

# MONOTONICITY OF PRIMAL AND DUAL OBJECTIVE VALUES IN PRIMAL-DUAL INTERIOR-POINT ALGORITHMS \*

SHINJI MIZUNO<sup>†</sup>, MICHAEL J. TODD<sup>‡</sup>, AND LEVENT TUNÇEL<sup>§</sup>

**Abstract.** We study monotonicity of primal and dual objective values in the framework of primal-dual interior-point methods. The primal-dual affine-scaling algorithm is monotone in both objectives. We derive a condition under which a primal-dual interior-point algorithm with a centering component is monotone. Then we propose primal-dual algorithms that are monotone in both primal and dual objective values and achieve polynomial time bounds. We also provide some arguments showing that several existing primal-dual algorithms use parameters close to one that almost always improves both objectives.

**Key words.** Linear programming, primal-dual, interior-point methods, monotonicity.

**AMS subject classifications.** 90C05, 65K05

**1. Introduction.** Recent interest in interior-point methods for linear programming has focused on primal-dual interior-point algorithms. Numerical experiments seem to indicate that primal-dual algorithms are superior to primal-only or dual-only methods. Convergence of most of the primal-dual interior-point algorithms is based on either the decrease in the duality gap or the decrease in a potential function. Neither of these two measures tells us anything about the change in the true primal and true dual objective values. In primal-only (or dual-only) settings some work has been done to show that polynomiality and strict improvement in the objective function at each iteration can be achieved (see Anstreicher [An86, An91], Mizuno and Nagasawa [MN92a]). In this paper we address the following question: Is it possible to achieve the best polynomial bound with monotonicity in both primal and dual objective functions for primal-dual algorithms?

When one implements a primal-dual interior-point algorithm, the step sizes are often determined as a fixed ratio of the distance from the current iterate to the boundary in the primal and dual feasible regions respectively, see [MMS89, LMS91] for example. In that case, the next iterate may not be well-defined because one of the primal and dual directions may be an unbounded direction of the corresponding feasible region. In addition, when separate step sizes are taken in the primal and dual spaces, the duality gap may increase at the next iterate. Since these phenomena do not occur if the algorithm is monotone in both objective values, the analysis of monotone algorithms is important from this practical point of view.

---

\* Copyright (C) by the Society for Industrial and Applied Mathematics, in SIAM Journal on Optimization, 4 (1994), pp. 613–625.

<sup>†</sup> The Institute of Statistical Mathematics, 4–6–7 Minami-Azabu, Minato-ku, Tokyo 106, JAPAN. Research performed while visiting Cornell University (April 92 — January 93) as an overseas research scholar of the Ministry of Science, Education and Culture of Japan.

<sup>‡</sup> School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853–3801, U.S.A. Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS–8920550.

<sup>§</sup> Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, CANADA. Research performed while this author was a Ph.D. student at Cornell University and was supported in part by the United States Army Research Office through the Army Center of Excellence for Symbolic Methods in Algorithmic Mathematics (ACSyAM), Mathematical Sciences Institute of Cornell University, Contract DAAL03-91-C-0027 and also by NSF, AFOSR and ONR through NSF Grant DMS–8920550.

First, we show that the primal-dual affine-scaling algorithm of Monteiro, Adler and Resende [MAR90] (see also [KMNY91, MN92b, Tu92]) is (usually strictly) monotone in both primal and dual objective values. This was proved by Monteiro et al.; see p. 213 in [MAR90]. Then we derive a condition on the centering parameter under which a primal-dual interior-point algorithm with a centering component is monotone in both objective values. It seems that most primal-dual interior-point algorithms (except for the affine-scaling algorithm) are not guaranteed to satisfy this condition. We show how to control the centering parameter in a way that allows monotonicity in both objectives and does not hurt the polynomial time bound proven for the non-monotone version of a primal-dual algorithm. In particular, we study algorithms using one-sided infinity neighborhoods and two-norm neighborhoods of the central path and show that monotonicity can be achieved while keeping the iteration bounds  $O(nt)$  and  $O(\sqrt{nt})$  respectively (where  $t$  denotes the desired improvement in precision). The algorithm proposed is not only monotone, but it may also enable us to take a longer step in the sense that the centering parameter is less than in the non-monotone version of the algorithm.

From the theoretical point of view, the complexity of primal-dual interior-point algorithms is based on a bound on the second-order term in Newton's method (see [MTY90] for example). In this paper, we obtain a new bound on the second-order term to show the polynomiality of the algorithm (see Lemma 9).

Finally, we study the anticipated critical value of the centering parameter and show that several existing primal-dual algorithms use parameters close to one that improves both objectives at almost every iteration.

**2. Preliminaries.** We consider linear programming problems in the following primal ( $P$ ) and dual ( $D$ ) forms:

$$\begin{aligned} (P) \quad & \text{minimize} && c^T x \\ & && Ax = b, \\ & && x \geq 0, \\ \\ (D) \quad & \text{maximize} && b^T y \\ & && A^T y + s = c, \\ & && s \geq 0, \end{aligned}$$

where  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , and  $c \in \mathbb{R}^n$ . We assume  $A$  has full row rank (without loss of generality) and that there exist interior solutions for both problems, i.e.,

$$\mathcal{F}_0 := \{(x, s) > 0 : x \in F(P), s \in F(D)\} \neq \emptyset,$$

where  $F(P)$  and  $F(D)$  denote the sets of feasible solutions for the primal and dual problems respectively. Most of the time we deal only with  $s$  as a dual feasible solution. So, whenever we say  $s \in F(D)$ , we mean that  $s \geq 0$  and there exists a  $y \in \mathbb{R}^m$  such that  $A^T y + s = c$ . Given a vector denoted by a lower-case roman letter (e.g.  $x$ ), the corresponding upper-case letter (e.g.  $X$ ) will denote the diagonal matrix whose entries are the components of that vector (i.e.  $X = \text{diag}(x)$ ), and  $e$  will denote the vector of ones. We denote the components of a vector using subscripts and the iterate

numbers using superscripts. Whenever we ignore superscripts it will be clear from the context what the iterate number should be.

Next we describe the central path and its neighborhoods. The central path is given by the set of solutions to the following system of equations and inequalities (for  $\mu > 0$ ):

$$\begin{aligned} (1) \quad & Ax = b, \quad x \geq 0, \\ (2) \quad & A^T y + s = c, \quad s \geq 0, \\ (3) \quad & Xs = \mu e. \end{aligned}$$

Note that when  $\mu = 0$ , (1) – (3) give necessary and sufficient conditions for optimality. Analyses of the central path have been performed by several authors (see, for instance, Sonnevend [So85], Megiddo [Meg88], and Bayer and Lagarias [BL89]). Our objective is to follow this path approximately to an optimal solution. We define some neighborhoods of the central path using the notation of [MTY90]. Let  $\beta \in (0, 1)$  be a constant; then a 2-norm neighborhood of the central path can be defined as

$$\mathcal{N}_2(\beta) := \{(x, s) \in \mathcal{F}_0 : \|Xs - \mu e\|_2 \leq \beta\mu \text{ for } \mu = \frac{x^T s}{n}\}.$$

Henceforth,  $\mu$  always denotes  $\frac{x^T s}{n}$ . Kojima et al. [KMY89], Monteiro and Adler [MA89], and Mizuno et al. [MTY90] designed algorithms that use a 2-norm neighborhood. Using the  $\infty$ -norm or just one side of the  $\infty$ -norm, wider neighborhoods have been defined and used by Kojima et al. [KMY88], Mizuno et al. [MTY90], and Zhang and Tapia [ZT90]. Using the  $\infty$ -norm we have

$$\mathcal{N}_\infty(\beta) := \{(x, s) \in \mathcal{F}_0 : \|Xs - \mu e\|_\infty \leq \beta\mu\}$$

and using only one side of the  $\infty$ -norm,

$$\mathcal{N}_\infty^-(\beta) := \{(x, s) \in \mathcal{F}_0 : \|Xs - \mu e\|_\infty^- \leq \beta\mu\}.$$

Here, for  $u \in \mathbb{R}^n$ ,  $\|u\|_\infty^- := -\min\{0, \min\{u_j\}\}$ . Clearly, for a given  $\beta \in (0, 1)$ ,  $\mathcal{N}_2(\beta)$  is the smallest and  $\mathcal{N}_\infty^-(\beta)$  the largest of the three neighborhoods defined here.

Suppose we have an interior-point solution  $(x, s) \in \mathcal{F}_0$ . Then a search direction  $(dx, ds)$  can be generated by solving the following set of equalities (see Kojima et al. [KMY88]):

$$\begin{aligned} (4) \quad & Adx = 0, \\ (5) \quad & A^T dy + ds = 0, \\ (6) \quad & Sdx + Xds = \gamma\mu e - Xs, \end{aligned}$$

where  $\gamma \in [0, 1]$  is a constant. The solution of (4) – (6) is

$$\begin{aligned} dx(\gamma) &:= -X^{1/2}S^{-1/2}P_{\bar{A}}(X^{1/2}S^{1/2} - \gamma\mu X^{-1/2}S^{-1/2})e, \\ ds(\gamma) &:= -X^{-1/2}S^{1/2}(I - P_{\bar{A}})(X^{1/2}S^{1/2} - \gamma\mu X^{-1/2}S^{-1/2})e, \end{aligned}$$

where  $\bar{A} := AX^{1/2}S^{-1/2}$ , and  $P_{\bar{A}} := I - \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A}$ , the projection matrix into the null space of  $\bar{A}$ . The next iterates in terms of  $\alpha$  can be written as

$$\begin{aligned} x(\alpha) &= x + \alpha dx(\gamma), \\ s(\alpha) &= s + \alpha ds(\gamma). \end{aligned}$$

A primal-dual interior-point algorithm is one that, given an iterate  $(x, s) \in \mathcal{F}_0$ , defines the next iterate as  $(x(\alpha), s(\alpha))$  derived as above for some  $\gamma \in [0, 1]$  and some  $\alpha > 0$  such that  $(x(\alpha), s(\alpha)) \in \mathcal{F}_0$  (the centering parameter  $\gamma$  and the step size  $\alpha$  may vary from one iteration to the next). We define

$$\begin{aligned} v &:= \frac{1}{\sqrt{\mu}}X^{1/2}S^{1/2}e, \\ w &:= \sqrt{\mu}X^{-1/2}S^{-1/2}e. \end{aligned}$$

Note that  $\|v\|_2^2 = n$  and  $v^T w = n$ . Also observe that  $(x, s)$  is on the central path if and only if  $v = w = e$ . We use  $v_p$  to denote the projection of  $v$  onto the null space of  $\bar{A}$ , etc., and  $v_q$  to denote the projection of  $v$  onto the image space of  $\bar{A}^T$ , etc. Then we have that

$$(7) \quad \begin{aligned} Xs - \mu e &= \mu V(v - w), \\ dx(\gamma) &= -\sqrt{\mu}X^{1/2}S^{-1/2}(v_p - \gamma w_p), \\ ds(\gamma) &= -\sqrt{\mu}X^{-1/2}S^{1/2}(v_q - \gamma w_q), \end{aligned}$$

where  $V = \text{diag}(v)$ . From these equalities, we see that

$$\begin{aligned} (8) \quad \mathcal{N}_2(\beta) &= \{(x, s) \in \mathcal{F}_0 : \|V(v - w)\|_2 \leq \beta\}, \\ (9) \quad \mathcal{N}_\infty^-(\beta) &= \{(x, s) \in \mathcal{F}_0 : \|V(v - w)\|_\infty^- \leq \beta\}, \\ (10) \quad dX(\gamma)ds(\gamma) &= \mu(V_p - \gamma W_p)(v_q - \gamma w_q), \end{aligned}$$

where  $dX(\gamma) = \text{diag}(dx(\gamma))$ ,  $V_p = \text{diag}(v_p)$ , and  $W_p = \text{diag}(w_p)$ .

**3. Monotonicity.** First we show that the primal-dual affine-scaling algorithms (see Monteiro, Adler and Resende [MAR90], Kojima, Megiddo, Noma, and Yoshise [KMNY91], Mizuno and Nagasawa [MN92b], and Tunçel [Tu92]) are (usually strictly) monotone in both objectives. As mentioned in the introduction, this was essentially proved in [MAR90].

Note that for any  $(x, s) \in \mathcal{F}_0$  we have  $X^{1/2}S^{-1/2}A^T y + X^{1/2}S^{1/2}e = X^{1/2}S^{-1/2}c$  or

$$\bar{A}^T y + \sqrt{\mu}v = \bar{c},$$

where  $\bar{c} := X^{1/2}S^{-1/2}c$ . Projecting the vectors on both sides of the equality onto the null space of  $\bar{A}$ , we see that

$$(11) \quad v_p = \frac{1}{\sqrt{\mu}} \bar{c}_p.$$

An affine-scaling algorithm results from setting  $\gamma = 0$  at each iteration. Then we get

$$(12) \quad \begin{aligned} c^T x(\alpha) - c^T x &= \alpha c^T dx(0) \\ &= -\alpha \sqrt{\mu} (X^{1/2}S^{-1/2}c)^T v_p \\ &= -\alpha \sqrt{\mu} \bar{c}^T v_p \\ &= -\alpha \mu \|v_p\|_2^2. \end{aligned}$$

For each  $s(\alpha)$ , there is a  $y(\alpha)$  such that  $(y(\alpha), s(\alpha))$  is feasible for (D). In the same way, we get

$$(13) \quad \begin{aligned} b^T y(\alpha) - b^T y &= (Ax)^T y(\alpha) - (Ax)^T y \\ &= -x^T s(\alpha) + x^T s \\ &= -\alpha x^T ds(0) \\ &= \alpha \mu \|v_q\|_2^2. \end{aligned}$$

We immediately have the following theorem.

**THEOREM 1.** *A primal-dual affine-scaling algorithm is strictly monotone in both problems unless all the primal or all the dual solutions are optimal.*

*Proof.* Since a primal-dual affine-scaling algorithm is monotone in both objectives by (12) and (13), we only prove strict monotonicity.

Suppose  $\|v_p\|_2 = 0$ . Then  $\bar{c}_p = 0$  by (11). So  $\hat{s} := 0$  is feasible in (D); therefore, for any  $x \in F(P)$ ,  $x$  is optimal in (P) and  $\hat{s}$  in (D) by complementary slackness.

Now suppose  $\|v_q\|_2 = 0$ . Then  $b = Ax = \sqrt{\mu}\bar{A}v = 0$ , which implies  $\hat{x} := 0$  is feasible in (P). Hence, for any  $s$  that is feasible in (D),  $s$  is optimal in (D) and  $\hat{x}$  is optimal in (P) by complementary slackness.  $\square$

From now on, without loss of generality we assume that

$$0 < \|v_p\|_2^2 < n.$$

In the same way as we obtained (12), we have

$$\begin{aligned} c^T x(\alpha) - c^T x &= -\alpha\sqrt{\mu}(X^{1/2}S^{-1/2}c)^T(v_p - \gamma w_p) \\ &= -\alpha\mu v_p^T(v_p - \gamma w_p). \end{aligned}$$

So we have the following identities for improvements in the objective function values:

$$(14) \quad c^T x(\alpha) - c^T x = -\alpha\mu(\|v_p\|_2^2 - \gamma v_p^T w_p),$$

and, similarly,

$$(15) \quad b^T y(\alpha) - b^T y = \alpha\mu(\|v_q\|_2^2 - \gamma v_q^T w_q).$$

We define  $\gamma_P$  and  $\gamma_D$  as *constant-cost centering values* for the primal and dual problems respectively so that  $c^T x(\alpha) = c^T x$  and  $b^T y(\alpha) = b^T y$ :

$$(16) \quad \begin{aligned} \gamma_P &:= \frac{\|v_p\|_2^2}{v_p^T w_p}, \\ \gamma_D &:= \frac{\|v_q\|_2^2}{v_q^T w_q}, \end{aligned}$$

where we define  $\gamma_P := +\infty$  or  $\gamma_D := +\infty$  when the corresponding denominator is 0. Note that if  $(x, s)$  is on the central path ( $v = w$ ) then  $\gamma_P = \gamma_D = 1$ . The next lemma describes a relation between  $\gamma_P$  and  $\gamma_D$ .

LEMMA 2. *Either  $\gamma_P = \gamma_D = 1$*

*or*

*$\gamma_P \in (0, 1)$  and  $\gamma_D \in (-\infty, 0) \cup (1, \infty)$*

*or*

*$\gamma_D \in (0, 1)$  and  $\gamma_P \in (-\infty, 0) \cup (1, \infty)$ .*

*Proof.* We see that  $n = v^T v = \|v_p\|_2^2 + \|v_q\|_2^2$  and  $n = v^T w = v_p^T w_p + v_q^T w_q$ . If  $v_p^T w_p = \|v_p\|_2^2$  then  $v_q^T w_q = \|v_q\|_2^2$ , so we have  $\gamma_P = \gamma_D = 1$ . If  $v_p^T w_p > \|v_p\|_2^2$  then  $\gamma_P \in (0, 1)$  and  $\gamma_D \in (-\infty, 0) \cup (1, \infty)$ . Otherwise  $\gamma_D \in (0, 1)$  and  $\gamma_P \in (-\infty, 0) \cup (1, \infty)$ .  $\square$

Define

$$\bar{\gamma} := \gamma_P \text{ if } \gamma_P \in (0, 1], \text{ and } \bar{\gamma} := \gamma_D \text{ otherwise.}$$

LEMMA 3. *If  $0 \leq \gamma \leq \bar{\gamma}$  ( $0 \leq \gamma < \bar{\gamma}$ ), then the primal objective function is non-increasing (decreasing) along the direction  $dx(\gamma)$  and the dual objective function is non-decreasing (increasing) along the direction  $ds(\gamma)$ .*

*Proof.* Follows from Lemma 2 and equations (14) and (15).  $\square$

Now we have a sufficient condition under which a primal-dual interior-point algorithm is monotone in both objective values.

**THEOREM 4.** *An algorithm whose search directions are  $dx(\gamma)$  and  $ds(\gamma)$  for  $0 \leq \gamma \leq \bar{\gamma}$  ( $0 \leq \gamma < \bar{\gamma}$ ) at each iterate is (strictly) monotone in both primal and dual objective values.*

*Proof.* Directly follows from Lemma 3.  $\square$

If the exceptional case ( $\|v_p\|_2^2 = 0$  or  $\|v_p\|_2^2 = n$ ) is excluded, Theorem 1 is a special case of Theorem 4. Although the condition in Theorem 4 is simple, none of the primal-dual interior-point algorithms, except for the affine-scaling algorithm, seems to use values for the centering parameter that are guaranteed to satisfy it. Here we describe a strictly monotone algorithm, which is a simple variant of a primal-dual path-following algorithm [KMY88, KMY89, MA89, MTY90]. Let  $\mathcal{N}$  be a neighborhood of the central path and  $\gamma' \in (0, 1)$  be a constant. Suppose  $(x^0, s^0) \in \mathcal{N}$  with  $(x^0)^T s^0 \leq 2^t$  is given.

### Algorithm:

$k := 0$

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

$(x, s) := (x^k, s^k),$

compute  $v_p, v_q, w_p, w_q, \gamma_P$  and  $\gamma_D,$

$\bar{\gamma} := \gamma_P$  if  $\gamma_P \in (0, 1]$  and  $\bar{\gamma} := \gamma_D$  otherwise,

$\gamma := \gamma' \bar{\gamma},$

choose the maximum step size  $\alpha \in (0, 1)$  such that  $(x + \alpha dx(\gamma), s + \alpha ds(\gamma)) \in \mathcal{N},$

$(x^{k+1}, s^{k+1}) := (x + \alpha dx(\gamma), s + \alpha ds(\gamma)),$

$k := k + 1$

end

If we use  $\gamma := \gamma'$  in the algorithm above, it becomes a well known path-following algorithm. In that case, the algorithm terminates in  $O(nt)$  iterations when  $\mathcal{N} = \mathcal{N}_{\infty}^-(\beta)$  and  $\gamma' \in (0, 1)$  is a constant and it terminates in  $O(\sqrt{nt})$  iterations when  $\mathcal{N} = \mathcal{N}_2(\beta)$  and  $\gamma' = 1 - \delta/\sqrt{n}$  for a positive constant  $\delta$  (for suitable values of

the constants see [KMY88, KMY89, MA89, MTY90]). Taking  $\mathcal{N} := \mathcal{N}_\infty^-(\beta)$  and  $\gamma' \in (0, 1)$  in the algorithm above will define Algorithm I. Algorithm II will be defined by setting  $\mathcal{N} := \mathcal{N}_2(\beta)$  and  $\gamma' := 1 - \delta/\sqrt{n}$ . In the next section, we show that Algorithm I terminates in  $O(nt)$  iterations and Algorithm II terminates in  $O(\sqrt{nt})$  iterations.

Note that if  $\bar{\gamma}$  is small, the centering parameter  $\gamma$  in the algorithm is small even if  $\gamma'$  is close to 1. This observation distinguishes Algorithms I and II from the path-following algorithms in [KMY88, KMY89, MA89, MTY90].

**4. Analysis.** The following results (Lemma 5 and Lemma 6) are standard in the primal-dual framework (see for instance Mizuno, Todd and Ye [MTY90]).

LEMMA 5. (a)  $x_j(\alpha)s_j(\alpha) = \mu[(1-\alpha)v_j^2 + \alpha\gamma] + \alpha^2 dx_j(\gamma)ds_j(\gamma)$ .

(b)  $\mu(\alpha) := x(\alpha)^T s(\alpha)/n = (1-\alpha + \alpha\gamma)\mu$ .

*Proof.* Both (a) and (b) directly follow from the definitions and  $dx(\gamma)^T ds(\gamma) = 0$ .

□

Part (b) of this lemma shows how the duality gap varies with the step size  $\alpha$ . We would like  $\alpha(1-\gamma)$  to be large. The next result shows how large  $\alpha$  can be.

LEMMA 6. (a)  $(x, s) \in \mathcal{N}_\infty^-(\beta)$  and  $\alpha \leq \min \left\{ \frac{\beta\gamma\mu}{\|dX(\gamma)ds(\gamma)\|_\infty}, 1 \right\}$  implies  $(x(\alpha), s(\alpha)) \in \mathcal{N}_\infty^-(\beta)$ .

(b)  $(x, s) \in \mathcal{N}_2(\beta)$  and  $\alpha \leq \min \left\{ \frac{\beta\gamma\mu}{\|dX(\gamma)ds(\gamma)\|_2}, 1 \right\}$  implies  $(x(\alpha), s(\alpha)) \in \mathcal{N}_2(\beta)$ .

*Proof.* (a) Suppose  $(x, s) \in \mathcal{N}_\infty^-(\beta)$  and  $\alpha \leq \min \left\{ \frac{\beta\gamma\mu}{\|dX(\gamma)ds(\gamma)\|_\infty}, 1 \right\}$ . By (9),  $\|V(v-w)\|_\infty \leq \beta$ . By using  $Vw = e$ , we see that

$$(17) \quad v_j^2 \geq 1 - \beta \quad \text{for each } j.$$

So, using Lemma 5 and the bound on  $\alpha$ , we get (for all  $j$ )

$$\begin{aligned} x_j(\alpha)s_j(\alpha) &\geq \mu[(1-\alpha)(1-\beta) + \alpha\gamma] - \alpha\beta\gamma\mu \\ &= \mu(1-\alpha + \alpha\gamma)(1-\beta) \\ &= \mu(\alpha)(1-\beta). \end{aligned}$$

Since  $\alpha \in [0, 1]$  and  $\gamma > 0$ , the right hand side is always positive. So  $(x(\alpha), s(\alpha)) \in \mathcal{F}_0$ . Hence  $(x(\alpha), s(\alpha)) \in \mathcal{N}_\infty^-(\beta)$ .

(b) Suppose  $(x, s) \in \mathcal{N}_2(\beta)$  and  $\alpha \leq \min \left\{ \frac{\beta\gamma\mu}{\|dX(\gamma)ds(\gamma)\|_2}, 1 \right\}$ . Then feasibility of  $(x(\alpha), s(\alpha))$  can be verified as in (a). Using (6), the bound on  $\alpha$ , and Lemma 5, we obtain



$$\begin{aligned}
\|X(\alpha)s(\alpha) - \mu(\alpha)e\|_2 &= \|(1 - \alpha)Xs + \alpha\gamma\mu e + \alpha^2 dX(\gamma)ds(\gamma) - (1 - \alpha + \alpha\gamma)\mu e\|_2 \\
&= \|(1 - \alpha)(Xs - \mu e) + \alpha^2 dX(\gamma)ds(\gamma)\|_2 \\
&\leq (1 - \alpha)\|Xs - \mu e\|_2 + \alpha\beta\gamma\mu \\
&\leq (1 - \alpha)\beta\mu + \alpha\beta\gamma\mu \\
&= \beta\mu(\alpha).
\end{aligned}$$

□

One way of showing a bound on the number of iterations while keeping the algorithm monotone is to show that there exists an  $\epsilon > 0$  such that  $\bar{\gamma} \geq \epsilon$  for all iterates. It seems hard to find such an  $\epsilon$  that is close to 1 (or even bounded away from zero). Instead, we will show that the second-order terms  $\|dXds\|_\infty^-$  and  $\|dXds\|_2$  can be bounded by a multiple of  $\bar{\gamma}\mu$ . This is a new way of estimating the second-order terms. Since the bound we will show is a multiple of  $\bar{\gamma}$ , it will imply that whenever  $\bar{\gamma}$  is small, the second-order terms will be small, hence allowing us to take a larger step and still stay in the desired neighborhood. Since  $dX(\gamma)ds(\gamma) = \mu(V_p - \gamma W_p)(v_q - \gamma w_q)$  by (10), we get bounds on  $\|v_p - \gamma w_p\|_2$  and  $\|v_q - \gamma w_q\|_2$  in the following two lemmas.

**LEMMA 7.** *If  $\gamma_P < +\infty$ , then  $\|v_p - \gamma_P w_p\|_2 \leq |\gamma_P| \|v_p - w_p\|_2$ ; if  $\gamma_D < +\infty$ , then  $\|v_q - \gamma_D w_q\|_2 \leq |\gamma_D| \|v_q - w_q\|_2$ .*

*Proof.* Note that  $v_p^T(v_p - \gamma_P w_p) = 0$  from the definition (16) of  $\gamma_P$ . So,

$$\begin{aligned}
\|v_p - \gamma_P w_p\|_2^2 &= (v_p - \gamma_P w_p)^T (v_p - \gamma_P w_p) \\
&= (\gamma_P v_p - \gamma_P w_p)^T (v_p - \gamma_P w_p) \\
&\leq |\gamma_P| \|v_p - w_p\|_2 \|v_p - \gamma_P w_p\|_2.
\end{aligned}$$

If  $\|v_p - \gamma_P w_p\|_2 = 0$  the first inequality holds; otherwise, we divide both sides by  $\|v_p - \gamma_P w_p\|_2$  to get the result. In the same way, we can prove the second inequality. □

**LEMMA 8.** (a) *Let  $\gamma \in (0, 1]$ . If  $(x, s) \in \mathcal{N}_\infty^-(\beta)$  then  $\|v - \gamma w\|_2 \leq \sqrt{\frac{2n}{1-\beta}}$ .*

(b) *Let  $\gamma \in [1 - \delta/\sqrt{n}, 1]$ . If  $(x, s) \in \mathcal{N}_2(\beta)$  then  $\|v - \gamma w\|_2 \leq \frac{\beta + \delta}{\sqrt{1-\beta}}$  and*

$$\|v - w\|_2 \leq \frac{\beta}{\sqrt{1-\beta}}.$$

*Proof.* (a) If  $(x, s) \in \mathcal{N}_\infty^-(\beta)$ , we have (17). Using (17) and  $Vw = e$ , we get

$$(18) \quad w_j^2 \leq 1/(1-\beta) \quad \text{for each } j.$$

Hence  $\|v - \gamma w\|_2^2 = v^T v - 2\gamma v^T w + \gamma^2 w^T w \leq n + w^T w \leq 2n/(1-\beta)$ .

(b) Since  $(x, s) \in \mathcal{N}_2(\beta) \subset \mathcal{N}_\infty^-(\beta)$ , we have (17) and  $\|V(v - w)\|_2 \leq \beta$  (see (8)).

So,

$$\begin{aligned}
\|v - \gamma w\|_2 &\leq (1 - \gamma)\|v\|_2 + \gamma\|v - w\|_2 \\
&\leq (1 - \gamma)\sqrt{n} + \gamma\|V(v - w)\|_2\|V^{-1}e\|_\infty \\
&\leq \delta + \gamma\beta/\sqrt{1 - \beta} \\
&\leq (\delta + \beta)/\sqrt{1 - \beta}.
\end{aligned}$$

If  $\gamma = 1$ , we can set  $\delta = 0$  to get the last inequality.  $\square$

**LEMMA 9.** *Let  $\beta \in (0, 1)$  and  $\delta > 0$  be constants, and let  $\bar{\gamma} = \gamma_P$  if  $\gamma_P \in (0, 1]$  and  $\bar{\gamma} = \gamma_D$  otherwise.*

(a) *If  $(x, s) \in \mathcal{N}_\infty^-(\beta)$ ,  $\gamma' \in (0, 1]$ , and  $\gamma = \gamma'\bar{\gamma}$ , then we have*

$$\|dX(\gamma)ds(\gamma)\|_\infty^- \leq 4\bar{\gamma}n\mu/(1 - \beta).$$

(b) *If  $(x, s) \in \mathcal{N}_2(\beta)$ ,  $\gamma' = 1 - \delta/\sqrt{n}$ , and  $\gamma = \gamma'\bar{\gamma}$ , then we have*

$$\|dX(\gamma)ds(\gamma)\|_2 \leq \bar{\gamma}\mu(3 + \delta)^2/(1 - \beta).$$

*Proof.* Without loss of generality, we assume that  $\gamma_P \in (0, 1]$  and  $\bar{\gamma} = \gamma_P$  in the analysis below.

(a) Let  $(x, s) \in \mathcal{N}_\infty^-(\beta)$  and  $\gamma' \in (0, 1]$ . Since  $\gamma = \gamma'\gamma_P$ , we have

$$(19) \quad \|v_p - \gamma w_p\|_2 \leq \|v_p - \gamma_P w_p\|_2 + \gamma_P(1 - \gamma')\|w_p\|_2.$$

Using Lemma 7, Lemma 8(a), and (18), we get

$$\begin{aligned}
\|v_p - \gamma w_p\|_2 &\leq \gamma_P\sqrt{\frac{2n}{1 - \beta}} + \gamma_P(1 - \gamma')\sqrt{\frac{n}{1 - \beta}} \\
&\leq 2\gamma_P\sqrt{\frac{2n}{1 - \beta}}.
\end{aligned}$$

Using Lemma 8(a) we also have

$$\|v_q - \gamma w_q\|_2 \leq \sqrt{\frac{2n}{1 - \beta}}.$$

By (10), we obtain

$$\begin{aligned}
\|dX(\gamma)ds(\gamma)\|_\infty^- &= \mu\|(V_p - \gamma W_p)(v_q - \gamma w_q)\|_\infty^- \\
&\leq \mu\|v_p - \gamma w_p\|_2\|v_q - \gamma w_q\|_2 \\
&\leq 4\gamma_P n\mu/(1 - \beta).
\end{aligned}$$

(b) Let  $(x, s) \in \mathcal{N}_2(\beta)$  and  $\gamma' = 1 - \delta/\sqrt{n}$ . Using (18), (19), and Lemma 7 we have

$$\begin{aligned} \|v_p - \gamma w_p\|_2 &\leq \gamma_P \|v - w\|_2 + \gamma_P(1 - \gamma') \sqrt{\frac{n}{1 - \beta}} \\ &\leq \gamma_P \left( \frac{\beta}{\sqrt{1 - \beta}} + \frac{\delta}{\sqrt{1 - \beta}} \right). \end{aligned}$$

The last inequality follows from Lemma 8(b). We also see

$$\begin{aligned} \|v_q - \gamma w_q\|_\infty &= \|(v - \gamma w) - (v_p - \gamma w_p)\|_\infty \\ &\leq \|v\|_\infty + \|w\|_\infty + \|v_p - \gamma w_p\|_2 \\ &\leq \sqrt{1 + \beta} + \frac{1}{\sqrt{1 - \beta}} + \frac{\beta + \delta}{\sqrt{1 - \beta}} \\ &\leq \frac{(3 + \delta)}{\sqrt{1 - \beta}}. \end{aligned}$$

So

$$\begin{aligned} \|dX(\gamma)ds(\gamma)\|_2 &= \mu \|(V_q - \gamma W_q)(v_p - \gamma w_p)\|_2 \\ &\leq \mu \|v_q - \gamma w_q\|_\infty \|v_p - \gamma w_p\|_2 \\ &\leq \gamma_P \mu (3 + \delta)^2 / (1 - \beta). \end{aligned}$$

□

**THEOREM 10.** *Let  $\beta \in (0, 1)$  and  $\delta > 0$  be constants independent of the input data of (P) and (D).*

(a) *Algorithm I ( $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$  and  $\gamma' \in (0, 1)$  is independent of the input data) terminates in  $O(nt)$  iterations and is strictly monotone in both objectives.*

(b) *Algorithm II ( $\mathcal{N} = \mathcal{N}_2(\beta)$  and  $\gamma' = 1 - \delta/\sqrt{n}$ ) terminates in  $O(\sqrt{nt})$  iterations and is strictly monotone in both objectives.*

*Proof.* The strict monotonicity follows from Theorem 4.

(a) Since  $(x, s) \in \mathcal{N}_\infty^-(\beta)$  throughout Algorithm I, by Lemma 6(a), Lemma 9(a), and  $\gamma = \gamma' \bar{\gamma}$  we get

$$(x(\alpha), s(\alpha)) \in \mathcal{N}_\infty^-(\beta) \quad \text{for } \alpha \leq \frac{\beta(1 - \beta)\gamma'}{4n}.$$

Then by Lemma 5(b), we have

$$\begin{aligned} (x^{k+1})^T s^{k+1} &\leq (1 - (1 - \gamma)\beta(1 - \beta)\gamma'/4n)(x^k)^T s^k \\ &\leq (1 - (1 - \gamma')\beta(1 - \beta)\gamma'/4n)(x^k)^T s^k \end{aligned}$$

which implies  $(x^k)^T z^k \leq 2^{-t}$  for  $k = O(nt)$ .

(b) As in the proof of (a), the result follows from Lemma 5(b), Lemma 6(b), Lemma

9(b), and  $\gamma = \gamma' \bar{\gamma}$ . In particular, since  $(x, s) \in \mathcal{N}_2(\beta)$  throughout Algorithm II and  $\gamma = \gamma' \bar{\gamma}$ , we find

$$(x(\alpha), s(\alpha)) \in \mathcal{N}_2(\beta) \quad \text{for } \alpha \leq \frac{\beta(1-\beta)\gamma'}{(3+\delta)^2}.$$

□

**5. Typical values of  $\gamma_P$  and  $\gamma_D$ .** Consider Algorithm II. The proof of Theorem 10(b) above shows that we can choose  $\alpha$  at least equal to some constant (independent of  $n$ ) value  $\hat{\alpha} \in (0, 1]$  when  $(x, s) \in \mathcal{N}_2(\beta)$ ,  $\gamma' = 1 - \delta/\sqrt{n}$  and  $\gamma = \gamma' \bar{\gamma}$ . Hence, whenever  $\bar{\gamma} \leq \hat{\gamma}$  for some constant  $\hat{\gamma} \in (0, 1)$ , we could achieve a constant factor reduction in the duality gap by Lemma 5. In this section we consider whether this is likely to occur.

Note that in Algorithm I, we always have  $\gamma \leq \hat{\gamma}$  for constant  $\hat{\gamma} \in (0, 1)$  (in fact,  $\hat{\gamma} = \gamma'$ ), but we can only guarantee that  $\alpha$  is at least some constant divided by  $n$  by Lemma 6 and Lemma 9; hence typical values of  $\gamma_P$  and  $\gamma_D$  are of less interest.

For definiteness, choose  $\beta = 1/3$ , and suppose  $(x, s) \in \mathcal{N}_2(\beta)$ . By (8), we have

$$\|Vv - e\|_2 \leq \beta,$$

and so

$$(20) \quad \|v - e\|_2 \leq \beta$$

since each  $v_j$  is at least as close to 1 as is  $v_j^2$ . In addition,  $w - e = W(e - v)$ , so

$$(21) \quad \begin{aligned} \|w - e\|_2 &\leq \|w\|_\infty \|v - e\|_2 \\ &\leq \frac{\beta}{\sqrt{1-\beta}} \end{aligned}$$

using (18). Since  $\beta = 1/3$ , we get

$$(22) \quad \|v - e\|_2 \leq \frac{1}{2}, \quad \|w - e\|_2 \leq \frac{1}{2},$$

and from  $\|e\|_2 = \sqrt{n}$  we deduce

$$(23) \quad \theta(v, e) \leq \arcsin\left(\frac{1}{2\sqrt{n}}\right), \quad \theta(w, e) \leq \arcsin\left(\frac{1}{2\sqrt{n}}\right),$$

where  $\theta(t, u)$  denotes the angle between vectors  $t$  and  $u$ .

Now  $\gamma_P$  and  $\gamma_D$  are related to the angles between the projections of  $v$  and  $w$  into the null space  $U$  of  $\bar{A}$  or its orthogonal complement. If  $v \neq w$ , it is possible to choose  $U$  (or  $U^\perp$ ) so that  $v_p^T w_p$  (or  $v_q^T w_q$ ) is negative. However, since from (23)

$$\theta(v, w) \leq \arcsin\left(\frac{1}{\sqrt{n}}\right),$$

it appears that for “most” subspaces  $U$ ,  $v_p^T w_p$  will be close to  $\|v_p\|_2^2$  and hence  $\gamma_P$  close to 1. Here we would like to make this statement in some sense precise. As in [MTY90], we assume that

- (24)  $U$  is a random subspace of  $\mathbb{R}^n$  of dimension  $d := n - m$ , drawn from the unique distribution on such subspaces that is invariant under orthogonal transformations.

In order to analyze this situation, we proceed as in the proof of Lemma 6 of [MTY90]. The orthogonal invariance of the subspace  $U$  implies that we can instead assume that  $U$  is the fixed subspace

$$\tilde{U} := \{x \in \mathbb{R}^n : x_{d+1} = \dots = x_n = 0\}$$

and that  $e$ ,  $v$ , and  $w$  are replaced by  $\tilde{e}$ ,  $\tilde{v}$ , and  $\tilde{w}$ , where these vectors are the results of applying the orthogonal transformation  $\vartheta$  taking  $U$  to  $\tilde{U}$  to the vectors  $e$ ,  $v$ , and  $w$ . Hence

- (25)  $\tilde{e}$  is uniformly distributed on the unit sphere of radius  $\sqrt{n}$  in  $\mathbb{R}^n$ ,

and

$$(26) \quad \|\tilde{v} - \tilde{e}\|_2 \leq \frac{1}{2}, \quad \|\tilde{w} - \tilde{e}\|_2 \leq \frac{1}{2},$$

$$(27) \quad \theta(\tilde{v}, \tilde{e}) \leq \arcsin\left(\frac{1}{2\sqrt{n}}\right), \quad \theta(\tilde{w}, \tilde{e}) \leq \arcsin\left(\frac{1}{2\sqrt{n}}\right).$$

Let  $\tilde{e}_p$ ,  $\tilde{v}_p$ , and  $\tilde{w}_p$  denote the projections of  $\tilde{e}$ ,  $\tilde{v}$ , and  $\tilde{w}$  onto  $\tilde{U}$ ; they are also the images under  $\vartheta$  of  $e_p$ ,  $v_p$ , and  $w_p$ . Hence

$$(28) \quad \gamma_P = \frac{\|\tilde{v}_p\|_2^2}{\tilde{v}_p^T \tilde{w}_p}.$$

Now  $\|\tilde{v}_p - \tilde{e}_p\|_2 \leq 1/2$  from (26), so that

$$(29) \quad \|\tilde{v}_p\|_2 \in \left[ \|\tilde{e}_p\|_2 - \frac{1}{2}, \|\tilde{e}_p\|_2 + \frac{1}{2} \right],$$

and similarly

$$(30) \quad \|\tilde{w}_p\|_2 \in \left[ \|\tilde{e}_p\|_2 - \frac{1}{2}, \|\tilde{e}_p\|_2 + \frac{1}{2} \right].$$

Also,  $\tilde{v}_p^T \tilde{w}_p = \frac{1}{2} (\|\tilde{v}_p\|_2^2 + \|\tilde{w}_p\|_2^2 - \|\tilde{v}_p - \tilde{w}_p\|_2^2)$ , so since  $\|\tilde{v}_p - \tilde{w}_p\|_2 \leq \|\tilde{v} - \tilde{w}\|_2 \leq 1$ ,

$$(31) \quad \tilde{v}_p^T \tilde{w}_p \in \left[ \left( \|\tilde{e}_p\|_2 - \frac{1}{2} \right)^2 - \frac{1}{2}, \left( \|\tilde{e}_p\|_2 + \frac{1}{2} \right)^2 \right].$$

Hence, if  $\|\tilde{e}_p\|_2 \geq 3/2$ ,

$$(32) \quad \gamma_P \in \left[ \frac{\left( \|\tilde{e}_p\|_2 - \frac{1}{2} \right)^2}{\|\tilde{e}_p\|_2 + \frac{1}{2}}, \frac{\left( \|\tilde{e}_p\|_2 + \frac{1}{2} \right)^2}{\left( \|\tilde{e}_p\|_2 - \frac{1}{2} \right)^2 - \frac{1}{2}} \right],$$

$$(33) \quad \gamma_P \geq 1 - \frac{2}{\|\tilde{e}_p\|_2 + \frac{1}{2}}.$$

**THEOREM 11.** *Suppose assumption (24) holds, and  $d = \Omega(n)$ . Then, with probability approaching 1 as  $n \rightarrow \infty$ ,*

$$\gamma_P \geq 1 - \frac{\xi}{\sqrt{n}}$$

for some constant  $\xi$ .

*Proof.* We use (33). As in Lemma 6 of [MTY90], we can suppose

$$\tilde{e} = \sqrt{n} \left( \frac{\lambda_1}{\|\lambda\|_2}, \dots, \frac{\lambda_n}{\|\lambda\|_2} \right)^T$$

from (25), where  $\lambda_1, \dots, \lambda_n$  are independent normal random variables with mean 0 and variance 1. Then

$$\tilde{e}_p = \sqrt{n} \left( \frac{\lambda_1}{\|\lambda\|_2}, \dots, \frac{\lambda_d}{\|\lambda\|_2}, 0, \dots, 0 \right)^T,$$

and

$$\|\tilde{e}_p\|_2^2 = n \left( \frac{\lambda_1^2 + \dots + \lambda_d^2}{\lambda_1^2 + \dots + \lambda_n^2} \right).$$

The quantity in parenthesis is a beta random variable with parameters  $d/2$  and  $m/2$ , which has mean  $d/n$  and variance  $2dm/n^2(n+2) \geq 1/(2n+4)$ . Using Chebyshev's inequality,

$$\|\tilde{e}_p\|_2^2 \geq n \left( \frac{d}{2n} \right)$$

with probability approaching 1 as  $n \rightarrow \infty$ . Since  $d = \Omega(n)$ , this shows  $\|\tilde{e}_p\|_2 = \Omega(\sqrt{n})$  with high probability, and hence completes the proof.  $\square$

Since  $U^-$  satisfies (24) if  $U$  does, the same result shows that  $\gamma_D \geq 1 - \xi/\sqrt{n}$  with high probability as long as  $m = \Omega(n)$ . These results suggest that we cannot expect to find  $\tilde{\gamma}$  ( $= \gamma' \gamma_P$  or  $\gamma' \gamma_D$ ) smaller than  $1 - \kappa/\sqrt{n}$  for some constant  $\kappa$  very often. Note that several primal-dual interior-point algorithms use a value for  $\gamma$  that is  $1 - \zeta/\sqrt{n}$  for some constant  $\zeta$  (e.g., [KMY89], [MA89]); our result indicates that this value is close to one that improves both objectives at almost every iteration.

**6. Concluding remarks.** In an implementation of the strictly monotone algorithm described in Section 3, we may use a small parameter value of  $\gamma'$  (for example  $\gamma' = 1/n$ ) and long step sizes such that the next primal and dual iterates are located a fixed ratio (say .99) of the way from the current point to the boundary of the corresponding feasible region. In that case, the next iterate is well-defined and the duality gap decreases strictly, although these phenomena have not been shown in non-monotone versions of the algorithm.

## REFERENCES

- [An86] K.M. Anstreicher, A Monotonic Projective Algorithm for Fractional Linear Programming, *Algorithmica* 1 (1986), pp. 483–498.
- [An91] K.M. Anstreicher, On Monotonicity in the Scaled Potential Algorithm for Linear Programming, *Linear Algebra Appl.* 152 (1991), pp. 223–232.
- [BL89] 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, *Trans. Amer. Math. Soc.* 314 (1989), pp. 499–581.
- [KMY88] 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, 1988, pp. 29–47.
- [KMY89] M. Kojima, S. Mizuno, and A. Yoshise, A Polynomial Time Algorithm for a Class of Linear Complementarity Problems, *Math. Programming* 44 (1989), pp. 1–26.
- [KMNY91] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise, *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, Lecture Notes in Computer Science No. 538, Springer-Verlag, New York, 1991.
- [LMS91] I. J. Lustig, R. E. Marsten, and D. F. Shanno, Computation Experience with a Primal-Dual Interior Point Method for Linear Programming, *Linear Algebra Appl.* 152 (1991), pp. 191–222.
- [MMS89] K. A. McShane, C. L. Monma, and D. Shanno, An Implementation of a Primal-Dual Interior Point Method for Linear Programming, *ORSA J. Comp.* 1 (1989), pp. 70–83.
- [Meg88] 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, pp. 131–158.
- [MN92a] S. Mizuno and A. Nagasawa, Strict Monotonicity in Todd's Low-Complexity Algorithm for Linear Programming, *Oper. Res. Lett.* 12 (1992), pp. 59–64.
- [MN92b] S. Mizuno and A. Nagasawa, A Primal-Dual Affine Scaling Potential Reduction Algorithm for Linear Programming, Research Memorandum No. 427 (1992), The Institute of Statistical Mathematics, Tokyo, Japan (to appear in *Math. Programming*).
- [MTY90] S. Mizuno, M.J. Todd, and Y. Ye, On Adaptive-Step Primal-Dual Interior-Point Algorithms for Linear Programming, TR-944 (1990), School of OR and IE, Cornell University, Ithaca, NY (to appear in *Math. Oper. Res.*).
- [MA89] R.C. Monteiro and I. Adler, Interior Path Following Primal-Dual Algorithms. Part I: Linear Programming, *Math. Programming* 44 (1989), pp. 27–42.
- [MAR90] 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, *Math. Oper. Res.* 15 (1990), pp. 191–214.
- [So85] 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-Verlag, New York, 1985, pp. 866–876.
- [Tu92] L. Tunçel, A Note on the Primal-Dual Affine Scaling Algorithms, TR-1004 (1992), SORIE Cornell University, Ithaca, NY.
- [ZT90] Y. Zhang, and R.A. Tapia, A Quadratically Convergent Polynomial Primal-Dual Interior Point Algorithm for Linear Programming, TR90-40 (1990), Department of Mathematical Sciences, Rice University, Houston, TX.