of the symbol ∞: $\ln(\infty) = \infty$, and $\ln(0) = -\infty$.

2. The infeasible-start problem and measures of closeness to the feasible and optimal regions

The central purpose of this paper is the development of an algorithm for linear programming whose efficiency (as measured in worst-case analysis) is sensitive to the distance from the initial starting point to the optimal solution of the linear program (measured in some suitable form). We will measure this distance in a number of different ways, as explained and discussed in this section.

Consider the linear programming problem:

$$\text{P:} \qquad \underset{x}{\text{minimize}} \quad c^T x$$
$$\text{subject to} \quad Ax = b,$$
$$x \geq 0,$$

whose dual is given by:

$$\text{D:} \qquad \underset{y,s}{\text{maximize}} \quad b^T y$$
$$\text{subject to} \quad A^T y + s = c,$$
$$s \geq 0.$$

The data for P (and D) is the triplet (A, b, c), where A is an $m \times n$ matrix, and b and c are m- and n-vectors, respectively. We make the following assumption regarding (A, b, c), which is elementary to verify and involves no loss of generality:

(A.1): A has rank m.

We make no assumption regarding the existence of feasible solutions or the boundedness of the set of optimal solutions.

If $\bar{x}$ is a feasible solution of P, $\bar{x}$ is defined to be a *nondegenerate* solution if A_β has rank m, where $\beta = \{j \mid \bar{x}_j > 0\}$ and A_γ is the submatrix of A whose columns are formed by choosing columns of A whose indices are in the set γ. Analogously for the dual, if $(\bar{y}, \bar{s})$ is a feasible solution to D, $(\bar{y}, \bar{s})$ is defined to be a *nondegenerate* solution if A_α has rank $|\alpha|$, where $\alpha = \{j \mid \bar{s}_j = 0\}$. We will say the triplet $(\bar{x}, \bar{y}, \bar{s})$ is *nondegenerate* if $\bar{x}$ is a nondegenerate solution of P and $(\bar{y}, \bar{s})$ is a nondegenerate solution of D.

2.1. THE INITIAL POINT

We suppose that we are given an initial starting point for P and D, i.e., a triplet $(\hat{x}, \hat{y}, \hat{s})$, where $\hat{x}$ is a starting point for P and $(\hat{y}, \hat{s})$ is a starting point for D. We make no assumption regarding the extent to which $\hat{x}$ and $(\hat{y}, \hat{s})$ satisfy the equations of P and D. That is, there is no assumption regarding the quantities $A\hat{x} - b$ and $A^T\hat{y} + \hat{s} - c$. However, we do assume the following properties of $(\hat{x}, \hat{y}, \hat{s})$:

(P.1) $\hat{x} > 0$.

(P.2) $\hat{s} > 0$.

(P.3) $\hat{X}\hat{S}e = \hat{\theta}e$ for some scalar $\hat{\theta}$.

Properties (P.1) and (P.2) state that $\hat{x}$ and $\hat{s}$ are strictly positive vectors, and (P.3) states that $\hat{x}$ and $\hat{s}$ are "centered" vis-à-vis one another, i.e., $\hat{x}_i\hat{s}_i = \hat{x}_j\hat{s}_j$ for all $i, j = 1,\ldots,n$, and $\hat{\theta}$ is equal to the common value of $\hat{x}_j\hat{s}_j$.

Properties (P.1) and (P.2) assume the positivity of the slack vectors for the primal and for the dual. From the standpoint of interior-point methodology, there needs to be something that is positive about the initial point, and this positivity is guaranteed by (P.1) and (P.2). Property (P.3) is a centering condition, that aligns the primal and dual slacks with one another.

Although these properties may seem restrictive (particularly property (P.3)), we now argue that there are a number of ways that these three properties can arise in practice, either naturally or with only a little bit of coercion. Suppose, for example, that the starting point $(\hat{x}, \hat{y}, \hat{s})$ is the output from an interior-point algorithm solution (within a small tolerance) of a previous version of the linear program. Then (P.1) and (P.2) will be satisfied, and (P.3) will be satisfied to a small tolerance if the algorithm had produced a solution on the central trajectory, and can easily be "smoothed" to satisfy (P.3) exactly. Suppose instead that $(\hat{x}, \hat{y}, \hat{s})$ is the output from the simplex method as applied to a previous version of the linear program. Given a desired value of the scalar $\hat{\theta}$ (e.g., $\hat{\theta} = 0.001$), if the solution $(\hat{x}, \hat{y}, \hat{s})$ is strictly complementary (i.e., $\hat{x} + \hat{s} > 0$) one can replace all zero slacks in the primal by $\hat{\theta}/\hat{s}_j$ and all zero slacks in the dual by $\hat{\theta}/\hat{x}_j$, and in this way (P.1), (P.2), and (P.3) will be satisfied. If the solution $(\hat{x}, \hat{y}, \hat{s})$ is not strictly complementary, then in all indices for which $\hat{x}_j$ and $\hat{s}_j$ are zero, one can replace $\hat{x}_j$ and $\hat{s}_j$ by the value $\sqrt{\hat{\theta}}$, and (P.1), (P.2), and (P.3) will be satisfied. These arguments are intended to show the reasonableness of the assumption that $(\hat{x}, \hat{y}, \hat{s})$ satisfies the three properties (P.1), (P.2), and (P.3).

Since the purpose of this study is the development of an algorithm for linear programming whose behavior is sensitive to how close the starting point is to the feasible region and to the optimal solution, we need to explore how this closeness is measured. We begin by looking at how to measure closeness to feasibility.

2.2. MEASURING CLOSENESS TO FEASIBILITY

Let the feasible regions of P and D be denoted as

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$$

and

$$S = \{s \in \mathbb{R}^n \mid s \geq 0 \text{ and } A^T y + s = c \text{ for some } y \in \mathbb{R}^m\},$$

where it should be noted that, technically, S is the projection of the feasible values of (y, s) onto the space of s-variables. We denote the optimal regions of P and D by

$$\mathcal{X}^* = \{x \in \mathcal{X} \mid x \text{ solves } P\}, \quad \text{and}$$

$$S^* = \{s \in S \mid (y, s) \text{ solves } D \text{ for some } y \in \mathbb{R}^m\}.$$

From the duality theory of linear programming, $\mathcal{X}^* \neq \emptyset$, if and only if $S^* \neq \emptyset$.

We will use two types of measures to indicate how close $\hat{x}$ (respectively, $\hat{s}$) is to the feasible region $\mathcal{X}$ (respectively, S). Given the triplet $(\hat{x}, \hat{y}, \hat{s})$ satisfying (P.1), (P.2), and (P.3), then $\hat{x} > 0$ and $\hat{s} > 0$, and so $\hat{x}$ is feasible for P if $A\hat{x} = b$ (and $(\hat{y}, \hat{s})$ is feasible for D if $A^T\hat{y} + \hat{s} = c$). Thus, one way to measure how close $\hat{x}$ is to the feasible region $\mathcal{X}$ is to compute the norm $\|b - A\hat{x}\|$. Similarly for the dual, one way to measure how close $(\hat{y}, \hat{s})$ is to the dual feasible region S is to compute the norm $\|c - A^T\hat{y} - \hat{s}\|$. Thus, our first measures of closeness to the feasible regions $\mathcal{X}$ and S are the norms $\|b - A\hat{x}\|$ and $\|c - A^T\hat{y} - \hat{s}\|$.

We develop a second measure of closeness as follows. Considering the primal feasible region $\mathcal{X}$ as an example, one can measure the Euclidean distance from $\hat{x}$ to $\mathcal{X}$ by considering the program

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|x - \hat{x}\| \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0. \end{aligned}$$

However, we can relax (and save considerable effort) by ignoring the inequality constraints and (noting that $\hat{x} > 0$ from (P.1)), we can instead measure closeness to $\mathcal{X}$ by considering the program

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|x - \hat{x}\| \\ \text{subject to} \quad & Ax = b. \end{aligned}$$

This program measures how close the positive vector $\hat{x}$ is to the affine manifold of solutions to the system $Ax = b$. Since the Euclidean norm is somewhat arbitrary, and scaling will be important in our algorithm, we will consider a version of the above program scaled by the components $(\hat{x}_j)^{-1}$, $j = 1, \ldots, n$:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|\hat{X}^{-1}(x - \hat{x})\| \\ & Ax = b. \end{aligned}$$

Arguing in a parallel manner for the dual, we define the following measures of closeness to the feasible region. Define

$$\rho_1 = \underset{x}{\text{minimize}} \ \|\hat{X}^{-1}(x - \hat{x})\|, \quad \text{subject to } Ax = b, \qquad \sigma_1 = \underset{y,s}{\text{minimize}} \ \|\hat{S}^{-1}(s - \hat{s})\| \quad \text{subject to } A^T y + s = c \tag{2.1a}$$

and

$$\tilde{x} = \arg\min_{x} \ \|\hat{X}^{-1}(x - \hat{x})\|, \quad \text{subject to } Ax = b, \qquad (\tilde{y}, \tilde{s}) = \arg\min_{y,s} \ \|\hat{S}^{-1}(s - \hat{s})\| \quad \text{subject to } A^T y + s = c \qquad (2.1b)$$

and

$$\delta_1 = \max\{\rho_1, \sigma_1\}. \qquad (2.1c)$$

Note that because A has full row rank, all the above quantities are well-defined and easily computed. The quantity ρ_1 measures the weighted (by $\hat{X}^{-1}$) Euclidean distance from the positive point $\hat{x}$ to the affine space of x that satisfies $Ax = b$, and is a positive number. One might ask why is the Euclidean norm used in (2.1a, b, c)? As it turns out, better bounds in the analysis are obtained if the infinity norm is used throughout, but this norm does not lend itself to easy computation if used in (2.1a), whereas the Euclidean norm does. In fact, because (2.1) uses the Euclidean norm, we can obtain the following closed form solutions for (2.1):

$$\rho_1 = \sqrt{(b - A\hat{x})^T (A\hat{X}^2 A^T)^{-1}(b - A\hat{x})} \qquad (2.2a)$$

and

$$\sigma_1 = \sqrt{(c - A^T\hat{y} - \hat{s})^T \hat{S}^{-1} P \hat{S}^{-1}(c - A^T\hat{y} - \hat{s})}, \qquad (2.2b)$$

where

$$P = [I - \hat{S}^{-1}A^T(A\hat{S}^{-2}A^T)^{-1}A\hat{S}^{-1}] \qquad (2.2c)$$

and

$$\tilde{x} = \hat{x} + \hat{X}^2 A^T (A\hat{X}^2 A^T)^{-1}(b - A\hat{x}), \qquad (2.3a)$$

$$(\tilde{y}, \tilde{s}) = \big((A\hat{S}^{-2}A^T)^{-1}A\hat{S}^{-2}(c - \hat{s}), \hat{s} + \hat{S}P\hat{S}^{-1}(c - A^T\hat{y} - \hat{s})\big). \qquad (2.3b)$$

Note that $\rho_1 = 0$ if $\hat{x}$ is feasible for the primal, and is positive otherwise. Similarly, $\sigma_1 = 0$ if $(\hat{y}, \hat{s})$ is feasible for the dual, and is positive otherwise. The overall measure of infeasibility is δ_1 which is the maximum of ρ_1 and σ_1. The following proposition demonstrates some limiting properties of the measure δ_1.

PROPOSITION 2.1

Suppose $(\hat{x}^k, \hat{y}^k, \hat{s}^k)$ is a sequence of starting points and $(\hat{x}^k, \hat{y}^k, \hat{s}^k) \to (\bar{x}, \bar{y}, \bar{s})$ as $k \to \infty$, where $\bar{x} \in \mathcal{X}$ and $\bar{s} \in S$. Let δ_1^k denote the value of δ_1 for the triplet $(\hat{x}^k, \hat{y}^k, \hat{s}^k)$. Then

(i) $\limsup_{k\to\infty}\{\delta_1^k\} \le \sqrt{n}$, and

(ii) if $\bar{x}$ and $(\bar{y}, \bar{s})$ are nondegenerate solutions, then $\lim_{k\to\infty}\{\delta_1^k\} = 0$.

Proof

We first prove (i). For each k, consider the computation ρ_1^k.

$$\rho_1^k = \min_x \{\|(\hat{X}^k)^{-1}(x - \hat{x}^k)\| \text{ subject to } Ax = b\} \leq \sqrt{n} \min_x \{\|e - (\hat{X}^k)^{-1}x\|_\infty \text{ subject to } Ax = b\}$$

$$\leq \sqrt{n}\,\|e - (\hat{X}^k)^{-1}\bar{x}\|_\infty.$$

As $k \to \infty$, $\bar{x}_j/\hat{x}_j^k \to 1$ if $\bar{x}_j > 0$, and $\bar{x}_j/\hat{x}_j^k \to 0$ if $\bar{x}_j = 0$. Therefore, $\limsup_{k\to\infty}\{\rho_1^k\} \leq \sqrt{n}$. A similar argument holds for σ_1^k, and so $\limsup_{k\to\infty}\{\delta_1^k\} \leq \sqrt{n}$.

To prove (ii), let $\beta = \{j \mid \bar{x}_j > 0\}$, $\alpha = \{j \mid \bar{x}_j = 0\}$. For each k, consider the computation of ρ_1^k:

$$\rho_1^k = \min_x \{\|(\hat{X}^k)^{-1}(x - \hat{x}^k)\| \text{ subject to } Ax = b\} \leq \min_x \{\|(\hat{X}^k)^{-1}(x - \hat{x}^k)\| \text{ subject to } Ax = b,\ x_\alpha = \hat{x}_\alpha^k\}$$

$$= \min_{x_\beta} \{\|(\hat{X}^k)_{\beta\beta}^{-1}(x_\beta - \hat{x}_\beta)\| \text{ subject to } A_\beta x_\beta = b - A_\alpha \hat{x}_\alpha^k\}$$

$$= \sqrt{(b - A\hat{x}^k)^T (A_\beta (\hat{X}_{\beta\beta}^k)^2 A_\beta^T)^{-1}(b - A\hat{x}^k)}.$$

As $k \to \infty$, $b - A\hat{x}^k \to b - A\bar{x} = 0$, and $A_\beta(\hat{X}_{\beta\beta}^k)A_\beta^T \to A\bar{X}^2A^T$, which is invertible. Thus, $\rho_1^k \to 0$ as $k \to \infty$. A similar argument holds for σ_1^k, and so $\lim_{k\to\infty}\{\delta_1^k) = 0$. □

Note in the proof of the proposition that if the infinity norm were used to define δ_1 in (2.1), then the lim sup in (i) would be 1 instead of $\sqrt{n}$.

2.3. MEASURING CLOSENESS TO OPTIMALITY

We now turn our attention to two measures of the closeness of $(\hat{x}, \hat{y}, \hat{s})$ to the set of optimal solutions of P and D. One way to measure how close $\hat{x}$ and $(\hat{y}, \hat{s})$ are to the set of optimal solutions is by measuring the complementary slackness of $\hat{x}$ and $(\hat{y}, \hat{s})$ by computing

$$\hat{x}^T\hat{s} = n\hat{\theta}.$$

If $\|\hat{x} - x^*\| < \varepsilon$ and $\|\hat{s} - s^*\| < \varepsilon$, where $x^* \in \mathcal{X}^*$ and $s^* \in S^*$, then $n\hat{\theta} = \hat{x}^T\hat{s} \leq n\varepsilon^2 + n\varepsilon\|x^* + s^*\|_\infty$, which goes to zero as ε goes to zero.

We develop a second measure as follows. Let

$$\rho_2 = \inf_{x^* \in \mathcal{X}^*} \| \hat{X}^{-1} x^* \|_\infty, \quad \sigma_2 = \inf_{s^* \in S^*} \| \hat{S}^{-1} s^* \|_\infty, \tag{2.4a}$$

and

$$\delta_2 = \max\{\rho_2, \sigma_2, 1\}. \tag{2.4b}$$

Note in (2.4b) that $\delta_2 \geq 1$. As it turns out, the complexity analysis uses $\ln(\delta_2)$, and so $\delta_2 \geq 1$ conveniently ensures that $\ln(\delta_2) \geq 0$. Note that if $\hat{x} \geq x^*$ and $\hat{s} \geq s^*$ for some $x^* \in \mathcal{X}^*$, $s^* \in S^*$, then $\delta_2 = 1$ and so $\ln(\delta_2) = 0$.

PROPOSITION 2.2

Suppose $(\hat{x}^k, \hat{y}^k, \hat{s}^k)$ is a sequence of starting points and $(\hat{x}^k, \hat{y}^k, \hat{s}^k) \to (x^*, y^*, s^*)$ as $k \to \infty$, where $x^* \in \mathcal{X}^*$ and $s^* \in S^*$. Let δ_2^k denote the value of δ_2 for the triplet $(\hat{x}^k, \hat{y}^k, \hat{s}^k)$.
Then $\lim_{k \to \infty} \ln(\delta_2) = 0$. □

In (2.4), note that it is possible to have $\rho_2 = +\infty$, $\sigma_2 = +\infty$, and $\delta_2 = +\infty$. If any one of these three quantities is finite, all three are finite and all limits are attained. We close this section with the following:

PROPOSITION 2.3

If $\bar{x} > 0$ and $\bar{s} > 0$, then

$$\| \bar{S}(\hat{x} - \bar{x}) \| + \| \bar{X}(\hat{s} - \bar{s}) \| \leq \delta_1 (\| \hat{X}\bar{s} \| + \| \hat{S}\bar{x} \|).$$

Proof

The quantity δ_1 satisfies

$$\delta_1 \geq (\hat{x}_j)^{-1} |\hat{x}_j - \bar{x}_j|, \quad \delta_1 \geq (\hat{s}_j)^{-1} |\hat{s}_j - \bar{s}_j| \quad \text{for } j = 1, \ldots, n, \tag{2.5}$$

see (2.1c), and therefore

$$\| \bar{S}(\hat{x} - \bar{x}) \| \leq \delta_1 \| \bar{S}\hat{x} \| = \delta_1 \| \hat{X}\bar{s} \|$$

and

$$\| \bar{X}(\hat{s} - \bar{s}) \| \leq \delta_1 \| \bar{X}\hat{s} \| = \delta_1 \| \hat{S}\bar{x} \|.$$

Combining these last two statements gives the desired result. □

Remark 2.3

Note that if the infinity norm were used to define δ_1, then proposition 2.3 would still be valid.

3. The infeasible-start barrier problem

Given the data (A, b, c) for P and D, and the starting solution $(\hat{x}, \hat{y}, \hat{s})$, we create the following parametric family of linear programs:

$$\begin{aligned} \text{P}(\varepsilon): \quad & \underset{x}{\text{minimize}} && [c + \varepsilon(A^T\hat{y} + \hat{s} - c)]^T x && (3.1a) \\ & \text{subject to} && Ax = b + \varepsilon[A\hat{x} - b], \\ & && x \geq 0, \end{aligned}$$

and its dual:

$$\begin{aligned} \text{D}(\varepsilon): \quad & \underset{y,s}{\text{maximize}} && (b + \varepsilon[A\hat{x} - b])^T y && (3.1b) \\ & \text{subject to} && A^T y + s = c + \varepsilon[A^T\hat{y} + \hat{s} - c], \\ & && s \geq 0. \end{aligned}$$

Note that $\hat{x}$ and $(\hat{y}, \hat{s})$ are feasible solutions of $\text{P}(\varepsilon)$ and $\text{D}(\varepsilon)$, respectively, at $\varepsilon = 1$, and we would like to find an optimal solution to $\text{P}(\varepsilon)$ and $\text{D}(\varepsilon)$ at $\varepsilon = 0$, since P(0) is P and D(0) is D. In order to accomplish this, we create the following logarithmic barrier problem:

$$\begin{aligned} \text{BP}(\varepsilon, \mu): \quad & \underset{x}{\text{minimize}} && (c + \varepsilon[A^T\hat{y} + \hat{s} - c])^T x - \varepsilon\mu \sum_{j=1}^{n} \ln(x_j) && (3.2) \\ & \text{subject to} && Ax = b + \varepsilon[A\hat{x} - b], \\ & && x > 0. \end{aligned}$$

Note that ε appears in three places in $\text{BP}(\varepsilon, \mu)$: on the RHS of the constraints, in the linear part of the objective function, and in the multiplier in front of the logarithmic barrier term. Note that μ appears only in the multiplier of the logarithmic barrier term. The optimality conditions for $\text{BP}(\varepsilon, \mu)$ state that x is a solution to $\text{BP}(\varepsilon, \mu)$ together with dual variables (y, s) if and only if the following Karush–Kuhn–Tucker conditions are satisfied:

$$Ax = b + \varepsilon[A\hat{x} - b], \tag{3.3a}$$

$$x > 0, \tag{3.3b}$$

$$A^T y + s = c + \varepsilon[A^T\hat{y} + \hat{s} - c], \tag{3.3c}$$

$$s > 0, \tag{3.3d}$$

$$\left(\frac{1}{\mu\varepsilon}\right) XSe - e = 0. \tag{3.4}$$

Note that (3.3a, b) indicates that x is feasible for P(ε) and (3.3c, d) indicates that (y, s) is feasible for D(ε).

Let γ be a scalar, $0 < \gamma < 1$. We will say that $(x, y, s, \varepsilon, \mu)$ is a γ-approximate solution to BP(ε, μ) if $(x, y, s, \varepsilon, \mu)$ satisfies (3.3a)–(3.3d), plus

$$\left\| \left(\frac{1}{\mu \varepsilon} \right) XSe - e \right\| \leq \gamma. \tag{3.3e}$$

Note that (3.3e) is just a relaxation of (3.4), and is in the class of approximation measures used in Roos and Vial [24] and Tseng [28]. (In the algorithm of the next section, we will use $\gamma = 1/4$ typically.)

We have the following properties of a γ-approximate solution to BP(ε, μ):

PROPOSITION 3.1

Suppose that $(x, y, s, \varepsilon, \mu)$ is a γ-approximate solution to BP(ε, μ) and $0 < \gamma < 1$. Then

(i) x is a feasible solution of P(ε),

(ii) (y, s) is a feasible solution of D(ε), and

(iii) the duality gap is $x^T s \leq \varepsilon \mu n(1 + \gamma)$.

Proof

Assertions (i) and (ii) follow directly from (3.3a–d). Elementary algebra reveals the duality gap is $(c + \varepsilon[A^T \hat{y} + \hat{s} - c])^T x - (b + \varepsilon[A\hat{x} - b]^T y = x^T s$. From (3.3e), $x_j s_j \leq \varepsilon \mu(1 + \gamma)$, and so $x^T s \leq \varepsilon \mu n(1 + \gamma)$. □

In view of proposition 3.1, one strategy for solving P and D is to find a sequence of γ-approximate solutions to BP(ε, μ) for a sequence of values of $\varepsilon \to 0$ and $(\varepsilon\mu) \to 0$. We will describe an algorithm that will accomplish this in the next section. We close this section with the following starting point for such an algorithm.

Let

$$\begin{aligned} (x^0, y^0, s^0) &= (\hat{x}, \hat{y}, \hat{s}), \\ \varepsilon^0 &= 1, \\ \mu^0 &= \hat{\theta}. \end{aligned} \tag{3.5}$$

PROPOSITION 3.2

$(x^0, y^0, s^0, \varepsilon^0, \mu^0)$ is a γ-approximate solution of BP(ε^0, μ^0) for any $\gamma \in (0, 1)$.

Proof

Direct algebraic substitution demonstrates that $(x, y, s, \varepsilon, \mu) = (x^0, y^0, s^0, \varepsilon^0, \mu^0)$ satisfies (3.3a)–(3.3d). Also,

$$\frac{1}{\varepsilon^0 \mu^0} X^0 S^0 e - e = \frac{1}{\hat{\theta}} \hat{X}\hat{S}e - e = 0$$

(from P.3), and so (3.3e) is satisfied by any $\gamma \in (0, 1)$. □

4. The path-following algorithm for the barrier problem

In this section, we present a path-following algorithm for solving for a γ-approximate solution of BP(ε, μ) for a sequence of values of ε and μ, where $\varepsilon \to 0$ and $\varepsilon\mu \to 0$, starting at the initial solution $(x, y, s, \varepsilon, \mu) = (x^0, y^0, s^0, \varepsilon^0, \mu^0) = (\hat{x}, \hat{y}, \hat{s}, 1, \hat{\theta})$, see (3.5) and proposition 3.2.

In view of the system (3.3a–e) and proposition 3.1, we would like to solve for a sequence of γ-approximate solutions $(x^k, y^k, s^k, \varepsilon^k, \mu^k)$ of BP(ε^k, μ^k), for a sequence of values of ε^k, where $\varepsilon^k \to 0$. Also, in light of proposition 3.1(iii), it would be advantageous if μ^k is kept as small as possible, so as to enforce a small duality gap. Ideally, it would be advantageous to shrink ε^k by a fractional quantity $\alpha \in (0, 1)$ at each iterate, so that $\varepsilon^{k+1} = \alpha\varepsilon^k$, and to maintain $\mu^{k+1} = \mu^k = \mu^0 = \hat{\theta}$ throughout the algorithm. However, we are not quite able to accomplish this. Instead, it may be necessary at some iterations to leave ε at its current value ε^k, and increase μ^k by a scalar quantity $\beta > 1$, and set $\mu^{k+1} = \beta\mu^k$. Nevertheless, we will be able to compute an upper bound on the number of iterations at which μ^k is increased. We will also compute an upper bound on the largest value of μ^k produced in the algorithm. The algorithm is given below, and an explanation and discussion of the steps of the algorithm follows.

ALGORITHM $(A, b, c, \hat{x}, \hat{y}, \hat{s}, \gamma)$

Step 0. *(Initialize)*

Set $(x^0, y^0, s^0) = (\hat{x}, \hat{y}, \hat{s})$, $\varepsilon^0 = 1$, $\mu^0 = \hat{\theta} = \hat{x}^T\hat{s}/n$.
Compute ρ_1, σ_1, δ_1 using (2.1)–(2.3).

$$\beta = 1 + \frac{\sqrt{\gamma} - \gamma}{\sqrt{n} - \sqrt{\gamma}},$$

$$\alpha = 1 - \frac{(\sqrt{\gamma} - \gamma)}{\sqrt{\gamma} + \sqrt{n} + n\delta_1(1 + \gamma + \beta)/(1 - \gamma)}. \tag{4.1}$$

Step 1. (*Test*)

$$(\bar{x}, \bar{y}, \bar{s}, \bar{\varepsilon}, \bar{\mu}) = (x^k, y^k, s^k, \varepsilon^k, \mu^k).$$

$$\begin{aligned} &\text{If } (\|\hat{S}\bar{x}\| + \|\hat{X}\bar{s}\|)/\bar{\mu} \leq n(1+\gamma+\beta), \text{ go to step 2.} \\ &\text{If } (\|\hat{S}\bar{x}\| + \|\hat{X}\bar{s}\|)/\bar{\mu} > n(1+\gamma+\beta), \text{ go to step 3.} \end{aligned} \tag{4.2}$$

Step 2. (*Decrease $\bar{\varepsilon}$ and take a Newton step*)

Set
$$\varepsilon = \alpha\bar{\varepsilon}, \quad \mu = \bar{\mu}. \tag{4.3a}$$

Solve the Newton equations (4.3b, c) in the variables (d, y):

$$Ad = -\bar{\varepsilon}(1-\alpha)[A\hat{x} - b], \tag{4.3b}$$

$$-\varepsilon\mu\bar{X}^{-2}d + A^T y = c + \varepsilon[A^T\hat{y} + \hat{s} - c] - \varepsilon\mu\bar{X}^{-1}e. \tag{4.3c}$$

Set
$$x = \bar{x} + d, \tag{4.3d}$$

$$s = \varepsilon\mu\bar{X}^{-1}(e - \bar{X}^{-1}d). \tag{4.3e}$$

Set $(x^{k+1}, y^{k+1}, s^{k+1}, \varepsilon^{k+1}, \mu^{k+1}) = (x, y, s, \varepsilon, \mu)$.

Set $k \leftarrow k+1$. Go to step 1.

Step 3. (*Increase $\bar{\mu}$ and take a Newton step*)

Set
$$\mu = \beta\bar{\mu}, \quad \varepsilon = \bar{\varepsilon}, \tag{4.4a}$$

Solve the Newton equations (4.4b, c) in the variables (d, y):

$$Ad = 0, \tag{4.4b}$$

$$-\varepsilon\mu\bar{X}^{-2}d + A^T y = c + \varepsilon[A^T\hat{y} + \hat{s} - c] - \varepsilon\mu\bar{X}^{-1}e. \tag{4.4c}$$

Set
$$x = \bar{x} + d, \tag{4.4d}$$

$$s = \varepsilon\mu\bar{X}^{-1}(e - \bar{X}^{-1}d). \tag{4.4e}$$

Set $(x^{k+1}, y^{k+1}, s^{k+1}, \varepsilon^{k+1}, \mu^{k+1}) = (x, y, s, \varepsilon, \mu)$.

Set $k \leftarrow k+1$. Go to step 1.

We now review the steps of the algorithm. At step 0, the algorithm is initialized with values of $(x^0, y^0, s^0, \varepsilon^0, \mu^0)$. The distance measures ρ_1, σ_1, and δ_1 that measure closeness to the feasible region are computed. The quantities β and α are computed. As discussed, the quantity β ($\beta > 1$) is used to increase μ^k to $\mu^{k+1} = \beta\mu^k$ at some iterations of the algorithm, and α ($\alpha < 1$) is used to decrease ε^k to $\varepsilon^{k+1} = \alpha\varepsilon^k$ at

other iterations of the algorithm. At step 1, a simple norm test is used to branch to either step 2 or step 3. If the algorithm branches to step 2, ε is decreased by the fractional quantity α, and μ is kept at its previous value (4.3a), and the Newton direction d is computed along with multipliers y in the Newton equations (4.3b, c). The current value $\bar{x}$ is updated by adding the Newton increment (4.3d), and new dual slack values s are defined in (4.3e). If the algorithm instead branches to step 3, μ is increased by the factor $\beta > 1$, and ε is kept at its previous value (4.4a), and the Newton direction d is computed along with dual multipliers y in the Newton equations (4.4b, c). The current value $\bar{x}$ is updated by adding the Newton increment in (4.4d), and new dual slack values s are defined in (4.4e).

We will prove the following two theorems regarding the behavior of this algorithm:

THEOREM 4.1

Suppose that $\gamma \in (0, 1)$. Then each iterate value $(x^k, y^k, s^k, \varepsilon^k, \mu^k)$, $k = 1,\ldots,$ is a γ-approximate solution of $BP(\varepsilon^k, \mu^k)$. □

THEOREM 4.2

Suppose that $\gamma \in (0, 1)$ and let δ_2 be defined in (2.4). Then for all iterations k,

$$\mu^k \leq \delta_2 \hat{\theta}. \tag{4.5}$$

□

While theorem 4.1 states that each iteration is a solution to (3.3a–e), theorem 4.2 states that the values of μ^k are bounded by $\delta_2 \hat{\theta}$ (an unknown quantity). Theorem 4.1 is proved in section 6, and theorem 4.2 is proved in section 7. One consequence of theorem 4.2 is the following:

COROLLARY 4.3

Suppose $\gamma = 1/4$. Let T_3 be the number of iterations where the algorithm visits step 3. Then $T_3 \leq 4\sqrt{n} \ln(\delta_2)$.

Proof of corollary 4.3

Since $\mu^0 = \hat{\theta}$, after the algorithm visits step 3 T_3 times, $\mu^k = \mu^0 \beta^{T_3}$ for some iteration k. From theorem 4.2,

$$\mu^0 \beta^{T_3} \leq \hat{\theta} \delta_2,$$

and since $\mu^0 = \hat{\theta}$, this implies that $\delta_2 (1/\beta)^{T_3} \geq 1$. Taking logarithms, we obtain

$$0 \le \ln(\delta_2) + T_3 \ln\left(\frac{1}{\beta}\right) \le \ln(\delta_2) + T_3\left(\frac{1}{\beta} - 1\right), \tag{4.6}$$

where the last inequality follows from the concavity of the logarithm function. Substituting $\gamma = 1/4$ in the definition of β in (4.1), we obtain from (4.6):

$$T_3 \le (4\sqrt{n} - 1)\ln(\delta_2) \le 4\sqrt{n}\ln(\delta_2),$$

since $\delta_2 \ge 1$. □

Corollary 4.3 places an upper bound on the number of iterations of the algorithm that visit step 3 and increase the value of μ, and this upper bound is given by $4\sqrt{n}\ln(\delta_2)$.

Another consequence of theorem 4.2 is a bound on the $\|\cdot\|_\infty$ ball containing an optimal solution to P or D. Corollary 8.1 in section 8 uses theorem 4.2 to derive a bound on the size of any primal or dual solution and yields a polynomial time test for infeasibility in P or D. See section 8 for details.

It should be noted that the Newton equations (4.3b, c) or (4.4b, c) use only primal scaling, as opposed to primal-dual scaling. They are derived by approximating the barrier problem $BP(\varepsilon, \mu)$ by its second-order (quadratic) approximation at the primal point $\bar{x}$, and so only use primal scaling. However, in view of the centering condition (3.3e) and theorem 4.1, one can easily show that the primal scaling matrix $(\bar{X})$ is not materially different from the primal-dual scaling matrix $(\bar{X}\bar{S}^{-1})^{1/2}$, and that the algorithm is in fact a primal-dual type of algorithm.

Finally, note that the algorithm is of the short fixed-step variety of path-following algorithms. There is the potential for the use of a line-search to enhance the convergence of the algorithm (if used with an aggressive choice for α and/or β at each iteration, or as part of a predictor-corrector variant of the algorithm), but such a line-search might not be very valuable in practice.

5. Complexity analysis of the algorithm

We say that x is an ε_1^*-feasible solution to the primal P if $x \ge 0$ and $\|b - Ax\| \le \varepsilon_1^*$. We say that (y, s) is an ε_1^*-feasible solution to the dual D if $s \ge 0$ and $\|c - A^T y - s\| \le \varepsilon_1^*$. We say that x and (y, s) are an ε_2^*-optimal solution to the primal-dual pair P and D if $x^T s \le \varepsilon_2^*$.

Now let us examine the algorithm. Because ε^k is a monotone decreasing sequence, if $\|b - A\hat{x}\| \le \varepsilon_1^*$, then $\|b - Ax^k\| \le \varepsilon_1^*$ for all iterate values. That is, if the starting primal solution $\hat{x}$ is ε_1^*-feasible, then all primal iterate values x^k will be ε_1^*-feasible. Similarly for the dual, if $(\hat{y}, \hat{s})$ is ε_1^*-feasible for the dual, then (y^k, s^k) will be ε_1^*-feasible for the dual for all iterate values (y^k, s^k). If $\|b - A\hat{x}\| > \varepsilon_1^*$ and/or $\|c - A^T\hat{y} - \hat{s}\| > \varepsilon_1^*$, we have the following complexity results:

THEOREM 5.1 (Primal feasibility)

Let $\gamma = 1/4$ in the algorithm, and let

$$T_P = \left\lceil (2 + 4\sqrt{n} + 15n\,\delta_1) \ln\left(\frac{\|b - A\hat{x}\|}{\varepsilon_1^*}\right)\right\rceil + \left\lfloor (4\sqrt{n}) \ln(\delta_2)\right\rfloor.$$

Then for all iterate values $k \geq T_P$, x^k is an ε_1^*-feasible solution to the primal P. □

THEOREM 5.2 (Dual feasibility)

Let $\gamma = 1/4$ in the algorithm, and let

$$T_D = \left\lceil (2 + 4\sqrt{n} + 15n\,\delta_1) \ln\left(\frac{\|c - A^T\hat{y} - \hat{s}\|}{\varepsilon_1^*}\right)\right\rceil + \left\lfloor (4\sqrt{n}) \ln(\delta_2)\right\rfloor.$$

Then for all iterate values $k \geq T_D$, (y^k, s^k) is an ε_1^*-feasible solution to the dual D. □

Before proving these two theorems, we offer the following interpretive remarks. For a fixed $\varepsilon_1^* > 0$, the quantity T_P (respectively, T_D) is a function only of n, δ_1, δ_2, and $\|b - A\hat{x}\|$ (respectively, $\|c - A^T\hat{y} - \hat{s}\|$), and is monotone increasing in all of these quantities. As developed in section 2, δ_1, $\|b - A\hat{x}\|$ and $\|c - A^T\hat{y} - \hat{s}\|$ are distance measures from the initial point $(\hat{x}, \hat{y}, \hat{s})$ to the feasible regions $\mathcal{X}$ and S, and δ_2 is a distance measure from the initial point $(\hat{x}, \hat{y}, \hat{s})$ to the optimal regions $\mathcal{X}^*$ and S^*.

Suppose $(\hat{x}^k, \hat{y}^k, \hat{s}^k)$ is a sequence of starting points whose limit is an optimal point (x^*, y^*, s^*), where $x^* \in \mathcal{X}^*$ and $s^* \in S^*$. Then from propositions 2.1 and 2.2, $\limsup_{k\to\infty}\{\delta_1^k\} \leq \sqrt{n}$, $\lim_{k\to\infty} \ln(\delta_2^k) = 0$, and since $\|b - A\hat{x}^k\| \to 0$, $T_P \to 0$ as $k \to \infty$. Thus, as the starting point $(\hat{x}, \hat{y}, \hat{s})$ approaches *any* optimal solution (x^*, y^*, s^*), T_P goes to zero. The exact same analysis holds for the dual problem. Thus, as $(\hat{x}, \hat{y}, \hat{s})$ approaches any optimal solution (x^*, y^*, s^*), T_D goes to zero.

Let us also consider the case when either the primal or the dual have no feasible solution. Then $\delta_2 = +\infty$, and so $T_P = +\infty$ and $T_D = +\infty$, as expected. (Note that when $\delta_2 = +\infty$, the algorithm cannot visit step 2 infinitely often. For if so, then $\lim_{k\to\infty} \varepsilon^k = 0$, but theorem 4.1 and proposition 3.1 would imply that $\min_{x\geq 0, s\geq 0, y}\{\|b - Ax\| + \|c - A^T y - s\|\} = 0$, contradicting the infeasibility of either the primal or the dual.)

We also have the following complexity result regarding ε_2^*-optimal solutions.

THEOREM 5.3 (Optimality)

Let $\gamma = 1/4$ in the algorithm, and let

$$T_O = \left\lceil (2 + 4\sqrt{n} + 15n\,\delta_1) \left(\ln(\delta_2) + \ln\left(\frac{(\hat{x}^T \hat{s})(1.25)}{\varepsilon_2^*} \right) \right) \right\rceil + \left\lfloor (4\sqrt{n}) \ln(\delta_2) \right\rfloor.$$

Then for all iterate values $k \geq T_O$, x^k and (y^k, s^k) are an ε_2^*-optimal solution to the primal-dual pair P and D. □

Again, before proving the theorem, we offer the following interpretive remarks. For a fixed duality gap tolerance $\varepsilon_2^* > 0$, T_O is a function only of the quantities n, δ_1, δ_2, and $\hat{x}^T \hat{s}$, and is monotone increasing in all of these quantities. As developed in section 2, $\hat{x}^T \hat{s}$ is also a distance measure from the initial point $(\hat{x}, \hat{y}, \hat{s})$ to the optimal regions $\mathcal{X}^*$ and $\mathcal{S}^*$. Suppose $(\hat{x}^k, \hat{y}^k, \hat{s}^k)$ is a sequence of starting points whose limit is an optimal point (x^*, y^*, s^*), where $x^* \in \mathcal{X}^*$ and $s^* \in \mathcal{S}^*$. Then $(\hat{x}^k)^T \hat{s}^k \to 0$, $\ln(\delta_2^k) \to 0$, and $\limsup_{k \to \infty} \{\delta_1^k\} \leq \sqrt{n}$, and therefore $T_O \to 0$.

We also note that if $\delta_2 = +\infty$, $T_O = +\infty$, as expected.

For a discussion of the ways different "hot" start and "cold" start strategies can affect these complexity bounds, see the discussion in section 8. We now prove these three theorems:

Proof of theorem 5.1

With T_P as given, and $k \geq T_P$, corollary 4.3 implies that the number of iterations in which the algorithm visits step 2 is at least

$$T_2 = \left\lceil (2 + 4\sqrt{n} + 15n\,\delta_1) \ln\left(\frac{\|b - A\hat{x}\|}{\varepsilon_1^*} \right) \right\rceil.$$

Therefore, $\varepsilon^k \leq (\alpha)^{T_2} \varepsilon^0 = (\alpha)^{T_2}$, and

$$\ln(\varepsilon^k) \leq T_2 \ln \alpha \leq T_2(\alpha - 1),$$

from the concavity of the logarithm function. With $\gamma = 1/4$, (4.1) yields

$$\beta = 1 + \frac{1}{4\sqrt{n} - 2} \leq 1.5$$

and

$$\alpha \leq 1 - \frac{1}{2 + 4\sqrt{n} + 15n\,\delta_1}.$$

Therefore,

$$\ln(\varepsilon^k) \leq -\ln\left(\frac{\|b - A\hat{x}\|}{\varepsilon_1^*} \right),$$

and so

$$\varepsilon^k \leq \varepsilon_1^* / \| b - A\hat{x} \|.$$

From theorem 4.1 and (3.3a),

$$x^k \geq 0 \quad \text{and} \quad \| b - Ax^k \| = \varepsilon^k \| b - A\hat{x} \| \leq \varepsilon_1^*. \qquad \square$$

Proof of theorem 5.2

This proof exactly parallels that of theorem 5.1, for the dual problem D. $\square$

Proof of theorem 5.3

Using the same logic as in the proof of Theorem 5.1, with

$$T_2 = \left\lceil (2 + 4\sqrt{n} + 15n\,\delta_1) \left(\ln(\delta_2) + \ln\left(\frac{(\hat{x}^T \hat{s})(1.25)}{\varepsilon_2^*} \right) \right) \right\rceil,$$

we obtain that for $k \geq T_O$,

$$\ln(\varepsilon^k) \leq -\ln(\delta_2) - \ln\left(\frac{(\hat{x}^T \hat{s})(1.25)}{\varepsilon_2^*} \right),$$

and so

$$\varepsilon^k \leq \frac{\varepsilon_2^*}{(\hat{x}^T \hat{s})\delta_2 (1.25)}.$$

From theorem 4.2, $\mu^k \leq \delta_2 \hat{\theta}$, and from proposition 3.1,

$$\begin{aligned} (x^k)^T (s^k) &\leq \varepsilon^k \mu^k n(1+\gamma) \\ &\leq \left(\frac{\varepsilon_2^*}{\hat{x}^T \hat{s}\, \delta_2 (1.25)} \right) (\delta_2 \hat{\theta})(n)(1+\gamma) \\ &= \varepsilon_2^*, \end{aligned}$$

since $\hat{x}^T \hat{s} = n\,\hat{\theta}$ and $\gamma = 0.25$. $\square$

6. Proof of theorem 4.1

We will split theorem 4.1 into the two separate theorems corresponding to when the algorithm branches to step 2 and step 3, respectively. We will therefore prove the following two theorems:

THEOREM 6.1

If the algorithm is at step 1 and $(\bar{x}, \bar{y}, \bar{s}, \bar{\varepsilon}, \bar{\mu})$ is a γ-approximate solution for a given $\gamma \in (0, 1)$, and the algorithm proceeds to step 2, then $(x, y, s, \varepsilon, \mu)$ is a γ-approximate solution.

THEOREM 6.2

If the algorithm is at step 1 and $(\bar{x}, \bar{y}, \bar{s}, \bar{\varepsilon}, \bar{\mu})$ is a γ-approximate solution for a given $\gamma \in (0, 1)$, and the algorithm proceeds to step 3, then $(x, y, s, \varepsilon, \mu)$ is a γ-approximate solution.

Theorem 4.1 is then an immediate consequence of theorems 6.1 and 6.2.

We first lay the foundation for the proof of theorem 6.1. Note that, from the definition of α in (4.1),

$$(1-\alpha)\left(\sqrt{\gamma} + \sqrt{n} + \frac{n\delta_1(1+\gamma+\beta)}{1-\gamma}\right) = \sqrt{\gamma} - \gamma.$$

Dividing by α and rearranging terms yields

$$\sqrt{\gamma} = \frac{\gamma}{\alpha} + \frac{(1-\alpha)}{\alpha}\left(\sqrt{n} + \frac{n\delta_1(1+\gamma+\beta)}{1-\gamma}\right). \tag{6.1}$$

Equation (6.1) will come in handy in the following proposition.

PROPOSITION 6.2

Under the assumptions of theorem 6.1,

$$\|\bar{X}^{-1}d\| \leq \sqrt{\gamma}.$$

Proof

Examining the Newton equations (4.3b) and (4.3c), we see that since A has full rank m, then (d, y) is determined uniquely. After some laborious manipulation, we can use (4.3b) and (4.3c) to assert that d must satisfy:

$$\bar{X}^{-1}d = P\left[e - \frac{1}{\bar{\varepsilon}\bar{\mu}}\bar{X}\bar{S}e\right] - [I-P][\bar{X}^{-1}(\hat{x}-\bar{x})](1-\alpha)\bar{\varepsilon}$$

$$+ P[\bar{X}(\hat{s}-\bar{s})]\left(\frac{1-\alpha}{\alpha\bar{\mu}}\right), \tag{6.2}$$

where

$$P = [I - \overline{X}A^T(A\overline{X}^2A^T)^{-1}A\overline{X}]. \tag{6.3}$$

To see the validity of (6.2), note that d in (6.2) satisfies

$$Ad = A\overline{X}\overline{X}^{-1}d = -A\overline{X}\overline{X}^{-1}(\hat{x} - \tilde{x})(1-\alpha)\bar{\varepsilon} = -\bar{\varepsilon}(1-\alpha)(A\hat{x} - b)$$

(since $A\tilde{x} = b$), which satisfies (4.3b), and that (6.2) implies (4.3c) is satisfied for some y.

Since both P and $[I - P]$ are projection matrices,

$$\|\overline{X}^{-1}d\| \le \left\| e - \frac{1}{\varepsilon\bar{\mu}}\overline{X}\overline{S}e \right\| + \|\overline{X}^{-1}(\hat{x} - \tilde{x})\|(1-\alpha)\bar{\varepsilon} + \|\overline{X}(\hat{s} - \tilde{s})\|\left(\frac{1-\alpha}{\alpha\bar{\mu}}\right). \tag{6.4}$$

$$\begin{aligned}\left\| e - \frac{1}{\varepsilon\bar{\mu}}\overline{X}\overline{S}e \right\| &= \left\| e - \frac{1}{\alpha\bar{\varepsilon}\bar{\mu}}\overline{X}\overline{S}e \right\| = \left\| \frac{1}{\alpha}\left(e - \frac{1}{\bar{\varepsilon}\bar{\mu}}\overline{X}\overline{S}e\right) + \frac{\alpha - 1}{\alpha}e \right\| \\ &\le \frac{1}{\alpha}\left\| e - \frac{1}{\bar{\varepsilon}\bar{\mu}}\overline{X}\overline{S}e \right\| + \frac{1-\alpha}{\alpha}\|e\| \\ &\le \frac{\gamma}{\alpha} + \frac{1-\alpha}{\alpha}\sqrt{n}, \end{aligned} \tag{6.5}$$

where the last inequality follows from (3.3e). Also

$$\begin{aligned}(1-\alpha)\bar{\varepsilon}\|\ \overline{X}^{-1}(\hat{x} - \tilde{x})\| &\le (1-\alpha)\bar{\varepsilon}\|\overline{S}(\hat{x} - \tilde{x})\|\ \frac{1}{\bar{\varepsilon}\bar{\mu}(1-\gamma)} \\ &\le \frac{(1-\alpha)}{\alpha\bar{\mu}(1-\gamma)}\|\overline{S}(\hat{x} - \tilde{x})\| \end{aligned} \tag{6.6}$$

and

$$\left(\frac{1-\alpha}{\alpha\bar{\mu}}\right)\|\overline{X}(\hat{s} - \tilde{s})\| \le \frac{(1-\alpha)}{\alpha(1-\gamma)\bar{\mu}}\|\overline{X}(\hat{s} - \tilde{s})\|. \tag{6.7}$$

Combining (6.4), (6.5), (6.6), (6.7), and proposition 2.3 yields:

$$\begin{aligned}\|\overline{X}^{-1}d\| &\le \frac{\gamma}{\alpha} + \frac{(1-\alpha)}{\alpha}\sqrt{n} + \left(\frac{(1-\alpha)}{\alpha(1-\gamma)\bar{\mu}}\right)\delta_1(\|\hat{X}\bar{s}\| + \|\hat{S}\bar{x}\|) \\ &\le \frac{\gamma}{\alpha} + \frac{(1-\alpha)}{\alpha}\sqrt{n} + \left(\frac{(1-\alpha)}{\alpha(1-\gamma)}\right)\delta_1 n(1+\gamma+\beta) \quad \text{(from step 1 of the algorithm (4.2))} \\ &\le \sqrt{\gamma} \quad \text{(from (6.1)).} \end{aligned}$$

□

Proof of theorem 6.1

Direct substitution shows that with $x = \bar{x} + d$, then

$$Ax = b + \varepsilon(A\hat{x} - b)$$

and $x = \bar{x} + d = \bar{X}(e + \bar{X}^{-1}d) \geq 0$ since $\|\bar{X}^{-1}d\| \leq \sqrt{\gamma} < 1$ from proposition 6.1. Thus, (3.3a) and (3.3b) are satisfied. Also note that, from (4.3c) and (4.3e),

$$A^T y + s = c + \varepsilon(A^T \hat{y} + \hat{s} - c),$$

verifying (3.3c), and $s = \varepsilon\mu\bar{X}^{-1}(e - \bar{X}^{-1}d) > 0$ since $\|\bar{X}^{-1}d\| \leq \sqrt{\gamma} < 1$ from proposition 6.1, verifying (3.3d). From (4.3d) and (4.3e), we obtain

$$\begin{aligned} \frac{x_j s_j}{\varepsilon\mu} - 1 &= (1 - (\bar{X}^{-1}d)_j)(1 + (\bar{X}^{-1}d)_j) - 1 \\ &= -(\bar{X}^{-1}d)_j^2, \end{aligned}$$

and so

$$\begin{aligned} \left\| e - \frac{1}{\varepsilon\mu} XSe \right\| \leq \left\| e - \frac{1}{\varepsilon\mu} XSe \right\|_1 &= \sum_{j=1}^{n} (\bar{X}^{-1}d)_j^2 \\ &= \|\bar{X}^{-1}d\|^2 \leq (\sqrt{\gamma})^2 = \gamma. \end{aligned}$$

This demonstrates (3.3e), and so $(x, y, s, \varepsilon, \mu)$ is a γ-approximate solution. □

PROPOSITION 6.3

Under the assumptions of theorem 6.2,

$$\|\bar{X}^{-1}d\| \leq \sqrt{\gamma}.$$

Proof

First note that from definition of β in (4.1) that

$$(\beta - 1)(\sqrt{n} - \sqrt{\gamma}) = \sqrt{\gamma} - \gamma,$$

and dividing by β and rearranging yields

$$\frac{\gamma}{\beta} + \frac{(\beta - 1)}{\beta}(\sqrt{n}) = \sqrt{\gamma}. \tag{6.8}$$

Equations (4.4b)–(4.4c) can be manipulated to yield

$$\bar{X}^{-1}d = P\left[e - \frac{1}{\varepsilon\mu}\bar{X}\bar{S}e\right],$$

where P is defined in (6.3), and so

$$\begin{aligned}
\|\bar{X}^{-1}d\| &\leq \left\| e - \frac{1}{\varepsilon\mu}\bar{X}\bar{S}e \right\| \\
&= \left\| e - \frac{1}{\beta\bar{\varepsilon}\bar{\mu}}\bar{X}\bar{S}e \right\| = \left\| \frac{1}{\beta}\left(e - \frac{1}{\bar{\varepsilon}\bar{\mu}}\bar{X}\bar{S}e\right) + \frac{\beta-1}{\beta}e \right\| \\
&\leq \frac{1}{\beta}\left\| e - \frac{1}{\bar{\varepsilon}\bar{\mu}}\bar{X}\bar{S}e \right\| + \frac{\beta-1}{\beta}\|e\| \\
&\leq \frac{\gamma}{\beta} + \frac{\beta-1}{\beta}\sqrt{n} \qquad \text{(from (3.3e))} \\
&= \sqrt{\gamma} \qquad \text{(from (6.8)).}
\end{aligned}$$

□

Proof of theorem 6.2

The proof here is almost identical to that of theorem 6.1. Direct substitution shows that

$$Ax = A\bar{x} + Ad = A\bar{x} = b + \bar{\varepsilon}(A\hat{x} - b) = b + \varepsilon(A\hat{x} - b)$$

and

$$A^T y + s = c + \varepsilon(A^T\hat{y} + \hat{s} - c)$$

follows from (4.4b–e). $x > 0$ and $s > 0$ follow as in the proof of theorem 6.1 from the equations (4.4d) and (4.4e) and the fact that $\|\bar{X}^{-1}d\| \leq \sqrt{n} < 1$, as shown in proposition 6.3. Finally, the identical analysis as in the proof of theorem 6.1 shows that

$$\left\| e - \frac{1}{\varepsilon\mu}XSe \right\| \leq \gamma.$$

□

7. Proof of theorem 4.2

We begin the proof of theorem 4.2 by proving three intermediate results. We start with an algebraic equality motivated from the proof of lemma 3.3 of Mizuno [16]:

PROPOSITION 7.1 (see also Mizuno [16])

Suppose $\bar{x}$ and $(\bar{y}, \bar{s})$ are feasible solutions of $P(\bar{\varepsilon})$ and $D(\bar{\varepsilon})$, respectively,

for some $\bar{\varepsilon} \in [0,1]$, and suppose x^* and (y^*, s^*) are optimal solutions of P and D, respectively. Let $\hat{x}$ and $(\hat{y}, \hat{s})$ satisfy (P.1)–(P.3). Then

$$(\bar{\varepsilon}\hat{s} + (1-\bar{\varepsilon})s^*)^T \bar{x} + (\bar{\varepsilon}\hat{x} + (1-\bar{\varepsilon})x^*)^T \bar{s}$$
$$= (\bar{\varepsilon}\hat{s} + (1-\bar{\varepsilon})s^*)^T (\bar{\varepsilon}\hat{x} + (1-\bar{\varepsilon})x^*) + \bar{x}^T \bar{s}. \tag{7.1}$$

Proof

Direct arithmetic substitution establishes that

$$A(\bar{\varepsilon}\hat{x} + (1-\bar{\varepsilon})x^* - \bar{x}) = 0$$

and

$$A^T(\bar{\varepsilon}\hat{y} + (1-\bar{\varepsilon})y^* - \bar{y}) + (\bar{\varepsilon}\hat{s} + (1-\bar{\varepsilon})s^* - \bar{s}) = 0.$$

Therefore,

$$(\bar{\varepsilon}\hat{x} + (1-\bar{\varepsilon})x^* - \bar{x})^T (\bar{\varepsilon}\hat{s} + (1-\bar{\varepsilon})s^* - \bar{s}) = 0,$$

and rearranging proves the result. □

The next lemma can be considered as a generalization of lemma 3.3 of Mizuno [16].

LEMMA 7.2

Suppose $(\bar{x}, \bar{y}, \bar{s}, \bar{\varepsilon}, \bar{\mu})$ are a γ-approximate solution of BP$(\bar{\varepsilon}, \bar{\mu})$. Then

$$(\|\hat{S}\bar{x}\| + \|\hat{X}\bar{s}\|)/\bar{u} \leq n\left(\frac{\hat{\theta}}{\bar{\mu}}\delta_2 + 1 + \gamma\right).$$

Proof

If $\delta_2 = +\infty$, the result is trivial. If $\delta_2 < +\infty$, then P and D have optimal solutions. In light of (2.4), let $x^* \in \mathcal{X}^*$ and $s^* \in S^*$ satisfy $\delta_2 \geq \|\bar{X}^{-1}x^*\|_\infty$ and $\delta_2 \geq \|\bar{S}^{-1}s^*\|_\infty$.

We have

$$\begin{aligned}
\bar{\varepsilon}(\|\hat{S}\bar{x}\| + \|\hat{X}\bar{s}\|) &\leq \bar{\varepsilon}(\|\hat{S}\bar{x}\|_1 + \|\hat{X}\bar{s}\|_1) \\
&= \bar{\varepsilon}\hat{s}^T\bar{x} + \bar{\varepsilon}\hat{x}^T\bar{s} \\
&\leq (\bar{\varepsilon}\hat{s} + (1-\bar{\varepsilon})s^*)^T\bar{x} + (\bar{\varepsilon}\hat{x} + (1-\bar{\varepsilon})x^*)^T\bar{s} \\
&= (\bar{\varepsilon}\hat{s} + (1-\bar{\varepsilon})s^*)^T(\bar{\varepsilon}\hat{x} + (1-\bar{\varepsilon})x^*) + \bar{x}^T\bar{s} \\
&\qquad \text{(from proposition 7.1)} \\
&= \bar{\varepsilon}^2\hat{s}^T\hat{x} + \bar{\varepsilon}(1-\bar{\varepsilon})(\hat{x}^Ts^* + \hat{s}^Tx^*) + \bar{x}^T\bar{s}
\end{aligned}$$

$$= \bar{\varepsilon}^2 n\hat{\theta} + \bar{\varepsilon}(1-\bar{\varepsilon})(\hat{\theta}\, e^T \hat{S}^{-1} s^* + \hat{\theta}\, e^T \hat{X}^{-1} x^*) + \bar{x}^T \bar{s}$$

(from (P.1) – (P.3))

$$\leq \bar{\varepsilon}^2 n\hat{\theta}\,\delta_2 + \bar{\varepsilon}(1-\bar{\varepsilon})\hat{\theta}\, e^T(\hat{S}^{-1} s^* + \hat{X}^{-1} x^*) + n\bar{\mu}\bar{\varepsilon}(1+\gamma)$$

(using $\delta_2 \geq 1$ and proposition 3.1).

However, due to the complementarity of s^* and x^*, at most one component of $(\hat{S}^{-1} s^* + \hat{X}^{-1} x^*)_j$ is nonzero for each j, and each term is bounded by δ_2. Therefore,

$$\bar{\varepsilon}(\|\hat{S}\bar{x}\| + \|\hat{X}\bar{s}\|) \leq \bar{\varepsilon}^2 n\hat{\theta}\,\delta_2 + \bar{\varepsilon}(1-\bar{\varepsilon}) n\hat{\theta}\,\delta_2 + n\bar{\mu}\bar{\varepsilon}(1+\gamma).$$

Dividing by $\bar{\mu}\bar{\varepsilon}$ and simplifying gives the result. □

Proof of theorem 4.2

Note that $\mu^0 = \hat{\theta} \leq \delta_2\hat{\theta}$, since $\delta_2 \geq 1$. Therefore, (4.5) is satisfied at $k = 0$. We now proceed by induction on k. Suppose $\mu^k \leq \delta_2\hat{\theta}$. If at the kth iteration the algorithm branches to step 2, $\mu^{k+1} = \mu^k \leq \delta_2\hat{\theta}$. Therefore, suppose instead that the algorithm branches to step 3. In this case, for the current values $(\bar{x}, \bar{y}, \bar{s}, \bar{\varepsilon}, \bar{\mu})$,

$$n(1+\gamma+\beta) < (\|\hat{S}\bar{x}\| + \|\hat{X}\bar{s}\|)/\bar{\mu} \leq n((\hat{\theta}/\bar{\mu})\delta_2 + 1 + \gamma),$$

where the first inequality follows from step 1 of the algorithm (4.2), and the second inequality is from lemma 7.2. From the above, we obtain

$$\frac{\hat{\theta}\delta_2}{\beta} > \bar{\mu} = \mu^k. \tag{7.2}$$

From (4.4a), we obtain

$$\mu^{k+1} = \beta\mu^k < \hat{\theta}\delta_2.$$

Therefore, by induction, the theorem is proved. □

8. Discussion

8.1. THE ROLE OF THE MEASURE δ_1 IN HOT AND COLD START STRATEGIES

From section 2, and in particular (2.1a–c), δ_1 is defined as the maximum of the two quantities ρ_1 and σ_1, where ρ_1 and σ_1 are the solution to the Euclidean norm minimization problems (2.1a). Tracing the analysis through the paper, we see that the only property of δ_1 that is used is in proposition 2.3, which is used in the proof of theorem 4.1 (shortly following equation (6.7)). Furthermore, in the proof of

proposition 2.3, the only property of δ_1 that is used is the property that δ_1 satisfies $\delta_1 \geq (\hat{x}_j)^{-1}(\hat{x}_j - \tilde{x}_j)$, $\delta_1 \geq (\hat{s}_j)^{-1}(\hat{s}_j - \tilde{s}_j)$, for some $(\tilde{x}, \tilde{y}, \tilde{s})$ that satisfy the equations $A\tilde{x} = b$, $A^T\tilde{y} + \tilde{s} = c$, see (2.5). In fact, we have:

Remark 8.1

The scalar δ_1 that appears in theorems 5.1, 5.2, 5.3, can be replaced by any scalar $\tilde{\delta}_1$ that satisfies the criterion:

$$\tilde{\delta}_1 \geq \|\hat{X}^{-1}(\tilde{x} - \hat{x})\|_\infty, \quad \tilde{\delta}_1 \geq \|\hat{S}^{-1}(\tilde{s} - \hat{s})\|_\infty \tag{8.1}$$

for some $(\tilde{x}, \tilde{y}, \tilde{s})$ satisfying $A\tilde{x} = b$, $A^T\tilde{y} + \tilde{s} = c$. □

Remark 8.1 is important in analyzing and interpreting theorems 5.1, 5.2, and 5.3 as follows.

Let us consider a "cold" start approach to solving the linear program P via the following strategy: find *any* $(\tilde{x}, \tilde{y}, \tilde{s})$ that solves $A\tilde{x} = b$, $A^T\tilde{y} + \tilde{s} = c$. (This can be done easily via Gaussian elimination.) Then set $\hat{x} = \|\tilde{x}\|_\infty e$, $\hat{s} = \|\tilde{s}\|_\infty e$, and $\hat{\theta} = \|\tilde{x}\|_\infty \|\tilde{s}\|_\infty$. Then $\hat{x}$ and $\hat{s}$ satisfy (P.1), (P.2), and (P.3), and $\tilde{\delta}_1 = 2$ will work in (8.1). Then under the reasonable presumption that $\|\tilde{x}\|_\infty \leq 2^L$ and $\|\hat{s}\|_\infty \leq 2^L$ (where L is the bit-size representation of the linear programming data (A, b, c)), theorems 5.1, 5.2, and 5.3 imply that the algorithm is an $O(nL)$ iteration algorithm for solving P, where $\varepsilon_1^* = \varepsilon_2^* \leq 2^{-L}$ and a suitable procedure is used to round to an exact optimal solution at the termination of the algorithm.

This "cold start" strategy sacrifices good initial complementarity slackness $(\hat{x}^T\hat{s} = n\|\hat{x}\|_\infty\|\hat{s}\|_\infty)$ in order to achieve a small value of $\tilde{\delta}_1 = 2$ in theorems 5.2, 5.2, and 5.3. Because a higher complementarity slackness value is not as harmful as a high value of $\tilde{\delta}_1$ in the complexity bound (due to the appearance of $\ln(\hat{x}^T\hat{s})$ in theorem 5.3), this seems reasonable. One interpretation of this "cold start" strategy is that it conservatively assigns a large initial slack to every inequality constraint. That is, this strategy makes no guess regarding any active constraints and assumes all constraints are inactive. In practice, this is a strategy that has worked well in the OB1 interior-point code [13]. Thus, from this loose perspective, the complexity bounds derived in theorems 5.1, 5.2, and 5.3 are consistent with practice.

Now consider instead a "hot start" strategy based on guessing a good starting basis. Such a strategy would assign some initial values $\hat{x}_j$ or $\hat{s}_j$ to be a very small positive number ε. Suppose that the set $\{1,\ldots,n\}$ is partitioned into B and N, where $B \cup N = \{1,\ldots,n\}$ and $B \cap N = \emptyset$, and that A_B is invertible. If $\hat{x}_j = \varepsilon$ for $j \in N$, then the only way to prevent $\|\hat{X}^{-1}(\hat{x} - \tilde{x})\|_\infty = \|e - \hat{X}^{-1}\tilde{x}\|_\infty$ from being very large is to set $\tilde{x}_j \approx \varepsilon$ for $j \in N$. This leaves $\tilde{x}_B \approx A_B^{-1}(b - A_N e_N \varepsilon)$, and $(\hat{X}^{-1}(\tilde{x} - \hat{x}))_B$ could still be quite large, even for reasonably sized $\hat{x}$. This argument indicates a way in which a "hot start" guess of a good basis might fail, causing $\tilde{\delta}_1$ to be very

large. In fact, this type of phenomenon has occurred in practice, where an initial guess of the active constraints has been wrong, and has considerably slowed the progress of the interior-point code OB1 [13].

8.2. DETECTING INFEASIBILITY

Note that an immediate consequence of theorem 4.2 is:

COROLLARY 8.1

Suppose that $\gamma \in (0, 1)$ and that δ_2 is finite. Suppose μ is increased at some iteration. Then for all k,

$$\inf_{x^* \in \mathcal{X}^*} \| \hat{X}^{-1} x^* \|_\infty \geq \mu^k / \hat{\theta}$$

or

$$\inf_{s^* \in \mathcal{S}^*} \| \hat{S}^{-1} s^* \|_\infty \geq \mu^k / \hat{\theta}.$$

Proof

From theorem 4.2, we have $\delta_2 \geq \mu^k / \hat{\theta}$. Note that $\mu^0 = \hat{\theta}$ from (3.5). Then if μ^k ever increases, $\delta_2 > 1$. Thus, from (2.4b), $\delta_2 = \rho_2$ or $\delta_2 = \sigma_2$. Suppose $\delta_2 = \rho_2$. Then $\inf_{x^* \in \mathcal{X}^*} \| \hat{X}^{-1} x^* \|_\infty = \rho_2 = \delta_2 \geq \mu^k / \hat{\theta}$. A parallel argument proves the results when $\delta_2 = \sigma_2$. □

This corollary can be used to detect infeasibility of P or D as follows. Suppose the algorithm visits step 3 a total number of $T_3 = O(\sqrt{n}L)$ times. If each $\hat{x}_j \geq 2^{-L}$, $\hat{s}_j \geq 2^{-L}$, then after T_3 visits to step 3, $\mu^k / \hat{\theta} \geq (\beta)^{T_3} = O(2^L)$, and corollary 8.1 implies that $\inf_{x^* \in \mathcal{X}^*} \| x^* \|_\infty \geq O(2^L)$ or $\inf_{s^* \in \mathcal{S}^*} \| \hat{S}^{-1} s^* \|_\infty \geq O(2^L)$. Since this cannot happen (see [19]), δ_2 cannot be finite, and so the algorithm will detect infeasibility in $O(\sqrt{n}L)$ iterations.

Acknowledgement

I am very grateful to Shinji Mizuno and Masakazu Kojima for stimulating discussions that have provided the background of the results contained herein.

References

[1] K.M. Anstreicher, A combined phase I–phase II projective algorithm for linear programming, Mathematical Programming 43(1989)209–223.

[2] K.M. Anstreicher, A combined phase I–phase II scaled potential algorithm for linear programming, Mathematical Programming 52(1991)429–439.

[3] K.M. Anstreicher, On interior algorithms for linear programming with no regularity assumptions, Operations Research Letters 11(1992)209–212.

[4] R.M. Freund, Theoretical efficiency of a shifted barrier function algorithm for linear programming, Linear Algebra and its Applications 152(1991)19–41.
[5] R.M. Freund, A potential-function reduction algorithm for solving a linear program directly from an infeasible "warm start", Mathematical Programming 52(1991)441–466.
[6] R.M. Freund, A potential reduction algorithm with user-specified phase I– phase II balance, for solving a linear program from an infeasible warm start, to appear in SIAM Journal on Optimization.
[7] R.M. Freund, Following a "balanced" trajectory from an infeasible point to an optimal linear programming solution with a polynomial-time algorithm, to appear in Mathematics of Operations Research.
[8] G. de Ghellinck and J.-P. Vial, A polynomial Newton method for linear programming, Algorithmica 1(1986)425–453.
[9] P.E. Gill, W. Murray, M.A. Saunders and M.H. Wright, Shifted barrier methods for linear programming, Technical Report SOL 88-9, Department of Operations Research, Stanford University (1988).
[10] N. Karmarkar, A new polynomial time algorithm for linear programming, Combinatorica 4 (1984)373–395.
[11] M. Kojima, N. Megiddo and S. Mizuno, A primal-dual infeasible-interior-point algorithm for linear programming, Research Report RJ 8500, IBM Almaden Research Center, San Jose, CA (1991), Mathematical Programming 61(1993)261–280.
[12] I.J. Lustig, R.E. Marsten and D.F. Shanno, Computational experience with a primal-dual interior point method for linear programming, Linear Algebra and its Applications 152(1991)191–222.
[13] I.J. Lustig and D.F. Shanno, private communication (April, 1993).
[14] R. Marsten, R. Subramanian, M. Saltzman, I.J. Lustig and D. Shanno, Interior point methods for linear programming: Just call Newton, Lagrange, and Fiacco and McCormick!, Interfaces 20 (1990)105–116.
[15] J. Mitchell, An interior point column generation method for linear programming using shifted barriers, SIAM Journal on Optimization 4(1994)423–440.
[16] S. Mizuno, Polynomiality of infeasible interior point algorithms for linear programming, Mathematical Programming 67(1994)109–119.
[17] S. Mizuno, M. Kojima and M. J. Todd, Infeasible-interior-point primal-dual potential-reduction algorithms for linear programming, to appear in SIAM Journal on Optimization.
[18] S. Mizuno, M. J. Todd and Y. Ye, A surface of analytic centers and infeasible-interior-point algorithms for linear programming, Mathematics of Operations Research 20(1995)135–162.
[19] C.H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization* (Prentice–Hall, NJ, 1982).
[20] R.A. Polyak, Modified barrier functions (theory and methods), Mathematical Programming 54 (1992)177–222.
[21] F.A. Potra, A quadratically convergent infeasible interior-point algorithm for linear programming, Report No. 28, Dept. of Mathematics, The University of Iowa, Iowa City, IA (1992).
[22] F.A. Potra, An infeasible interior-point predictor-corrector algorithm for linear programming, to appear in SIAM Journal on Optimization.
[23] F.A. Potra, A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points, Mathematical Programming 67(1994)383–406.
[24] C. Roos and J.P. Vial, A polynomial method of approximate centers for linear programming, Mathematical Programming 54(1992)295–305.
[25] M.J. Todd, On Anstreicher's combined phase I–phase II projective algorithm for linear programming, Mathematical Programming 55(1992)1–15.
[26] M.J. Todd, Combining phase I and phase II in a potential reduction algorithm for linear programming, Mathematical Programming 59(1993)133–150.
[27] M.J. Todd and Y. Wang, On combined phase 1–phase 2 projective methods for linear programming, Algorithmica 9(1993)64–83.

[28] P. Tseng, A simple complexity proof for a polynomial-time linear programming algorithm, Operations Research Letters 8(1989)155–159.
[29] J.P. Vial, A projective algorithm for linear programming with no regularity condition, Operations Research Letters 12(1992)1–2.
[30] Y. Ye, M.J. Todd and S. Mizuno, An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm, Mathematics of Operations Research 19(1994)53–67.
[31] Y. Zhang, On the convergence of a class of infeasible interior-point methods for the horizontal linear complementarity problem, SIAM Journal on Optimization 4(1994)208–227.
[32] Y. Zhang, Superlinear convergence of infeasible interior-point methods for linear programming, Mathematical Programming 66(1994)361–377.