SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 944

October 1990

ON ADAPTIVE-STEP PRIMAL-DUAL INTERIOR-POINT ALGORITHMS FOR LINEAR PROGRAMMING

by

S. Mizuno[1], M.J. Todd[2], Y. Ye[3]

[1]Department of Prediction and Control, The Institute of Statistical Mathematics, 4-6-7, Minami-Azabu, Minato-ku, Tokyo 106, JAPAN.

[2]School of OR&IE, Cornell University, Ithaca, NY 14853-3801. Research was supported in part by NSF grant DMS-8904406 and ONR contract N00014-87-K0212 and by NSF, AFORS, and ONR through NSF grant DMS-8920550.

[3]Department of Management Sciences, The University of Iowa, Iowa City, Iowa 52242. Research was supported in part by NSF grant DDM-8922636.

On adaptive-step primal-dual interior-point algorithms for linear programming

Shinji MIZUNO
Department of Prediction and Control
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Michael J. TODD *
School of Operations Research and Industrial Engineering
Cornell University, Ithaca, New York 14853, USA.

Yinyu YE †
Department of Management Sciences
The University of Iowa, Iowa City, Iowa 52242, USA.

October 1990, revised May, November 1991

Abstract: We describe several adaptive-step primal-dual interior point algorithms for linear programming. All have polynomial time complexity while some allow very long steps in favorable circumstances. We provide heuristic reasoning for expecting that the algorithms will perform much better in practice than guaranteed by the worst-case estimates, based on an analysis using a nonrigorous probabilistic assumption.

Key words: Linear Programming, interior point algorithms, path-following algorithms, potential reduction algorithms.

AMS 1980 subject classification. Primary: 90C05.

*Research was supported in part by NSF grant DMS-8904406 and ONR contract N00014-87-K0212 and by NSF, AFORS, and ONR through NSF grant DMS-8920550.

†Research was supported in part by NSF grant DDM-8922636.

1 Introduction

This paper describes and analyzes several primal-dual interior-point algorithms for linear programming. In some sense these methods follow the central path (defined below), but certain of our algorithms allow a very loose approximation to the path. All methods discussed have polynomial worst-case complexity. We also provide heuristic arguments that the algorithms will typically require far fewer iterations than guaranteed by the worst-case bounds. Our reasoning considers a single iteration and is based on a nonrigorous probabilistic assumption on the data at that iteration. We hope that such analyses can be made more rigorous in the future.

We consider linear programming problems in the following standard form:

$$\begin{array}{lll} \text{(P)} & \text{minimize} & c^T x \\ & \text{subject to} & Ax = b,\ x \geq 0, \end{array}$$

where $c \in R^n$, $A \in R^{m \times n}$ and $b \in R^m$ are given data, and $x \in R^n$ is the decision vector. The dual to (P) can be written as

$$\begin{array}{lll} \text{(D)} & \text{maximize} & b^T y \\ & \text{subject to} & A^T y + s = c,\ s \geq 0, \end{array}$$

with variables $y \in R^m$ and $s \in R^n$. The components of s are called dual slacks. Denote by $\mathcal{F}$ the set of all pairs (x, s) such that x is feasible in (P) and s (with some y) in (D). We also denote by $\mathcal{F}^0$ the set of all pairs $(x, s) > 0$ in $\mathcal{F}$. Assume that $\mathcal{F}^0 \neq \phi$. Note that the duality gap associated with $(x, s) \in \mathcal{F}$ is $x^T s = c^T x - b^T y$ for any y with $A^T y + s = c$.

Megiddo [13], Sonnevend [25], and Bayer and Lagarias [2] analyzed the central path which is expressed as

$$\mathcal{C} = \left\{ (x, s) \in \mathcal{F}^0 : Xs = \frac{x^T s}{n} e \right\}$$

in primal-dual form, where e denotes the vector of ones and $X = \text{diag}(x)$ denotes the diagonal matrix with diagonal entries equal to the elements of x. Renegar [22], Gonzaga [6], Vaidya [29], Kojima, Mizuno and Yoshise [9, 10], and Monteiro and Adler [19, 20] proposed algorithms that generate a sequence of points in a neighborhood of the central path $\mathcal{C}$ and move in a direction that tries to find a new central point. These are called central path-following methods. The original primal-dual path-following algorithm presented by Kojima, Mizuno and Yoshise [9] uses the neighborhood

$$\left\{ (x, s) \in \mathcal{F}^0 : Xs \geq 0.5 \frac{x^T s}{n} e \right\}.$$

Given a pair (x^0, s^0) with $(x^0)^T s^0 \leq 2^t$, their algorithm required $O(nt)$ iterations to reach a pair (x^k, s^k) with $(x^k)^T s^k \leq 2^{-t}$ — briefly, to attain precision t. (This implies that an exact solution to a problem (P) with integer data with input length L can be obtained in $O(nL)$ iterations. We will use the precision terminology since we regard (P) as having real data.) Then Kojima, Mizuno, and Yoshise [10] and Monteiro

and Adler [19, 20] modified the algorithm by using the neighborhood

$$\mathcal{N}_2(\beta) = \left\{(x,s) \in \mathcal{F}^0 : \|Xs - \mu e\| \leq \beta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n}\right\}$$

for some $\beta \in (0,1)$ so that it requires at most $O(n^{0.5}t)$ iterations to attain precision t. Here $\|\cdot\|$ without subscript designates the ℓ_2 norm.

In this paper we will first analyze an adaptive-step path-following algorithm that generates a sequence of iterates in $\mathcal{N}_2(1/4)$. Actually, the algorithm has a predictor-corrector form, so that it also generates intermediate iterates in $\mathcal{N}_2(1/2)$. Such a primal-dual predictor-corrector algorithm was first proposed in Sonnevend, Stoer, and Zhao [26]. They take a number of corrector steps after each predictor step to get a point sufficiently close to the central path, while our algorithm requires just one corrector step after each predictor step. (Later papers of Sonnevend, Stoer, and Zhao [27] and Zhao and Stoer [34] provide a more complete analysis of their method and allow less exact centering.) Like that of [26], our algorithm has worst-case complexity $O(n^{0.5}t)$ iterations to attain precision t, since it reduces the duality gap by a factor of $(1 - \Omega(n^{-0.5}))$ at each iteration. However we anticipate that the algorithm will perform better in practice. To justify this, we analyze a single iteration. We show that, under a plausible but nonrigorous probabilistic assumption, the decrease in the duality gap will be $(1-\Omega(n^{-0.25}))$ with probability approaching one as $n \to \infty$. Thus we might argue heuristically that the algorithm will typically require only $O(n^{0.25}t)$ iterations. In contrast, Sonnevend, Stoer, and Zhao [26, 27], who were also trying to improve the worst-case bound of $O(n^{0.5}t)$ iterations, prove a rigorous bound of $O(n^{0.25}t)$ iterations for some special classes of linear programming problems.

Next we consider adaptive-step algorithms generating sequences of iterates in either

$$\mathcal{N}_\infty(\beta) = \left\{(x,s) \in \mathcal{F}^0 : \|Xs - \mu e\|_\infty \leq \beta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n}\right\}$$

or

$$\mathcal{N}_\infty^-(\beta) = \left\{(x,s) \in \mathcal{F}^0 : \|Xs - \mu e\|_\infty^- \leq \beta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n}\right\},$$

for any $\beta \in (0,1)$. Here, for any $z \in R^n$,

$$\|z\|_\infty^- := \|z^-\|_\infty$$

and

$$\|z\|_\infty^+ := \|z^+\|_\infty,$$

where $(z^-)_j := \min\{z_j, 0\}$ and $(z^+)_j := \max\{z_j, 0\}$ and $\|\cdot\|_\infty$ is the usual ℓ_∞ norm. Note that $\|z\|_\infty = \max\{\|z\|_\infty^+, \|z\|_\infty^-\}$ and that neither $\|\cdot\|_\infty^-$ nor $\|\cdot\|_\infty^+$ is a norm, although they obey the triangle inequality. The neighborhood used by Kojima, Mizuno, and Yoshise [9] is $\mathcal{N}_\infty^-(1/2)$ and our algorithms have the same worst-case complexity as theirs. Indeed, our algorithms reduce the duality gap by a factor of $(1 - \Omega(n^{-1}))$ at each iteration, and thus require $O(nt)$ iterations to attain precision t. However, if we analyze a single iteration using the same nonrigorous probabilistic assumption, the decrease of the duality gap will be $(1 - \Omega(1/\log n))$ with probability approaching one as $n \to \infty$.

We easily see that

$$\mathcal{C} \subset \mathcal{N}_2(\beta) \subset \mathcal{N}_\infty(\beta) \subset \mathcal{N}_\infty^-(\beta) \subset \mathcal{F}^0 \text{ for each } \beta \in (0,1).$$

Our results indicate that when we use a wider neighborhood of the central path, the worst-case number of iterations grows, while the typical behavior might be expected to improve.

Our adaptive-step algorithms for the neighborhoods $\mathcal{N}_\infty(\beta)$ and $\mathcal{N}_\infty^-(\beta)$ generate sequences of points lying in the boundaries of these sets. Since our results hold for arbitrary $\beta \in (0,1)$, our algorithms can generate sequences of points in a wide area of the feasible region. In particular,

$$\mathcal{N}_\infty^-(1) = \mathcal{F}^0,$$

so when β is close to 1, the neighborhood $\mathcal{N}_\infty^-(\beta)$ spreads almost all over the feasible region $\mathcal{F}$ and the points generated by the algorithm based on $\mathcal{N}_\infty^-(\beta)$ are close to the boundary rather than the central path. Moreover the search direction of our algorithm approaches that of the original primal-dual affine scaling algorithm (Monteiro, Adler, and Resende [18]) if β goes to 1. Finally, our algorithm approaches the conditions required for fast local convergence given by Zhang, Tapia, and Dennis [33] for β close to 1.

We also describe a potential reduction algorithm. Although the search direction of the potential reduction algorithm is the same as that of the path-following algorithm, the step size is determined as the minimum point of the potential function in the neighborhood $\mathcal{N}_\infty^-(\beta)$. The primal-dual potential function introduced by Todd and Ye [28] is

$$\psi(x,s) = \rho \log(x^T s) - \sum_{j=1}^{n} \log(x_j s_j), \tag{1}$$

where $\rho > n$. Using this, Ye [32], Freund [5], Anstreicher and Bosch [1], and Kojima, Mizuno and Yoshise [8] have developed $O(n^{0.5}t)$-iteration potential reduction algorithms with the choice of $\rho = n + \Theta(n^{0.5})$. However, practical experiments indicate that a larger ρ is much better (McShane et al. [12] and Lustig et al. [11]). We show that if we set $\rho = n + \Theta(n^2)$, the potential reduction algorithm has the same properties as our adaptive-step method based on $\mathcal{N}_\infty^-(\beta)$.

Section 2 describes the basic step on which our algorithms are built and gives some worst-case bounds on the second-order term that arises. In section 3, we describe our first algorithm, an adaptive-step predictor-corrector method based on $\mathcal{N}_2(\beta)$, and analyze its worst-case behavior. Section 4 introduces two adaptive-step algorithms, based on $\mathcal{N}_\infty(\beta)$ and $\mathcal{N}_\infty^-(\beta)$ respectively, and establishes their polynomial-time complexity. In section 5 we describe our potential reduction algorithm in $\mathcal{N}_\infty^-(\beta)$. Section 6 is concerned with a probabilistic analysis of the second-order term, under a strong probabilistic assumption on the data at a particular iteration. If certain inequalities which we prove to hold with high probability under our assumption do indeed hold, we show that the decrease in the duality gap or potential function will be far larger than predicted by the worst case analysis. The final section contains concluding comments.

This paper is based on the two reports [16, 17].

2 Preliminaries

Suppose we have a pair $(x, s) \in \mathcal{N}$, where $\mathcal{C} \subset \mathcal{N} \subset \mathcal{F}^0$. Let γ be a constant in $[0, 1]$. Let μ denote $x^T s/n$. Then we can generate a search direction $d = (d_x, d_s)$ using the primal-dual scaling method of Kojima et al. [9] by solving:

$$\begin{aligned} Sd_x + Xd_s &= \gamma\mu e - Xs, \\ Ad_x &= 0, \\ A^T d_y + d_s &= 0. \end{aligned} \tag{2}$$

Here $X = \mathrm{diag}(x)$, $S = \mathrm{diag}(s)$, and e denotes the vector of ones in R^n. It is well-known that d is the Newton step from (x, s) to find the point on $\mathcal{C}$ with duality gap $n\gamma\mu$ (see, e.g., [19]). Note that $d_x^T d_s = -d_x^T A^T d_y = 0$ from (2). To show the dependence of d on the current pair and parameter γ, we write $d = d(x, s, \gamma)$.

Having obtained the search direction d, we let

$$\begin{aligned} x(\theta) &:= x + \theta d_x, \\ s(\theta) &:= s + \theta d_s. \end{aligned} \tag{3}$$

Note that $Ax(\theta) = b$ for all θ, while if (y, s) is feasible in (D), d_y is given by (2), and $y(\theta) := y + \theta d_y$, then $A^T y(\theta) + s(\theta) = c$. In our adaptive-step algorithms, we will frequently let our next iterate be $(x^+, s^+) = (x(\bar{\theta}), s(\bar{\theta}))$, where $\bar{\theta}$ is as large as possible so that $(x(\theta), s(\theta))$ remains in the neighborhood $\mathcal{N}$ for $\theta \in [0, \bar{\theta}]$.

Let $\mu(\theta) = x(\theta)^T s(\theta)/n$ and $X(\theta) = \mathrm{diag}(x(\theta))$. In order to get bounds on $\bar{\theta}$, we first note that

$$\mu(\theta) = (1 - \theta)\mu + \theta\gamma\mu, \tag{4}$$

$$X(\theta)s(\theta) - \mu(\theta)e = (1 - \theta)(Xs - \mu e) + \theta^2 D_x d_s, \tag{5}$$

where $D_x = \mathrm{diag}(d_x)$. Thus $D_x d_s$ is the second-order term in Newton's method to compute a new point of $\mathcal{C}$. Hence we can usually choose $\bar{\theta}$ larger (and get a larger decrease in the duality gap) if $D_x d_s$ is smaller. In this section we obtain several bounds on the size of $D_x d_s$.

First, it is helpful to reexpress $D_x d_s$. Let

$$\begin{aligned} p &:= X^{-0.5} S^{0.5} d_x, \\ q &:= X^{0.5} S^{-0.5} d_s, \\ r &:= (XS)^{-0.5}(\gamma\mu e - Xs), \end{aligned} \tag{6}$$

and

$$U := \{z : AX^{0.5} S^{-0.5} z = 0\}.$$

Then the system (2) can be rewritten as

$$\begin{aligned} p + q &= r, \\ p \in U, \quad q &\in U^\perp, \end{aligned} \tag{7}$$

where $U^\perp$ denotes the orthogonal complement of U. Thus p is the projection of r on U and q its projection on $U^\perp$, so that $p^T q = 0$. Also

$$D_x d_s = Pq \tag{8}$$

where $P = \mathrm{diag}(p)$.

Lemma 1 *With the notation above,*

a)

$$\|Pq\| \leq \frac{\sqrt{2}}{4}\|r\|^2; \tag{9}$$

b)

$$-\frac{\|r\|^2}{4} \leq p_j q_j \leq \frac{r_j^2}{4} \quad \textit{for each} \quad j;$$

c)

$$\begin{aligned}
\|Pq\|_\infty^- &\leq \frac{\|r\|^2}{4} \leq \frac{n\|r\|_\infty^2}{4}, \\
\|Pq\|_\infty^+ &\leq \frac{\|r\|_\infty^2}{4}, \quad \textit{and} \\
\|Pq\|_\infty &\leq \frac{\|r\|^2}{4} \leq \frac{n\|r\|_\infty^2}{4}.
\end{aligned}$$

Proof. a) is proved in Mizuno [15]. See also [10, 20]. The right hand inequality in (b) follows from $p_j + q_j = r_j$ for each j. For the left hand inequality,

$$\begin{aligned}
p_j q_j &\geq \sum_{p_i q_i < 0} p_i q_i \\
&= -\sum_{p_i q_i \geq 0} p_i q_i \quad (\text{since} \quad p^T q = 0) \\
&\geq -\sum_{p_i q_i \geq 0} \frac{r_i^2}{4} \\
&\geq -\|r\|^2/4.
\end{aligned}$$

(c) is an immediate consequence of (b). □

The bounds in lemma 1 cannot improved by much in the worst case: consider the case where

$$\begin{aligned}
r &= e = (1, 1, \cdots, 1)^T, \\
p &= (1/2, 1/2, \cdots, 1/2, (1+\sqrt{n})/2)^T, \quad \text{and} \\
q &= (1/2, 1/2, \cdots, 1/2, (1-\sqrt{n})/2)^T.
\end{aligned}$$

To use lemma 1 we also need to bound r. The following result is useful:

Lemma 2 *Let r be as above.*

a) *If $\gamma = 0$, then $\|r\|^2 = n\mu$.*

b) *If $\beta \in (0,1)$, $\gamma = 1$ and $(x,s) \in \mathcal{N}_2(\beta)$, then $\|r\|^2 \leq \beta^2\mu/(1-\beta)$.*

c) *If* $\beta \in (0,1)$, $\gamma \in (0,1)$, $\gamma \leq 2(1-\beta)$ *and* $(x,s) \in \mathcal{N}_\infty^-(\beta)$, *then* $\|r\|^2 \leq n\mu$. *Moreover, if* $(x,s) \in \mathcal{N}_\infty(\beta)$ *then* $\sqrt{1-\beta}\sqrt{\mu} \geq (\gamma/\sqrt{1-\beta} - \sqrt{1-\beta})\sqrt{\mu} \geq r_j \geq (\gamma/\sqrt{1+\beta} - \sqrt{1+\beta})\sqrt{\mu} \geq -\sqrt{1+\beta}\sqrt{\mu}$, *so* $\|r\|_\infty^2 \leq (1+\beta)\mu$.

Proof. a) If $\gamma = 0$, $r = -(XS)^{-0.5}Xs$, so $\|r\|^2 = x^T s = n\mu$.

b) Now $r = (XS)^{-0.5}(\mu e - Xs)$, so $\|r\| \leq \frac{1}{\sqrt{(1-\beta)\mu}}\beta\mu$, which yields the desired result.

c) In this case

$$\begin{aligned}
\|r\|^2 &= \sum_{j=1}^{n} \frac{(\gamma\mu - x_j s_j)^2}{x_j s_j} \\
&= \sum_{j=1}^{n} \left(\frac{(\gamma\mu)^2}{x_j s_j} - 2\gamma\mu + x_j s_j \right) \\
&\leq \frac{n(\gamma\mu)^2}{(1-\beta)\mu} - 2n\gamma\mu + n\mu \quad (\text{since} \quad x_j s_j \geq (1-\beta)\mu) \\
&\leq n\mu \quad (\text{since} \quad \gamma \leq 2(1-\beta)).
\end{aligned}$$

Now suppose $(x,s) \in \mathcal{N}_\infty(\beta)$, so that $x_j s_j \in [(1-\beta)\mu, (1+\beta)\mu]$ for each j. Thus, for each j,

$$\frac{\gamma\mu}{\sqrt{1-\beta}\sqrt{\mu}} - \sqrt{1-\beta}\sqrt{\mu} \geq r_j = \frac{\gamma\mu}{\sqrt{x_j s_j}} - \sqrt{x_j s_j} \geq \frac{\gamma\mu}{\sqrt{1+\beta}\sqrt{\mu}} - \sqrt{1+\beta}\sqrt{\mu},$$

which yields the final result. □

The algorithm of Section 3 uses parts (a) and (b) of this lemma, while those of Section 4 and 5 use part (c). Note that the proof of (c) shows that, if $(x,s) \in \mathcal{N}_\infty(\beta)$ and $\gamma \leq 1-\beta$, then $r_j \leq 0$ for each j, so that at least one of $(d_x)_j$ and $(d_s)_j$ is nonpositive for each j.

3 A predictor-corrector algorithm

In this section we describe and analyze an algorithm that takes a single "corrector" step to the central path after each "predictor" step to decrease μ. Although it is possible to use more general values of β, we will work with (nearly-centered) pairs in $\mathcal{N}_2(1/2)$ (intermediate iterates after a predictor step) and (very nearly centered) pairs in $\mathcal{N}_2(1/4)$ (after the corrector step).

Assume we have an initial pair $(x^0, s^0) \in \mathcal{N}_2(1/4)$ with $(x^0)^T s^0 \leq 2^t$. Many papers (e.g. [9, 10, 19, 20]) describe how to modify (P) and (D) so that such an initial pair can be found.

Algorithm 1 *Given* $(x^0, s^0) \in \mathcal{N}_2(1/4)$ *with* $(x^0)^T s^0 \leq 2^t$, *set* $k = 0$.

While $(x^k)^T s^k > 2^{-t}$ *do*

begin

set $(x,s) = (x^k, s^k)$;

(predictor step)

compute $d = d(x, s, 0)$ *from (2);*

compute the largest $\bar{\theta}$ *so that*

$$(x(\theta), s(\theta)) \in \mathcal{N}_2(1/2) \quad \textit{for} \quad \theta \in [0, \bar{\theta}];$$

set $(x', s') = (x(\bar{\theta}), s(\bar{\theta}))$;

(corrector step)

compute $d' = d(x', s', 1)$ *from (2);*

set $(x^{k+1}, s^{k+1}) = (x' + d'_x, s' + d'_s)$;

$k = k + 1$;

end.

This algorithm is very similar in spirit to algorithm 1 of Sonnevend, Stoer, and Zhao [26] – they take a fixed number of corrector steps after each predictor step to attain (to machine precision, say) a pair on the central path. The predictor step, with $\gamma = 0$, is exactly along the search direction of the primal-dual affine scaling algorithm of Monteiro, Adler and Resende [18]. The choice of the parameter $\bar{\theta}$ makes this an adaptive-step algorithm. Similar stepsize selection rules are discussed in Mizuno et al. [14], Sonnevend et al. [26] and Ye [31] and were alluded to in Section 6 of [19]. We can compute $\bar{\theta}$ by solving a quartic equation in one variable.

To analyze this method, we start by showing

Lemma 3 *For each* k, $(x^k, s^k) \in \mathcal{N}_2(1/4)$.

Proof. The claim holds for $k = 0$ by hypothesis. For $k > 0$, let (x', s') be the result of the predictor step at the kth iteration and let $d' = d(x', s', 1)$, as in the description of the algorithm. Let $x'(\theta)$ and $s'(\theta)$ be defined as in (3) and p', q' and r' as in (6) using x', s' and d'. Let $\mu'(\theta) := x'(\theta)^T s'(\theta)/n$ for all $\theta \in [0, 1]$ with $\mu' := \mu'(0) = (x')^T s'/n$ and $\mu^{k+1} := \mu'(1) = (x^{k+1})^T s^{k+1}/n$.

From (4),

$$\mu'(\theta) = \mu' \quad \text{for all} \quad \theta, \tag{10}$$

and in particular $\mu^{k+1} = \mu'$. From (5),

$$\begin{aligned} X'(\theta)s'(\theta) - \mu'(\theta)e &= (1-\theta)(X's' - \mu'e) + \theta^2 D'_x d'_s \\ &= (1-\theta)(X's' - \mu'e) + \theta^2 P'q', \end{aligned} \tag{11}$$

where $X'(\theta) = \mathrm{diag}(x'(\theta))$, etc. But by lemma 1(a) and lemma 2(b) with $\beta = 1/2$,

$$\|P'q'\| \leq \frac{\sqrt{2}}{4}\|r'\|^2 \leq \frac{\sqrt{2}}{4}\frac{(1/2)^2}{1-1/2}\mu' < \frac{1}{4}\mu'.$$

It follows that

$$\|X'(\theta)s'(\theta) - \mu' e\| \leq (1-\theta)\frac{\mu'}{2} + \theta^2\frac{\mu'}{4} \leq \frac{1}{2}\mu'. \tag{12}$$

Thus $X'(\theta)s'(\theta) \geq \frac{\mu'}{2}e > 0$ for all $\theta \in [0,1]$, and this implies that $x'(\theta) > 0$, $s'(\theta) > 0$ for all such θ by continuity. In particular, $x^{k+1} > 0$, $s^{k+1} > 0$, and (13) gives $(x^{k+1}, s^{k+1}) \in \mathcal{N}_2(1/4)$ as desired on setting $\theta = 1$. □

Now let $(x,s) = (x^k, s^k)$, $d = d(x,s,0)$, $\mu = \mu^k = x^T s/n$, and p, q and r be as in (6); these quantities all refer to the predictor step at iteration k. By (4),

$$\mu' = (1-\bar{\theta})\mu, \quad \text{or}$$

$$\mu^{k+1} = (1-\bar{\theta})\mu^k. \tag{13}$$

Hence the improvement in the duality gap at the kth iteration depends on the size of $\bar{\theta}$.

Lemma 4 *With the notation above, the stepsize in the predictor step satisfies*

$$\bar{\theta} \geq \theta_1 := \min\left\{\frac{1}{2}, \left(\frac{\mu}{8\|Pq\|}\right)^{0.5}\right\}.$$

Proof. By (5) applied to the predictor step,

$$\begin{aligned}
\|X(\theta)s(\theta) - \mu(\theta)e\| &= \|(1-\theta)(Xs - \mu e) + \theta^2 Pq\| \\
&\leq (1-\theta)\|Xs - \mu e\| + \theta^2\|Pq\| \\
&\leq \frac{1}{4}(1-\theta)\mu + \theta^2\|Pq\|,
\end{aligned}$$

using lemma 3. For $\theta \leq \theta_1$, $\theta^2\|Pq\| \leq \mu/8$, so

$$\begin{aligned}
\|X(\theta)s(\theta) - \mu(\theta)e\| &\leq \frac{1}{4}(1-\theta)\mu + \mu/8 \\
&\leq \frac{1}{4}(1-\theta)\mu\left(1 + \frac{1}{2(1-\theta)}\right) \\
&\leq \frac{1}{2}(1-\theta)\mu \\
&= \frac{1}{2}\mu(\theta),
\end{aligned}$$

using also $\theta_1 \leq 1/2$. Hence, as in the proof of lemma 3, $(x(\theta), s(\theta)) \in \mathcal{N}_2(1/2)$ for $\theta \leq \theta_1$, whence $\bar{\theta} \geq \theta_1$ as desired. □

We can now show

Theorem 1 *Algorithm 1 terminates in $O(n^{0.5}t)$ iterations.*

Proof. Using lemma 1(a) and lemma 2(a), we have

$$\|Pq\| \leq \frac{\sqrt{2}}{4}\|r\|^2 = \frac{\sqrt{2}}{4}n\mu,$$

so that $\theta_1 \geq 8^{-0.25}n^{-0.5}$ at each iteration. Then (13) and lemma 4 imply that

$$\mu^{k+1} \leq (1 - 8^{-0.25}n^{-0.5})\mu^k$$

for each k. This yields the desired result. □

Note that, if $\|Pq\|$ and $\|P'q'\|$ are typically much smaller than the bound given by lemma 1, then the improvement will be far greater. From (10) and (11), each iterate will be much better centered than is guaranteed by lemma 3, and the predictor step will be much larger than $O(n^{-0.5})$ by lemma 4. We will provide heuristic arguments for this improvement in section 6. For now we merely note

Corollary 1 *If at a particular iteration we have $\|Pq\| \leq n^{0.5}\mu$, then the duality gap at that iteration will decrease at least by a factor of $(1 - 8^{-0.5}n^{-0.25})$. Algorithm 1 requires only $O(n^{0.25}t)$ iterations of this type.*

Proof. This follows immediately from lemma 4 and (13). □

As stated in the introduction, Sonnevend, Stoer and Zhao [26, 27] establish a rigorous bound of $O(n^{0.25}t)$ iterations for their "strongly-centered" method applied to certain special classes of linear programming problems.

Most path-following algorithms (like those of the next section) take only predictor steps. Then, in order to achieve a worst-case complexity of $O(n^{0.5}t)$ iterations, they are forced to set $\gamma = 1 - \Theta(n^{-0.5})$, which limits their possible speed of convergence. We set $\gamma = 0$ to allow greater improvement, but then, in order to have a reasonable step, must enlarge the neighborhood of the central path for the result of the predictor step. A corrector step is then necessary. It seems that ℓ_2-neighborhoods are required in such algorithms. Lemma 1(c) shows that $\|Pq\|_\infty$ can only be bounded by a multiple of $\|r\|^2$, not $\|r\|^2_\infty$, unless an extra factor of n is introduced. But $\|r\|$ may be large compared to $\|r\|_\infty$, which is related to β with $(x, s) \in \mathcal{N}_\infty(\beta)$. Hence corrector steps may not behave well with the ℓ_∞ norm.

4 Algorithms in wider neighborhoods

Let $\mathcal{N}$ be a neighborhood of $\mathcal{C}$ with $\mathcal{C} \subset \mathcal{N} \subset \mathcal{F}^0$. In this section we consider algorithms of the following form based on $\gamma \in (0, 1)$ and $\mathcal{N}$:

Algorithm 2 *Given $(x^0, s^0) \in \mathcal{N}$ with $(x^0)^T s^0 \leq 2^t$, set $k = 0$;*

While $(x^k)^T s^k > 2^{-t}$ do

begin

set $(x, s) = (x^k, s^k)$;

compute $d = d(x, s, \gamma)$ from (2);

compute the largest $\bar{\theta}$ so that

$$(x(\theta), s(\theta)) \in \mathcal{N} \quad \text{for} \quad \theta \in [0, \bar{\theta}];$$

set $(x^{k+1}, s^{k+1}) = (x(\bar{\theta}), s(\bar{\theta}))$;

$k = k + 1$;

end.

For γ close to 0, the search direction approximates that of Monteiro, Adler and Resende's primal-dual affine scaling method [18]. Again the selection of $\bar{\theta}$ makes this an adaptive-step method. We will analyze this algorithm for $\mathcal{N} = \mathcal{N}_\infty(\beta)$ and $\mathcal{N}_\infty^-(\beta)$, where $\beta \in (0,1)$. In either case, computing $\bar{\theta}$ involves the solution of at most $2n$ single-variable quadratic equations.

Note that, if $\mu^k := (x^k)^T s^k/n$, (4) implies

$$\mu^{k+1} = (1 - \bar{\theta}(1-\gamma))\mu^k, \tag{14}$$

so we wish to bound $\bar{\theta}$ from below.

Lemma 5 *Let $\beta \in (0,1)$, $\gamma \in (0,1)$, and $\mathcal{N} = \mathcal{N}_\infty(\beta)$ or $\mathcal{N}_\infty^-(\beta)$. Let x, s, d and $\bar{\theta}$ be as in the kth iteration of Algorithm 2, and define p, q and r by (6). Then*

$$\bar{\theta} \geq \theta_2 := \min\left\{1, \frac{\beta\gamma\mu^k}{\|Pq\|_\infty}\right\} \quad \text{if } \mathcal{N} = \mathcal{N}_\infty(\beta),$$

$$\bar{\theta} \geq \theta_2^- := \min\left\{1, \frac{\beta\gamma\mu^k}{\|Pq\|_\infty^-}\right\} \quad \text{if } \mathcal{N} = \mathcal{N}_\infty^-(\beta).$$

Proof. Suppose first $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$. Then, for each $\theta \in [0, \theta_2^-]$, (4) and (5) imply

$$\begin{aligned} X(\theta)s(\theta) - \mu(\theta)e &= (1-\theta)(Xs - \mu e) + \theta^2 Pq, \\ &\geq -\left((1-\theta)\|Xs - \mu e\|_\infty^- + \theta^2\|Pq\|_\infty^-\right) e \\ &\geq -\left((1-\theta)\beta\mu^k + \theta\beta\gamma\mu^k\right) e \\ &= -\beta\mu(\theta)e. \end{aligned}$$

Hence, as in the proof of lemma 3, $(x(\theta), s(\theta)) \in \mathcal{N}_\infty^-(\beta)$ for $\theta \in [0, \theta_2^-]$, whence $\bar{\theta} \geq \theta_2^-$. If $\mathcal{N} = \mathcal{N}_\infty(\beta)$, a similar proof gives

$$\beta\mu(\theta)e \geq X(\theta)s(\theta) - \mu(\theta)e \geq -\beta\mu(\theta)e$$

for $\theta \in [0, \theta_2]$, which again implies $\bar{\theta} \geq \theta_2$. □

We can now prove

Theorem 2 *Let $\beta \in (0,1)$ and $\gamma \in (0,1)$ be constants with $\gamma \leq 2(1-\beta)$. Then Algorithm 2, with $\mathcal{N} = \mathcal{N}_\infty(\beta)$ or $\mathcal{N}_\infty^-(\beta)$, will terminate in $O(nt)$ iterations.*

Proof. In either case, each iterate lies in $\mathcal{N}_\infty^-(\beta)$, whence $\|Pq\|_\infty^- \leq \|Pq\|_\infty \leq \frac{\|r\|^2}{4} \leq \frac{n\mu^k}{4}$, using lemma 1(c) and lemma 2(c). Hence $\theta_2^- \geq \theta_2 \geq \frac{4\beta\gamma}{n}$. Then lemma 5 and (14) give

$$\mu^{k+1} \leq \left(1 - \frac{4\beta\gamma(1-\gamma)}{n}\right)\mu^k, \tag{15}$$

which yields the result. □

As in Section 3, we note that, if $\|Pq\|_\infty$ or $\|Pq\|_\infty^-$ is typically much smaller than is guaranteed by lemma 1(c), then much greater improvements can be made at that iteration. In fact, we have

Corollary 2 *Let β and γ be as in theorem 2. Suppose at some iteration we have $\|Pq\|_\infty \leq \log(n)\mu^k$ if $\mathcal{N} = \mathcal{N}_\infty(\beta)$ and $\|Pq\|_\infty^- \leq \log(n)\mu^k$ if $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$. Then the duality gap at that iteration will decrease at least by a factor $(1 - \beta\gamma(1-\gamma)/\log(n))$ using Algorithm 2 with either $\mathcal{N}_\infty(\beta)$ or $\mathcal{N}_\infty^-(\beta)$. The algorithm requires only $O(\log(n)t)$ iterations of this type.*

Proof. This follows immediately from lemma 5 and (14). □

It is possible to use $\mathcal{N} = \mathcal{N}_2(\beta)$ in algorithm 2, and also achieve an $O(nt)$ bound on the number of iterations. However, as we shall see in Section 6, the improvement of a typical value for $\|Pq\|$ compared to its worst-case bound may be only $O(n^{-0.5})$, so we are unlikely to be able to improve on algorithm 1 in this case.

5 A potential reduction algorithm

Algorithms 1 and 2 both used the longest step size that would keep the iterate in a certain neighborhood of the central path, and proved convergence using the reduction of the duality gap. Here we describe a variation of Algorithm 2 that uses a primal-dual potential function to choose the step size and to prove convergence. We shall see that we get parallel results to those of section 4.

The primal-dual potential function (1) can be written as

$$\psi(x,s) = (\rho - n)\log(x^T s) - \sum_{j=1}^{n} \log\left(\frac{x_j s_j}{x^T s/n}\right) + n\log n.$$

By the arithmetic-geometric mean inequality, the sum on the right hand side is nonpositive, so

$$\psi(x,s) \geq (\rho - n)\log(x^T s) + n\log n.$$

Hence, if we reduce ψ to $-(\rho - n)t + n\log n$, we shall have $x^T s \leq 2^{-t}$ as desired.

Assume we have $(x^0, s^0) \in \mathcal{F}^0$ such that $(x^0, s^0) \in \mathcal{N}$ and $\psi(x^0, s^0) \leq (\rho - n)t + n\log n$. We will give an algorithm that decreases ψ by a fixed amount depending on n each iteration. This then bounds the number of iterations required.

Algorithm 3 *As algorithm 2 except that $\bar{\theta}$ is chosen so that $(x(\bar{\theta}), s(\bar{\theta})) \in \mathcal{N}$ and*

$$\psi(x(\bar{\theta}), s(\bar{\theta})) \leq \psi(x(\theta), s(\theta)) \quad \textit{for each} \quad \theta \quad \textit{with} \quad (x(\theta), s(\theta)) \in \mathcal{N}.$$

Theorem 3 *Let β and γ be as in theorem 2. Let $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$ and*

$$\rho := n + \left(\frac{3}{\beta\gamma(1-\gamma)}\log\frac{1}{1-\beta}\right)n^2,$$

Then Algorithm 3 will terminate in $O(nt)$ iterations.

Proof. We show that $\theta = \theta_2^-$ defined in lemma 5 achieves a suitable reduction in ψ. Let x, s, μ be as in the kth iteration. Then we have:

$$\begin{aligned}
&\psi(x(\bar{\theta}), s(\bar{\theta})) - \psi(x, s) \\
&\leq \psi(x(\theta_2^-), s(\theta_2^-)) - \psi(x, s) \quad \text{(by lemma 5 and the definition of } \bar{\theta}\text{)} \\
&= (\rho - n)\log\frac{x(\theta_2^-)^T s(\theta_2^-)}{x^T s} - \sum_{j=1}^{n}\log\frac{x_j(\theta_2^-)s_j(\theta_2^-)}{x(\theta_2^-)^T s(\theta_2^-)} + \sum_{j=1}^{n}\log\frac{x_j s_j}{x^T s} \\
&= (\rho - n)\log\frac{\mu(\theta_2^-)}{\mu} - \sum_{j=1}^{n}\log\frac{x_j(\theta_2^-)s_j(\theta_2^-)}{\mu(\theta_2^-)} + \sum_{j=1}^{n}\log\frac{x_j s_j}{\mu} \\
&\leq (\rho - n)\log(1 - (1-\gamma)\theta_2^-) - \sum_{j=1}^{n}\log(1-\beta) \\
&\qquad \text{(from (4) and } (x(\theta_2^-), s(\theta_2^-)) \in \mathcal{N}_\infty^-(\beta)\text{)} \\
&\leq -\left(\frac{3n^2\theta_2^-}{\beta\gamma} - n\right)\log\frac{1}{1-\beta}. \qquad (16)
\end{aligned}$$

As in the proof of theorem 2, $\theta_2^- \geq \frac{4\beta\gamma}{n}$, whence $\psi(x(\bar{\theta}), s(\bar{\theta})) - \psi(x, s) \leq -11n\log\frac{1}{1-\beta}$. Since we need a decrease in ψ of $O(n^2 t)$ ($\rho - n$ is of order n^2), $O(nt)$ iterations suffice. □

As in section 4, we complement this worst-case result with

Corollary 3 *Let β and γ be as in theorem 2. Suppose at some iteration we have $\|Pq\|_\infty^- \leq \log(n)\mu^k$. Then ψ will decrease at least $\left(\frac{3n^2}{\beta\gamma\log n} - n\right)\log\frac{1}{1-\beta}$ at that iteration. The algorithm requires at most $O(\log(n)t)$ iterations of this type.*

Proof. Follows from lemma 5 and (16). □

6 Anticipated improved behavior

In section 3 to 5 we have introduced three algorithms, with worst-case complexities $O(n^{0.5}t)$ or $O(nt)$ iterations to attain precision t (Theorem 1 to 3). We have also stated corollaries indicating that better behavior will be obtained at any iteration where Pq is substantially smaller than what is guaranteed by a worst-case analysis. In this section we provide heuristic arguments why we might expect $\|Pq\|$, $\|Pq\|_\infty$ or $\|Pq\|_\infty^-$ to be of the sizes stated in these corollaries.

Recall that p and q are the projections of $r \in R^n$ onto the subspaces U and $U^\perp$ respectively. In this section we suppose r is fixed, but assume that

> U is a random subspace of R^n of dimension $d := n - m$, drawn from the unique distribution on such subspaces that is invariant under orthogonal transformations.

Given that U is the null space of $AX^{0.5}S^{-0.5} =: \tilde{A}$, this assumption would hold, for example, if each entry of the matrix $\tilde{A}$ were independently drawn from a standard normal distribution. Note that such assumptions, made at different iterations and hence values of X and S, are not consistent with one another. Further, for several interior-point algorithms the asymptotic behavior of (x^k, s^k) is known, and this behavior is also inconsistent with our assumption. We will comment further on our approach in the concluding section. For now, we examine the consequences on Pq of our assumption. Note that, to compensate for the deficiencies of our assumption, the results we obtain hold with probability approaching one as $n \to \infty$.

We now have

Theorem 4 *With the assumption above, if* $\rho = \|r\|_\infty / \|r\|$, *then*

$$\Pr\left(\|Pq\| \leq \frac{\|r\|^2}{4} \left(2\rho^2 + \frac{6.5}{n} \right)^{0.5} \right) \to 1 \quad as \quad n \to \infty.$$

Theorem 5 *With the assumption above,*

$$\Pr\left(\|Pq\|_\infty^- \leq (\log(n)/n)\|r\|^2 \right) \to 1 \quad as \quad n \to \infty.$$

Before we show how these results are proved, we indicate how they relate to the bounds on Pq that form the hypotheses of corollaries 1 to 3.

In corollary 1, we are analyzing the predictor step, so $r = -(XS)^{0.5}e$ and $(x, s) \in \mathcal{N}_2(1/4)$. Hence $\|r\|^2 = x^T s = n\mu$ and $\|r\|_\infty^2 = \|Xs\|_\infty \leq \mu + \|Xs - \mu e\| \leq \frac{5}{4}\mu$. Thus $\rho^2 \leq 5/(4n)$ and by theorem 4, with probability approaching 1

$$\|Pq\| \leq \frac{\|r\|^2}{4} \left(2\frac{5}{4n} + \frac{6.5}{n} \right)^{0.5} = \frac{3\|r\|^2}{4n^{0.5}} < n^{0.5}\mu,$$

which is the hypothesis of corollary 1.

For corollary 2, we consider first the case where $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$. Then by theorem 5 and lemma 2(c), with probability approaching 1

$$\|Pq\|_\infty^- \leq (\log(n)/n)\|r\|^2 \leq \log(n)\mu^k,$$

which gives the hypothesis of corollary 2 in this case. The same argument applies for corollary 3. Now suppose $\mathcal{N} = \mathcal{N}_\infty(\beta)$. Then with high probability

$$\|Pq\|_\infty^- \leq \log(n)\mu^k$$

as above. Also, by lemma 1(c) and lemma 2(c),

$$\|Pq\|_\infty^+ \leq \frac{\|r\|_\infty^2}{4} \leq \frac{(1+\beta)\mu^k}{4} \leq \frac{\mu^k}{2}.$$

Hence $\|Pq\|_\infty \leq \log(n)\mu^k$ with probability approaching 1, which gives the hypothesis for corollary 2 with $\mathcal{N} = \mathcal{N}_\infty(\beta)$.

Now we indicate the proofs of theorems 4 and 5. The proof of the former is long and technical and hence we omit it here. See Mizuno, Todd and Ye [16]. (A slightly weaker version of the theorem is proved at the end of the section.) However, we will give the first part of the proof, because it is also used in establishing theorem 5.

Because p and q are homogeneous of degree 1 in $\|r\|$, we assume henceforth without loss of generality that r is scaled so that

$$g = r/2 \text{ satisfies } \|g\| = 1.$$

Let $F = (g, H)$ be an orthogonal $n \times n$ matrix. If we express the vector p in terms of the basis consisting of the columns of F, we get

Lemma 6 *We can write*

$$p = (1+\zeta)g + \eta H v, \tag{17}$$

where

$\frac{1+\zeta}{2}$ *has a beta distribution with parameters* $\frac{d}{2}$ *and* $\frac{m}{2}$*;*
$\eta = \sqrt{1-\zeta^2}$*; and*
v *is uniformly distributed on the unit sphere in* R^{n-1}*.*

Proof. Since p and q are orthogonal with $p+q=r$, p lies on the sphere of center $r/2 = g$ and radius $\|g\| = 1$. Thus p can be written in the form (17), with $\eta = \sqrt{1-\zeta^2}$ and $\|v\| = 1$. We need to establish that ζ and v have the given distributions.

Note that $\|p\|^2 = (1+\zeta)^2 + \eta^2 = 2(1+\zeta)$. However, we can obtain the distribution of $\|p\|^2$ directly. The invariance under orthogonal transformations implies that we can alternatively take U as a fixed d-subspace, say $\{x \in R^n : x_{d+1} = \cdots = x_n = 0\}$, and r uniformly distributed on a sphere of radius 2. Then r can be generated as

$$\left(\frac{2\lambda_1}{\|\lambda\|}, \frac{2\lambda_2}{\|\lambda\|}, \cdots, \frac{2\lambda_n}{\|\lambda\|}\right)^T,$$

where $\lambda \sim N(0, I)$ in R^n (i.e., the components of λ are independent normal random variables with mean 0 and variance 1). But then

$$p = \left(\frac{2\lambda_1}{\|\lambda\|}, \frac{2\lambda_2}{\|\lambda\|}, \cdots, \frac{2\lambda_d}{\|\lambda\|}, 0, \cdots, 0\right)^T,$$

and $\|p\|^2 = 4(\lambda_1^2 + \cdots + \lambda_d^2)/(\lambda_1^2 + \cdots + \lambda_n^2)$. This has the distribution of four times a beta random variable with parameter $\frac{d}{2}$ and $\frac{m}{2}$ (see, e.g., Wilks [30]), which confirms the distribution of ζ.

Now let W be an orthogonal matrix with $Wg = g$. W can be thought of as rotating the sphere with center g around its diameter from 0 to $2g = r$. We can view the random d-subspace U as the null space of an $m \times n$ random matrix $\bar{A}$ with independent standard normal entries. The fact that p is the projection of r onto U is then equivalent to $\bar{A}p = 0$, $r - p = \bar{A}^T v$ for some v. But then $(\bar{A}W^T)Wp = 0$ and $r - Wp = Wr - Wp = (\bar{A}W^T)^T v$, so that Wp is the projection of r onto $U' = \{x : (\bar{A}W^T)x = 0\}$. If $\bar{A}$ has independent standard normal entries, so does $\bar{A}W^T$, so U' is also a random d-subspace. Thus Wp has the

same distribution as p. But writing W as $HW'H^T + gg^T$, where W' is an arbitrary orthogonal matrix of order $n-1$, we see that v has the same distribution as $W'v$. Since $\|v\| = 1$, v is uniformly distributed on the unit sphere R^{n-1}. □

Since $p + q = r = 2g$, (17) implies

$$\begin{aligned} q &= (1-\zeta)g - \eta Hv, \text{ so that} \\ Pq &= \eta^2 g^2 - 2\zeta\eta GHv - \eta^2 (Hv)^2 \qquad (18) \\ &= -(Hv)^2 + (\eta g - \zeta Hv)^2 \\ &\geq -\|Hv\|_\infty^2 e, \qquad (19) \end{aligned}$$

where $G := \text{diag}(g)$, and g^2, $(Hv)^2$, and $(\eta g - \zeta Hv)^2$ denote the vectors whose components are the squares of those of g, Hv, and $\eta g - \zeta Hv$ respectively.

The proof of theorem 4 proceeds by using (18) to evaluate $\|Pq\|^2$, and then analyzing all the terms in the resulting expression. See [16] for details. The proof of theorem 5 follows from (19) (which gives $\|Pq\|_\infty^- \leq \|Hv\|_\infty^2$) and the following result:

Lemma 7 *Let $F = [g, H]$ be an orthogonal matrix. If v is uniformly distributed on the unit sphere in R^{n-1},*

$$\Pr\left(\|Hv\|_\infty \leq \sqrt{3\frac{\log n}{n}}\right) \to 1 \quad as \quad n \to \infty.$$

Proof. Since v is uniformly distributed on the unit sphere in R^{n-1}, it can be generated as follows: $v = \lambda/\|\lambda\|$, where $\lambda \sim N(0, I)$ (the standard normal distribution in R^{n-1}). Hence we wish to obtain an upper bound on $\|H\lambda\|_\infty$ and a lower bound on $\|\lambda\|$, both of which hold with high probability. Now $\|\lambda\|^2$ is a χ^2 random variable with $n-1$ degrees freedom, so

$$\begin{aligned} \text{E}(\|\lambda\|^2) &= n-1, \\ \text{Var}(\|\lambda\|^2) &= 2(n-1). \end{aligned}$$

From Chebychev's inequality, we have

$$\Pr(\|\lambda\| \geq (1-\epsilon)\sqrt{n-1}) \to 1 \quad \text{as} \quad n \to \infty \qquad (20)$$

for any $\epsilon > 0$.

Let λ_0 be a standard normal variable, and let $\lambda' = (\lambda_0, \lambda)$, also $N(0, I)$ but in R^n. Then $\|\lambda'\|_\infty = \max\{\nu_j : j = 0, 1, 2, \cdots, n-1\}$ where $\nu_j = |\lambda_j|$ has the positive normal distribution. Then $1 - N_+(x) = 2(1 - N(x))$ where N_+ is the distribution function of ν, and N is the normal distribution function. It now follows from results in extreme value theory (Resnick [23], pp. 42 and 71) that

$$\Pr\left(\|\lambda'\|_\infty \leq \sqrt{2\log(2n)}\right) \to 1 \text{ as } n \to \infty.$$

Since $F\lambda'$ is also $N(0,I)$,

$$\Pr\left(\|F\lambda'\|_\infty \leq \sqrt{2\log(2n)}\right) \to 1 \text{ as } n \to \infty.$$

Now we have

$$\|H\lambda\|_\infty \leq \|F\lambda'\|_\infty + \|\lambda_0 g\|_\infty.$$

Since $\|g\| = 1$,

$$\Pr\left(\|\lambda_0 g\|_\infty \leq \epsilon\sqrt{\log n}\right) \to 1 \text{ as } n \to \infty$$

for any $\epsilon > 0$. From the above relations and (20), we get the result of the lemma. □

We conclude this section by showing how (18) and lemma 7 imply a slightly weaker form of theorem 4. Indeed, (18) yields

$$\begin{aligned} \|Pq\| &\leq \eta^2\|g^2\| + 2|\zeta\eta|\|g\|_\infty\|Hv\| + \eta^2\|(Hv)^2\| \\ &\leq \|g\|_\infty\|g\| + 2\|g\|_\infty + \|Hv\|_\infty\|Hv\| \\ &= 3\rho + \|Hv\|_\infty. \end{aligned}$$

By lemma 7, this is at most $3\rho + \sqrt{3\log(n)/n}$ with probability approaching 1 as $n \to \infty$. This bound would lead one to hope that $\|Pq\|$ would be at most $(n\log(n))^{0.5}\mu$ at a typical predictor step. Algorithm 1 would require at most $O((n\log(n))^{0.25}t)$ iterations of this type.

7 Concluding remarks

This paper has described a number of adaptive-step primal-dual algorithms for linear programming. They all have complexities of $O(n^{0.5}t)$ or $O(nt)$ iterations to attain precision t. We have also provided heuristic arguments for our expectation that, in practice, these algorithms will only require $O(n^{0.25}t)$ or $O(\log(n)t)$ iterations. In this section we will briefly comment on this approach.

There are two main viewpoints in the probabilistic analysis of algorithms. First one can develop randomized algorithms, and show that, on a worst-case instance of a problem, the average running time of the algorithm has a certain bound, or the running time satisfies a certain bound with high probability, or the running time always satisfies a certain bound and the algorithm gives a correct answer with high probability. As examples in linear programming, we cite the recent paper of Seidel [24], who gives a simple randomized algorithm whose expected running time for (P) is $O(m!n)$, and the references therein. Second one can consider the expected running time of a deterministic algorithm when applied to problem instances generated according to some probability distribution (or class of such distributions). See Borgwardt [3] and the references cited there for examples of such results in linear programming. This viewpoint is rather less compelling, since one can always argue that the distribution chosen for problem instances is inappropriate.

Our approach is distinct from the two just described. As we have noted, the assumptions we have made at each iteration are inconsistent with one another. Nevertheless, we feel that such an approach adds insight in the case where a more rigorous analysis seems intractable. Similar analyses have been made

for the simplex method (Dantzig [4]), for (a variant of) Karmarkar's method (Nemirovsky [21]), and for a primal-dual method (Gonzaga and Todd [7]).

Our results can be contrasted with those of Sonnevend, Stoer and Zhao [26, 27]. As we have noted, our predictor-corrector method is very similar to their algorithm, which also has a worst-case complexity of $O(n^{0.5}t)$ iterations to attain precision t. While we make a nonrigorous probabilistic analysis to conclude that $O(n^{0.25}t)$ iterations may typically suffice, they perform a sophisticated analysis relating the number of iterations to a curvature integral, and then conclude that for certain subclasses of problems $O(n^{0.25}t)$ iterations suffice.

Finally, let us describe a possible program to make our analysis rigorous. Suppose we assume that our original problem (P) is generated probabilistically as follows: the entries of A are independent standard normal random variables, $b = Ae$ and $c = A^T y + e$ for some y. Then $(x, s) = (e, e)$ is an initial point on the central path $\mathcal{C}$. Moreover, for all of our algorithms, r is a multiple of e and U is a random subspace with the orthogonal transformation-invariant distribution. Hence our analysis holds at the initial iteration. We now apply an algorithm similar to that of Sonnevend et al., so that each iterate lies in $\mathcal{C}$ and hence $r = e$ at each iteration. However, the null space U of $AX^{0.5}S^{-0.5}$ will have a different induced distribution at later iterations. We could hope that, before (x, s) gets too close to an optimal pair, this induced distribution is somewhat close to that we have assumed in section 6, so that its Radon-Nikodym derivative with respect to our distribution is suitably bounded. In this case, the probability that $\|Pq\|$ exceeds $n^{0.5}\mu$, which is small under the distribution we have assumed, will also be small under the distribution induced by the initial probabilistic generation of (P). Hence, for most iterations, the improvement in the duality gap would be as in corollary 1. A great many difficulties need to be resolved before such an approach could succeed. We would probably need bounds on how fast the probabilities in theorems 4 and 5 approach 1, and clearly as (x, s) approaches the optimum the induced distribution differs drastically from what we have assumed. But note that Sonnevend et al. have shown that asymptotically the required faster rate is achieved.

In the meantime, we hope that our nonrigorous analysis has lent some insight into the practical behavior of primal-dual algorithms. Our algorithms using $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$ for β close to 1 (and $\rho = n + O(n^2)$ for the potential reduction algorithm) are quite close to implemented primal-dual methods, and the result of our nonrigorous analysis, that $O(\log(n)t)$ iterations typically suffice, is borne out by the large-scale tests performed by Lustig et al. [11] with such an algorithm.

References

[1] K. M. Anstreicher and R. A. Bosch, "Long steps in an $O(n^3L)$ algorithm for linear programming," manuscript, Yale School of Management (New Haven, CT, 1989), to appear in *Mathematical Programming, Series A.*

[2] D. A. Bayer and J. C. Lagarias, "The nonlinear geometry of linear programming: I. Affine and projective scaling trajectories, II. Legendre transform coordinates and central trajectories," *Transactions of the*

American Mathematical Society 314 (1989) 499-581.

[3] K. H. Borgwardt, *The Simplex Method – A Probabilistic Approach* (Springer-Verlag, Berlin-Heidelberg-New York, 1987).

[4] G. B. Dantzig, "Expected number of steps of the simplex method for a linear program with a convexity constraint," Technical Report SOL 80-3, Department of Operations Research, Stanford University (Stanford, CA 1980).

[5] R. M. Freund, "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function," *Mathematical Programming* 51 (1991) 203-222.

[6] C. C. Gonzaga, "An algorithm for solving linear programming programs in $O(n^3L)$ operations," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1988) 1-28.

[7] C. C. Gonzaga and M. J. Todd, "An $O(\sqrt{n}L)$-iteration large-step primal-dual affine algorithm for linear programming," Technical Report No. 862, School of Operations Research and Industrial Engineering, Cornell University (Ithaca, New York, 1989), to appear in *SIAM J. on Optimization*.

[8] M. Kojima, S. Mizuno, and A. Yoshise, "An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems," *Mathematical Programming* 50 (1991) 331-342.

[9] M. Kojima, S. Mizuno, and A. Yoshise, "A primal-dual interior point algorithm for linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1989) 29-47.

[10] M. Kojima, S. Mizuno and A. Yoshise, "A polynomial-time algorithm for a class of linear complementarity problems," *Mathematical Programming* 44 (1989) 1-26.

[11] I. J. Lustig, R. E. Marsten and D. F. Shanno, "The primal-dual interior point method on the Cray supercomputer," in: T. F. Coleman and Y. Li, eds., *Large-scale Numerical Optimization* (SIAM, Philadelphia, 1990) 70-80.

[12] K. A. McShane, C. L. Monma, and D. Shanno, "An implementation of a primal-dual interior point method for linear programming," *ORSA Journal on Computing* 1 (1989) 70-83.

[13] N. Megiddo, "Pathways to the optimal set in linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1988) 131-158.

[14] S. Mizuno, A. Yoshise, and T. Kikuchi, "Practical polynomial time algorithms for linear complementarity problems," *Journal of the Operations Research Society of Japan* 32 (1989) 75-92.

[15] S. Mizuno, "A new polynomial time method for a linear complementarity problem," Research Report 16, Department of Industrial Engineering and Management, Tokyo Institute of Technology (Tokyo, Japan, 1989), to appear in *Mathematical Programming, Series A.*

[16] S. Mizuno, M. J. Todd, and Y. Ye, "Anticipated behavior of path-following algorithms for linear programming," Technical Report No. 878, School of Operations Research and Industrial Engineering, Cornell University (Ithaca, New York, 1989).

[17] S. Mizuno, M. J. Todd, and Y. Ye, "Anticipated behavior of long-step algorithms for linear programming," Research Report 24, Department of Industrial Engineering and Management, Tokyo Institute of Technology (Tokyo, Japan, 1990).

[18] R. C. Monteiro, I. Adler, and M. G. C. Resende, "A polynomial time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension," *Mathematics of Operations Research* 15 (1990) 191-214.

[19] R. C. Monteiro and I. Adler, "Interior path following primal-dual algorithms. Part I: Linear programming," *Mathematical Programming* 44 (1989) 27-42.

[20] R. C. Monteiro and I. Adler, "Interior path following primal-dual algorithms. Part II: Convex quadratic programming," *Mathematical Programming* 44 (1989) 43-66.

[21] A. S. Nemirovsky, "An algorithm of the Karmarkar type," *Tekhnicheskaya Kibernetika* 1 (1987) 105-118.

[22] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," *Mathematical Programming* 40 (1988) 59-94.

[23] S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes* (Springer-Verlag, New York, 1987).

[24] R. Seidel, "Small-Dimensional linear programming and convex hulls made easy," Discrete & Computational Geometry 6(5) 423-434, 1991.

[25] G. Sonnevend, "An analytical center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," in: *Lecture Notes in Control and Information Sciences* 84 (Springer, New York, 1985) 866-876.

[26] G. Sonnevend, J. Stoer, and G. Zhao, "On the complexity of following the central path of linear programs by linear extrapolation," *Methods of Operations Research* 63 (1989) 19-31.

[27] G. Sonnevend, J. Stoer, and G. Zhao, "On the complexity of following the central path of linear programs by linear extrapolation II," manuscript, University of Wűrzburg (Wűrzburg, Germany, 1990), to appear in *Mathematical Programming, Series B.*

[28] M. J. Todd and Y. Ye, "A centered projective algorithm for linear programming," *Mathematics of Operations Research* 15 (1990) 508-529.

[29] P. M. Vaidya, "An algorithm for linear programming which requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations," *Mathematical Programming* 47 (1990) 175-202.

[30] S. S. Wilks, *Mathematical Statistics* (John Wiley, New York, 1962).

[31] Y. Ye, "Further development on the interior algorithm for convex quadratic programming," manuscript, Stanford University (Stanford, CA), and Integrated Systems Inc. (Santa Clara, CA, revised November 1987).

[32] Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programming," in: J. C. Lagarias and M. J. Todd, eds., *Mathematical Developments Arising from Linear Programming* (American Mathematical Society, Providence, RI, 1990) 91-107.

[33] Y. Zhang, R. A. Tapia, and J. E. Dennis, "On the superlinear and quadratic convergence of primal-dual interior-point linear programming algorithms," manuscript, Department of Mathematical Sciences, Rice University (Houston, Texas, 1990).

[34] G. Zhao and J. Stoer, "Estimating the complexity of path-following methods for solving linear programs by curvature integrals," Report No. 225, Institut fűr Angewandte Mathematik und Statistik, Universität Wűrzburg (Wűrzburg, Germany, 1990).