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

TECHNICAL REPORT NO. 1014

July 1992

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

by

Shinji Mizuno*, Michael J. Todd**
and Levent Tuncel***

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

**Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

**Research supported in part by Mathematical Sciences Institute of Cornell University, and also by NSF, AFOSR and ONR through NSF grant DMS-8920550.

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

*Shinji MIZUNO**
Michael J. TODD†
Levent TUNÇEL‡

School of Operations Research and Industrial Engineering
Cornell University
Ithaca, NY 14853-3801

July 1992

Abstract

We study monotonicity of primal and dual objective values in the framework of primal-dual interior-point methods. We show that the primal-dual affine-scaling algorithm is monotone in both objectives. Then we derive a condition under which a primal-dual interior-point algorithm is monotone. Finally, 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 those that will almost always improve both objectives.

Keywords: Linear programming, primal-dual, interior-point methods, monotonicity.

---

*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.

†Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

‡Research supported in part by Mathematical Sciences Institute of Cornell University, and also by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

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 [MN91]). 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 is 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.

First, we show that the primal-dual affine-scaling algorithm [MAR90, KMNY91, MN92, Tu92] is strictly monotone in both primal and dual objective values. Then we derive a condition on the centering parameter under which a primal-dual interior-point algorithm 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 will allow monotonicity in both objectives and 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 $O(\sqrt{n}t)$ 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 4.5).

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 will improve both objectives at almost every iteration.

2 Preliminaries

We consider linear programming problems in the following primal (P) and dual (D) forms:

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

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

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Without loss of generality, we will assume A has full row rank 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 will 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. $X = \mathrm{diag}(x)$), and e will denote the vector of ones. We will 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. Now, 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 $\mu > 0$):

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

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 as given by Mizuno et al. [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, μ 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 ∞-norm or just one side of the ∞-norm, wider neighborhoods have been defined and used by Kojima et al. [KMY88], Mizuno et al. [MTY90], and Zhang and Tapia [ZT90]. Using the ∞-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 ∞-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 $\mathcal{N}_\infty^-(\beta)$ is 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]):

$$\begin{aligned} A dx &= 0, && (4) \\ A^T dy + ds &= 0, && (5) \\ S dx + X ds &= \gamma\mu e - Xs, && (6) \end{aligned}$$

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 α 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 γ and the step size α may vary from one iteration to the next). We will 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

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

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

$$\mathcal{N}_2(\beta) = \{(x, s) \in \mathcal{F}_0 : \|V(v - w)\|_2 \le \beta\}, \tag{8}$$

$$\mathcal{N}_\infty^-(\beta) = \{(x, s) \in \mathcal{F}_0 : \|V(v - w)\|_\infty^- \le \beta\}, \tag{9}$$

$$dX(\gamma)ds(\gamma) = \mu(V_p - \gamma W_p)(v_q - \gamma w_q), \tag{10}$$

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

3 Monotonicity

First we will show that the primal-dual affine-scaling algorithms (see Monteiro, Adler and Resende [MAR90], Kojima, Megiddo, Noma, and Yoshise [KMNY91], Mizuno and Nagasawa [MN92], and Tunçel [Tu92]) are strictly monotone in both objectives:

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

$$v_p = \frac{1}{\sqrt{\mu}}\,\bar{c}_p. \tag{11}$$

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

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

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

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

We immediately have the following theorem.

Theorem 3.1. *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. □

From now on, without loss of generality we will 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:

$$c^T x(\alpha) - c^T x = -\alpha\mu\left(\|v_p\|_2^2 - \gamma v_p^T w_p\right), \tag{14}$$

and, similarly,

$$b^T y(\alpha) - b^T y = \alpha\mu\left(\|v_q\|_2^2 - \gamma v_q^T w_q\right). \tag{15}$$

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

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

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 γ_P and γ_D.

Lemma 3.1. *Either*

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

or

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

Moreover, $\gamma_P = 1$ *if and only if* $\gamma_D = 1$.

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 $\gamma_P \in (-\infty, 0) \cup (1, \infty]$. □

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

Lemma 3.2. *If $0 \le \gamma \le \bar{\gamma}$ ($0 \le \gamma < \bar{\gamma}$), 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 3.1 and equations (14) and (15). □

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

Theorem 3.2. *An algorithm whose search directions are $dx(\gamma)$ and $ds(\gamma)$ for $0 \le \gamma \le \bar{\gamma}$ ($0 \le \gamma < \bar{\gamma}$) at each iterate is (strictly) monotone in both primal and dual objective values.*

Proof: Directly follows from Lemma 3.2. □

If the exceptional case ($\|v_p\|_2^2 = 0$ or $\|v_p\|_2^2 = n$) is excluded, Theorem 3.1 is a special case of Theorem 3.2. Although the condition in Theorem 3.2 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 $(x^0)^T s^0 \le 2^t$ 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, γ_P and γ_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 $O(\sqrt{n}t)$ iterations when $\mathcal{N} = \mathcal{N}_2(\beta)$ and $\gamma' = 1 - \delta/\sqrt{n}$ for a positive constant δ (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 will show that Algorithm I terminates in $O(nt)$ iterations and Algorithm II terminates in $O(\sqrt{n}t)$ iterations.

Note that if $\bar{\gamma}$ is small, the centering parameter γ in the algorithm is small even if γ' 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 4.1 - 4.2) are standard in the primal-dual framework (see for instance Mizuno, Todd and Ye [MTY90]).

Lemma 4.1.
(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 $dx(\gamma)^T ds(\gamma) = 0$. $\square$

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

Lemma 4.2.
(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

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

So, using Lemma 4.1 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 $\alpha \in [0,1]$, the right hand side is always non-negative. 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 Lemma 4.1 we obtain

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

$\square$

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 ϵ that is close to 1 (or bounded away from zero). Instead, we will show that the second-order terms $\|dX ds\|_\infty^-$ and $\|dX ds\|_2$ 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 4.3. *$\|v_p - \gamma_P w_p\|_2 \leq |\gamma_P| \|v_p - w_p\|_2$ and $\|v_q - \gamma_D w_q\|_2 \leq |\gamma_D| \|v_q - w_q\|_2$.*

Proof: Note that $v_p^T(v_p - \gamma_P w_p) = 0$ from the definition (16) of γ_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 4.4.
(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 $\|v - w\|_2 \leq \beta/\sqrt{1-\beta}$.

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

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

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 $\|V(v - w)\| \leq \beta$ (see (8)). So,

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

$$
\begin{aligned}
&\leq \delta + \gamma\beta/\sqrt{1-\beta} \\
&\leq (\delta+\beta)/\sqrt{1-\beta}.
\end{aligned}
$$

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

Lemma 4.5. *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

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

Using Lemma 4.3, Lemma 4.4(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 4.4(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 4.3 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 4.4(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 4.1. *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 $O(\sqrt{n}t)$ iterations and is strictly monotone in both objectives.

Proof: The strict monotonicity follows from Theorem 3.2.
(a) Since $(x,s) \in \mathcal{N}_\infty^-(\beta)$ throughout Algorithm I, by Lemma 4.2(a), Lemma 4.5(a), and $\gamma = \gamma'\bar{\gamma}$ we get

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

Then by Lemma 4.1(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 Lemmas 4.1(b), 4.2(b), 4.5(b), and $\gamma = \gamma' \bar{\gamma}$. □

5 Typical values of γ_P and γ_D

Consider Algorithm II. Lemmas 4.2 and 4.5 show that we can choose α 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 4.1. In this section we consider whether this is likely to occur.

Note that in Algorithm I, we always have $\gamma \leq \bar{\gamma}$ for constant $\bar{\gamma} \in (0, 1)$ (in fact, $\bar{\gamma} = \gamma'$), but we can only guarantee that α is at least some constant divided by n by Lemmas 4.2 and 4.5; hence typical values of γ_P and γ_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

$$\|v - e\|_2 \leq \beta \tag{20}$$

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

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

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

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

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

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

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

Now γ_P and γ_D are related to the angles between the projections of v and w into the null space of 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 γ_P close to 1. Here we would like to make this statement in some sense precise. As in [MTY90], we will assume that

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

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

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

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 ϑ taking U to $\tilde{U}$ to the vectors e, v, and w. Hence

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

and

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

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

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 ϑ of e_P, v_p, and w_p. Hence

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

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

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

and similarly

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

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

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

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

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

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

Theorem 5.1. *Suppose assumption (24) holds, and $d = \Omega(n)$. Then, with probability approaching 1 as $n \to \infty$,*

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

for some constant ξ.

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

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

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

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

and

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

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

Since $U^\perp$ 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 $\bar{\gamma}$ ($= \gamma'\gamma_P$ or $\gamma'\gamma_D$) smaller than $1 - \kappa/\sqrt{n}$ for some constant κ very often. Note that several primal-dual interior-point algorithms use a value for γ that is $1 - \zeta/\sqrt{n}$ for some constant ζ (e.g., [KMY89], [MA89]); our result indicates that this value is close to one that will improve 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 γ' (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, pp. 483-498.

[An91] K.M. Anstreicher, On Monotonicity in the Scaled Potential Algorithm for Linear Programming, *Linear Algebra and Its Applications* 152, 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, *Transactions of the American Mathematical Society 314*, 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, pp. 29-47.

[KMY89] M. Kojima, S. Mizuno, and A. Yoshise, A Polynomial Time Algorithm for a Class of Linear Complementarity Problems, *Mathematical Programming* 44, 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 Note in Computer Science* 538, Springer-Verlag, New York.

[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 and Its Applications* 152, 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 Journal on Computing* 1, 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, pp.131-158.

[MN91] S. Mizuno and A. Nagasawa, Strict Monotonicity in Todd's Low-Complexity Algorithm for Linear Programming, *(to appear in OR Letters)*.

[MN92] S. Mizuno and A. Nagasawa, A Primal-Dual Affine Scaling Potential Reduction Algorithm for Linear Programming, Research Memorandum No. 427, The Institute of Statistical Mathematics, Tokyo, Japan.

[MTY90] S. Mizuno, M.J. Todd, and Y. Ye, On Adaptive-Step Primal-Dual Interior-Point Algorithms for Linear Programming, *TR-944 (to appear in Math. of OR)*, School of OR and IE, Cornell University, Ithaca, NY.

[MA89] R.C. Monteiro and I. Adler, Interior Path Following Primal-Dual Algorithms. Part I: Linear Programming, *Mathematical Programming* 44, 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. of OR* 15, 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, New York, pp. 866-876.

[Tu92] L. Tunçel, A Note on the Primal-Dual Affine Scaling Algorithms, *TR-1004*, 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*, Department of Mathematical Sciences, Rice University, Houston, TX.