A PRIMAL-DUAL AFFINE SCALING ALGORITHM WITH NECESSARY CENTERING AS A SAFEGUARD

Gongyun Zhao †, Jie Sun ‡ and Jishan Zhu ⋆

November 1994

Technical Report No. 647 †

Abstract. An interior point algorithm for linear programming is proposed and analysed. At each iteration this algorithm finds a point in a wide neighborhood along a primal-dual affine scaling direction, and it adds a centering step only if the step along the affine scaling direction is too short. Moreover, the centering direction is computed with the old projection matrix which was obtained in the computation of the affine scaling direction. The best complexity upper bound to date is achieved by this algorithm.

Key Words. Complexity, linear programming, interior point method, predictor-corrector algorithm

Abbreviated Title. Safeguarded affine scaling algorithm

AMS(MOS) subject classifications. 90C60, 90C05

† Department of Mathematics, National University of Singapore, 0511, SINGAPORE. Email: matzgy@nus.sg

‡ Department of Decision Sciences, National University of Singapore, 0511, SINGAPORE. Email: fbasunj@nus.sg

⋆ Department of Decision Sciences, National University of Singapore, 0511, SINGAPORE. Email: fbazhujs@nus.sg

1. Introduction

We consider the following linear programming problem

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{LP}$$

and its dual

$$\begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & A^T y + s = c \\ & s \geq 0, \end{aligned} \tag{LD}$$

where A is an $m \times n$-matrix with $m \leq n$, b is an m-vector and c is an n-vector. We denote by $\mathcal{F}$ the feasible region of the problems (LP) and (LD)

$$\mathcal{F} := \{(x, s) \mid x \geq 0,\ s \geq 0,\ Ax = b,\ s + A^T y = c \ \text{ for some } y\},$$

and by $\mathcal{F}^0 := \{(x, s) \in \mathcal{F} \mid x > 0, s > 0\}$ the interior of $\mathcal{F}$. A commonly used wide neighborhood $\mathcal{N}_\infty^-(\beta)$ of the central path is defined as

$$\mathcal{N}_\infty^-(\beta) := \{(x, s) \in \mathcal{F} \mid x \circ s \geq \beta \frac{x^T s}{n} e\},$$

for some $\beta \in (0, 1)$.

A general search direction (d_x, d_s, d_y) in primal-dual interior point methods is the solution of the following equation system

$$\begin{aligned} & s \circ d_x + x \circ d_s = \gamma \mu e - x \circ s \\ & A d_x = 0 \\ & A^T d_y + d_s = 0, \end{aligned}$$

where $\mu := x^T s / n$ and the notation "$\circ$" represents the Hadamard product of vectors; that is, for any two vectors x and s

$$x \circ s := (x_1 s_1, \dots, x_n s_n)^T.$$

We also use the notation

$$x^v := (x_1^v, \dots, x_n^v)^T$$

throughout this paper. For $\gamma = 0$ and $\gamma = 1$ the directions are called *affine scaling direction* and *centering direction*, respectively. A direction combining these two directions corresponds to $\gamma \in (0, 1)$.

Many numerical experiments, eg. [1] [4], as well as the analysis of anticipated behavior, cf. [7], have shown the great efficiency of the *affine scaling algorithms* which use only the affine scaling direction and generate iterates in the neighborhood $\mathcal{N}_\infty^-(\beta)$. However, when starting at a point close to the boundary of the feasible region these algorithms may fail to solve a problem or may be inefficient, cf. [1] [6]. The polynomial complexity of such algorithms has been proved only if the starting point is close to the central path and short steps are used, cf. [3] [9].

An amendment to the affine scaling algorithms is to increase γ from 0 to some positive number. There have been a lot of works devoted to such algorithms with $\gamma \in (0,1)$, e.g. [5] [8] [10] [12], where a polynomial bound of $O(n\ln(\mu_0/\epsilon))$ is achieved for the number of iterations needed to reduce the parameter μ from μ_0 to less than ϵ. This upper bound, however, depends on γ, hence it is not valid for $\gamma = 0$. Actually it is only true for $0 < \gamma = O(1)$ independent of n. Thus there is a gap between the practice and the theory: In practice small γ makes algorithm more efficient, but theoretically small γ causes large upper bound of the complexity.

In this paper we propose an alternative amendment to the affine scaling algorithms. At each iteration, our algorithm finds a point in the neighborhood $\mathcal{N}_\infty^-(\beta)$ along an affine scaling direction, and it adds a centering step only if the step length θ in the affine scaling direction is too short, say $\theta \le \theta_{\min}$ where $\theta_{\min}$ is a given constant. The centering step is therefore regarded as a safeguard. Furthermore, the centering direction is computed with the old projection matrix which was obtained in the computation of the previous affine scaling direction. Hence these centering steps are cheap. We may also regard this algorithm as a predictor-corrector algorithm with incomplete correctors.

Practically, the proposed algorithm performs exactly the same as the efficient affine scaling algorithms if no short step occurs. On the other hand, if short steps are encountered in solving some tough problems, the corrector steps will then serve as a safeguard to prevent the algorithm from failure and to ensure the polynomial complexity. It is worth noticing that by implementing the affine scaling algorithm or the above mentioned algorithms with small $\gamma > 0$, if the process encounters a short step, then a number of such short steps will be likely to come successively. By our algorithm, however, as soon as it is alerted by $\theta \le \theta_{\min}$, a centering step will be performed to avoid the successive short steps. Hence such a safeguard is really necessary in practice.

Theoretically, although old projection matrices are used at the corrector-step, we can show that the iteration complexity of the algorithm is bounded by $\frac{30n}{\beta}\ln\frac{\mu_0}{\epsilon}$. This is one of the best complexity bounds known to date for the large-step algorithms using $\mathcal{N}_\infty^-(\beta)$, Moreover, this bound is independent of the parameter γ.

2. The Algorithm

In this section we describe the algorithm in details. We make the following two assumptions throughout this paper:

A1: Matrix A is of full rank, i.e. $\text{rank}(A) = m$;
A2: The interior of the feasible region $\mathcal{F}^0$ is nonempty.

The proposed algorithm follows the central path $(x(\mu), s(\mu))$ to a pair of optimal solutions of (LP) and (LD) as $\mu \downarrow 0$, where $(x(\mu), s(\mu))$ is the solution to the system

$$\begin{aligned} x \circ s &= \mu e \\ Ax &= b \\ A^T y + s &= c \\ x \geq 0&, s \geq 0. \end{aligned} \tag{LPD}$$

If $\mu = 0$, the solution of (LPD) satisfies the Kuhn-Tucker conditions and hence is a pair of optimal solutions of (LP) and (LD). For more details about the central path, the reader is referred to [2], [11], [13] and [14].

Each iteration in the algorithm consists of a predictor step and a possible corrector step in which the search directions are the *affine scaling direction* (d_x^a, d_s^a, d_y^a) and the *centering direction* (d_x^c, d_s^c, d_y^c) respectively. The affine scaling direction is defined by

$$\begin{aligned} s \circ d_x^a + x \circ d_s^a &= -x \circ s \\ A d_x^a &= 0 \\ A^T d_y^a + d_s^a &= 0. \end{aligned} \tag{2.1}$$

Denote

$$D = \operatorname{diag}\,(x^{1/2} \circ s^{-1/2}), \qquad P = DA^T(AD^2A^T)^{-1}AD, \qquad Q = I - P. \tag{2.2}$$

Here P and Q are orthogonal projection matrices. One can verify that

$$D^{-1} d_x^a + D d_s^a = -(x \circ s)^{1/2}, \tag{2.3}$$

and the solution of (2.1) is

$$d_s^a = -D^{-1} P (x \circ s)^{1/2}, \qquad d_x^a = -DQ(x \circ s)^{1/2}. \tag{2.4}$$

A predictor step at (x, s) results in the following point:

$$\bar{x} = x + \theta d_x^a, \qquad \bar{s} = s + \theta d_s^a, \qquad \bar{y} = y + \theta d_y^a. \tag{2.5}$$

Here the step length θ is determined such that $(\bar{x}, \bar{s})$ reaches the boundary of $\mathcal{N}_\infty^-(\beta)$.

If the step length θ is too short, say $\theta \leq \theta_{\min}$ where $\theta_{\min}$ is a given constant, then the algorithm will perform a corrector step at $(\bar{x}, \bar{s})$, using the *old* projection matrices P and Q as well as D at (x, s) defined in (2.2). Precisely, the centering direction (d_x^c, d_s^c) at the corrector step is defined as follows.

$$d_s^c = D^{-1} P \bar{\eta}, \qquad d_x^c = DQ\bar{\eta}, \qquad d_y^c = -(AD^2A^T)^{-1} AD\bar{\eta}, \tag{2.6}$$

where

$$\bar{\eta} := (\bar{x} \circ \bar{s})^{-1/2} \circ (\bar{\mu} e - \bar{x} \circ \bar{s}), \qquad \bar{\mu} := \frac{\bar{x}^T \bar{s}}{n}. \tag{2.7}$$

Lemma 2.1: The centering direction (d_x^c, d_s^c) defined in (2.6) satisfies

$$\begin{aligned} D^{-1}d_x^c + Dd_s^c &= \bar{\eta} \\ Ad_x^c &= 0 \\ A^T d_y^c + d_s^c &= 0. \end{aligned} \tag{2.8}$$

Proof: By (2.6),

$$D^{-1}d_x^c + Dd_s^c = Q\bar{\eta} + P\bar{\eta} = \bar{\eta}.$$

Using (2.2),

$$Ad_x^c = ADQ\bar{\eta} = (AD - ADP)\bar{\eta} = 0.$$

$$d_s^c + A^T d_y^c = D^{-1}P\bar{\eta} - A^T(AD^2A^T)^{-1}AD\bar{\eta} = 0. \qquad \#$$

Remark: The equations (2.8) define the exact centering direction at (x, s) if $\bar{\eta}$ is replaced by $\eta := (x \circ s)^{-1/2}(\mu e - x \circ s)$ where $\mu := x^T s/n$.

Now we are ready to present the algorithm.

Algorithm: Given $\epsilon > 0$, $\beta \in (0, 1)$, and $\theta_{\min} > 0$, (e.g., $\epsilon = 10^{-6}, \beta = 0.01$ and $\theta_{\min} = 0.1$). Let $(x, s) \in \mathcal{N}_\infty^-(\beta)$ be the current point.

0) **Stopping Test:** If $\mu := \frac{x^T s}{n} \leq \epsilon$, stop. Otherwise,
1) **Predictor Step:** Compute (d_x^a, d_s^a) by (2.4). Define $(\bar{x}, \bar{s}) = (x, s) + \theta(d_x^a, d_s^a)$, where θ is the largest number such that $(\bar{x}, \bar{s}) \in \mathcal{N}_\infty^-(\beta)$. If $\theta > \theta_{min}$, go to Stopping Test with $(x, s) := (\bar{x}, \bar{s})$. Otherwise,
2) **Corrector Step:** Compute (d_x^c, d_s^c) by (2.6). Determine $\alpha = \bar{\alpha}$ by

$$\begin{aligned} \max \quad & \lambda \\ \text{s.t.} \quad & (\bar{x} + \alpha d_x^c) \circ (\bar{s} + \alpha d_s^c) \geq (\bar{x} + \alpha d_x^c)^T(\bar{s} + \alpha d_s^c)\lambda e \\ & \alpha \geq 0. \end{aligned} \tag{2.9}$$

Set $x = \bar{x} + \bar{\alpha} d_x^c$, $s = \bar{s} + \bar{\alpha} d_s^c$. Go to Stopping Test. #

3. Complexity

The complexity analysis of the algorithm consists of two parts for the predictor step and the corrector step, respectively. After that, the main theorem will be presented showing that the iteration complexity of the algorithm is $O(n \ln(\mu_0/\epsilon))$. We denote $\theta_* := O(\beta/n)$ and assume that $\theta_{\min} \geq \theta_*$. If $\theta \geq \theta_{\min}$ at a predictor step, then no corrector step is performed; otherwise, if $\theta_* < \theta < \theta_{\min}$, then a corrector step follows and in both cases the duality gap is reduced by a factor of $1 - O(\beta/n)$. Finally, if $\theta \leq \theta_*$, then a corrector step is performed which, as we will show below, generates a point in $\mathcal{N}_\infty^-(\beta + O(\beta/n))$. The immediately followed predictor step will then have a step length greater than θ_*. This leads to the main theorem that the overall complexity of the algorithm is of the order $O(n\ln(\mu_0/\epsilon))$.

3.1. Analysis of the corrector step

Let $(x,s)\in\mathcal{N}_\infty^-(\beta)$ be the point we have at the beginning of a predictor step, and (d_x^a,d_s^a) be the affine scaling direction at (x,s) defined by (2.4). Let $(\bar{x},\bar{s})=(x,s)+\theta(d_x^a,d_s^a)$ be the resulting point at this predictor step. Suppose now $0\le\theta\le\theta_{min}$, then a corrector step should be carried out. Let (d_x^c,d_s^c) be the centering direction defined by (2.6) where D,P and Q are computed for (x,s) at the predictor step. We first give a list of estimates.

Lemma 3.1: *Using the notations defined in (2.1) – (2.7) and denoting $\mu:=x^Ts/n$, $\bar{\mu}:=\bar{x}^T\bar{s}/n$, we have the following estimates.*

$$\|x\circ s\|_\infty\le\|(x\circ s)^{1/2}\|_2^2=n\mu, \tag{3.1}$$

$$\|Dd_s^a\|_2\le\sqrt{n\mu},\qquad \|D^{-1}d_x^a\|_2\le\sqrt{n\mu}, \tag{3.2}$$

$$\|Dd_s^c\|_2\le\sqrt{\frac{n\bar{\mu}}{\beta}},\qquad \|D^{-1}d_x^c\|_2\le\sqrt{\frac{n\bar{\mu}}{\beta}}, \tag{3.3}$$

and

$$\bar{\mu}=(1-\theta)\mu. \tag{3.4}$$

In addition, if $\theta\le 1/15$, then we have

$$\|\bar{x}\circ\bar{s}-x\circ s\|_\infty\le\frac{16}{15}\theta n\mu. \tag{3.5}$$

Proof: The estimate (3.1) follows from $\|(x\circ s)^{1/2}\|_2^2=x^Ts$.

Using equalities in (2.4)

$$\|Dd_s^a\|_2=\|P(x\circ s)^{1/2}\|_2\le\|(x\circ s)^{1/2}\|_2=\sqrt{n\mu},$$

$$\|D^{-1}d_x^a\|_2=\|Q(x\circ s)^{1/2}\|_2\le\|(x\circ s)^{1/2}\|_2=\sqrt{n\mu},$$

we obtain (3.2).

Since $||Dd_s^c||=||P\bar{\eta}||\le||\bar{\eta}||$ and $||D^{-1}d_x^c||=||Q\bar{\eta}||\le||\bar{\eta}||$ by (2.6), for (3.3) we only need to estimate $\bar{\eta}$ which is defined by (2.7).

$$\begin{aligned}\|\bar{\eta}\|_2^2&=\sum_{i=1}^n\frac{(\bar{\mu}-\bar{x}_i\bar{s}_i)^2}{\bar{x}_i\bar{s}_i}\\&=\sum_{i=1}^n\left(\frac{\bar{\mu}^2}{\bar{x}_i\bar{s}_i}-2\bar{\mu}+\bar{x}_i\bar{s}_i\right)\\&=\sum_{i=1}^n\frac{\bar{\mu}^2}{\bar{x}_i\bar{s}_i}-2n\bar{\mu}+n\bar{\mu}\\&\le\frac{n\bar{\mu}}{\beta}-n\bar{\mu}\qquad\left(\text{ since }\frac{\bar{x}_i\bar{s}_i}{\bar{\mu}}\ge\beta\right)\\&\le\frac{n\bar{\mu}}{\beta}.\end{aligned}$$

We have the following for (3.4).

$$\begin{aligned}
n\bar{\mu} &= (x+\theta d_x^a)^T(s+\theta d_s^a)\\
&= x^T s + \theta(x^T d_s^a + s^T d_x^a) + \theta^2 (d_x^a)^T d_s^a\\
&= x^T s - \theta x^T s \qquad\qquad \text{(using (2.1))}\\
&= (1-\theta) n\mu.
\end{aligned}$$

The inequality (3.5) is proven below.

$$\begin{aligned}
\|\bar{x}\circ\bar{s} - x\circ s\|_\infty &= \|(x+\theta d_x^a)\circ(s+\theta d_s^a) - x\circ s\|_\infty\\
&= \|\theta(x\circ d_s^a + s\circ d_x^a) + \theta^2 d_x^a \circ d_s^a\|_\infty\\
&= \|-\theta x\circ s + \theta^2 d_x^a\circ d_s^a\|_\infty \qquad \text{(using (2.1))}\\
&\le \theta\|x\circ s\|_\infty + \theta^2\|D^{-1}d_x^a\|_2\|Dd_s^a\|_2\\
&\le \theta n\mu + \theta^2 n\mu \qquad \text{(using (3.1) and (3.2))}\\
&\le \frac{16}{15}\theta n\mu \qquad\qquad \text{(since } \theta \le \frac{1}{15}\text{)}. \qquad \#
\end{aligned}$$

Now we will estimate how far the new point $(\bar{x}+\bar{\alpha}d_x^c, \bar{s}+\bar{\alpha}d_s^c)$ can move away from the boundary of the feasible region in the case $0 \le \theta \le O(\beta/n)$.

Lemma 3.2: *Let*

$$0 \le \theta \le \theta_* := \frac{\beta}{15n},$$

and let $(x,s) \in \mathcal{N}_\infty^-(\beta)$ *and*

$$(\bar{x},\bar{s}) := (x,s) + \theta(d_x^a, d_s^a) \in \mathcal{N}_\infty^-(\beta).$$

Assume that $0 < \beta \le 0.1$. *Then for* $\bar{\alpha}$ *defined at the corrector step of the algorithm, the following estimate holds.*

$$(\bar{x}+\bar{\alpha}d_x^c)\circ(\bar{s}+\bar{\alpha}d_s^c) \ge (\beta + \frac{\beta}{20n})\bar{\mu}e, \tag{3.6}$$

where $\bar{\mu} = \frac{\bar{x}^T\bar{s}}{n}$, *which implies* $(\bar{x}+\bar{\alpha}d_x^c, \bar{s}+\bar{\alpha}d_s^c) \in \mathcal{N}_\infty^-(\beta+\beta/(20n))$.

Proof: Because $\bar{\alpha}$ is defined as the maximum by (2.9), hence if the inequality (3.6) can be shown for some α, it must also be true for $\bar{\alpha}$. In what follows we will show that the inequality (3.6) holds for

$$\alpha = \frac{\beta}{10n}.$$

We write the left hand side of (3.6) with the above α as

$$(\bar{x}+\alpha d_x^c)\circ(\bar{s}+\alpha d_s^c) = \bar{x}\circ\bar{s} + \alpha(\bar{x}\circ d_s^c + \bar{s}\circ d_x^c) + \alpha^2 d_x^c\circ d_s^c.$$

The middle term should be further decomposed,

$$\begin{aligned}
&\bar{x} \circ d_s^c + \bar{s} \circ d_x^c \\
=&x \circ d_s^c + s \circ d_x^c + \theta d_x^a \circ d_s^c + \theta d_s^a \circ d_x^c \\
=&(x \circ s)^{1/2} \circ (\bar{x} \circ \bar{s})^{-1/2} \circ (\bar{\mu} e - \bar{x} \circ \bar{s}) + \theta d_x^a \circ d_s^c + \theta d_s^a \circ d_x^c \qquad \text{(using (2.8))} \\
=&\bar{\mu} e - \bar{x} \circ \bar{s} + \theta d_x^a \circ d_s^c + \theta d_s^a \circ d_x^c \\
&+ \left[(x \circ s)^{1/2} - (\bar{x} \circ \bar{s})^{1/2}\right] \circ \left[\bar{\mu}(\bar{x} \circ \bar{s})^{-1/2} - (\bar{x} \circ \bar{s})^{1/2}\right].
\end{aligned}$$

So we have

$$(\bar{x} + \alpha d_x^c) \circ (\bar{s} + \alpha d_s^c) = (1-\alpha)\bar{x} \circ \bar{s} + \alpha \bar{\mu} e + \alpha(w_1 + w_2), \tag{3.7}$$

where

$$w_1 := \alpha d_x^c \circ d_s^c + \theta d_x^a \circ d_s^c + \theta d_s^a \circ d_x^c$$

and

$$w_2 := \left[(x \circ s)^{1/2} - (\bar{x} \circ \bar{s})^{1/2}\right] \circ \left[\bar{\mu}(\bar{x} \circ \bar{s})^{-1/2} - (\bar{x} \circ \bar{s})^{1/2}\right].$$

Using (3.2), (3.3) and (3.4) and noting that $\theta \le 1/15$, we obtain

$$\|w_1\|_2 \le \frac{\alpha n \bar{\mu}}{\beta} + 2\theta n \sqrt{\frac{\mu \bar{\mu}}{\beta}} = \frac{\alpha n \bar{\mu}}{\beta} + \frac{2\theta n \bar{\mu}}{\sqrt{(1-\theta)\beta}} \le (\alpha + 2.1\theta)\frac{n\bar{\mu}}{\beta}. \tag{3.8}$$

For w_2 we have

$$\begin{aligned}
\|w_2\|_\infty &= \|(x \circ s - \bar{x} \circ \bar{s}) \circ \left[(x \circ s)^{1/2} + (\bar{x} \circ \bar{s})^{1/2}\right]^{-1} \circ \left[\bar{\mu}(\bar{x} \circ \bar{s})^{-1/2} - (\bar{x} \circ \bar{s})^{1/2}\right]\|_\infty \\
&\le \|(x \circ s - \bar{x} \circ \bar{s}) \circ \left[\bar{\mu}(\bar{x} \circ \bar{s})^{-1} - e\right]\|_\infty \\
&\le \frac{16\theta n \mu}{15\beta} \qquad \left(\text{ using (3.5) and } \frac{\bar{x}_i \bar{s}_i}{\bar{\mu}} \ge \beta\right) \\
&\le \frac{16\theta n \bar{\mu}}{14\beta} \qquad \left(\text{ using (3.4) and } \theta \le \frac{1}{15}\right).
\end{aligned} \tag{3.9}$$

Notice that

$$(1-\alpha)\bar{x} \circ \bar{s} + \alpha \bar{\mu} e \ge (1-\alpha)\beta \bar{\mu} e + \alpha \bar{\mu} e = \beta \bar{\mu} e + \alpha(1-\beta)\bar{\mu} e. \tag{3.10}$$

It follows from (3.7), (3.8), (3.9) and (3.10) that

$$\begin{aligned}
(\bar{x} + \alpha d_x^c) \circ (\bar{s} + \alpha d_s^c) &\ge \beta \bar{\mu} e + \alpha\left[1 - \beta - \frac{(\alpha + 2.1\theta)n}{\beta} - \frac{16\theta n}{14\beta}\right]\bar{\mu} e \\
&\ge \beta \bar{\mu} e + \frac{\beta}{10n}\left[1 - \beta - \frac{1}{10} - \frac{3.25}{15}\right]\bar{\mu} e \\
&\ge \beta \bar{\mu} e + \frac{\beta \bar{\mu}}{20n} e \qquad (\text{ assuming } \beta \le 0.1).
\end{aligned}$$

The proof of Lemma 3.2 is completed. #

3.2. Analysis of the predictor step

We only analyse the predictor steps at those points which are just generated at a corrector step and whose preceding predictor step has a length $\theta \leq \theta_*$. By Lemma 3.2, such a point (x, s) satisfies

$$x \circ s \geq (\beta + \frac{\beta}{20n})\mu e, \tag{3.11}$$

where $\mu := x^T s/n$. We emphasize that (x, s) used in this subsection is not the point (x, s) used in subsection 3.1, actually, it is the point $(\bar{x} + \bar{\alpha} d_x^c, \bar{s} + \bar{\alpha} d_s^c)$ there. However, it should not impose any confusion because in what follows we do not care what point (x, s) is, but only assume that (x, s) satisfies (3.11).

We seek to show that at point (x, s) satisfying (3.11) the predictor step length $\tilde{\theta}$ is greater than θ_*.

Lemma 3.3: *Suppose that (x, s) satisfies (3.11). Let (d_x^a, d_s^a) be defined by (2.4). Then the predictor step length at (x, s) is greater than $\theta_* := \beta/(15n)$.*

Proof: We need only to show that for any $0 \leq \theta \leq \theta_*$ the following inequality holds.

$$(x + \theta d_x^a) \circ (s + \theta d_s^a) > \beta \bar{\mu} e, \tag{3.12}$$

where

$$\bar{\mu} = \frac{(x + \theta d_x^a)^T (s + \theta d_s^a)}{n}.$$

Using (3.2) we have

$$\|d_x^a \circ d_s^a\|_2 \leq n\mu, \tag{3.13}$$

where $\mu = x^T s/n$. It follows from (2.1) that

$$\begin{aligned}
&(x + \theta d_x^a) \circ (s + \theta d_s^a) \\
&\qquad = x \circ s + \theta(x \circ d_s^a + s \circ d_x^a) + \theta^2 d_x^a \circ d_s^a \\
&\qquad = (1 - \theta) x \circ s + \theta^2 d_x^a \circ d_s^a \\
&\qquad \geq (1 - \theta)(1 + \frac{1}{20n})\beta \mu e - \theta^2 n \mu e \qquad (\text{ using (3.11) and (3.13) }) \\
&\qquad = \left(1 + \frac{1}{20n} - \frac{\theta^2 n}{(1 - \theta)\beta}\right)\beta \bar{\mu} e \qquad (\text{using (3.4)})
\end{aligned}$$

Since $\theta \leq \beta/(15n)$ and $\beta < 1$,

$$\frac{\theta^2 n}{(1 - \theta)\beta} < \frac{1}{20n}.$$

This proves (3.12) and the proof of the lemma is completed. #

3.3. The main theorem

Theorem 3.4: *Let $\theta_{min} \geq \beta/(15n)$ and $\beta \leq 0.1$. Then the total number of iterations needed by the algorithm is not more than $\frac{30n}{\beta} \ln \frac{\mu_0}{\epsilon}$, where $\mu_0 := x_0^T s_0/n$ at the starting point (x_0, s_0).*

Proof: Let θ_k be the predictor step length at k-th iteration. We divide iterations into two sets $\mathcal{K}_i^+$ and $\mathcal{K}_i^-$ according to $\theta_* := \beta/(15n)$ (but not $\theta_{\min}$),

$$\mathcal{K}_i^+ := \{k \leq i \mid \theta_k > \theta_*\}, \qquad \mathcal{K}_i^- := \{k \leq i \mid \theta_k \leq \theta_*\},$$

and denote by K_i^+ and K_i^- the size of $\mathcal{K}_i^+$ and $\mathcal{K}_i^-$ respectively. Then we have $K_i^+ + K_i^- = i + 1$. By Lemma 3.3, if $k \in \mathcal{K}_i^-$, then $k+1$ must belong to $\mathcal{K}_i^+$. This implies that $K_i^- \leq K_i^+$, hence

$$i \leq 2K_i^+. \tag{3.14}$$

Let (x_k, s_k) be generated at the predictor step of k-th iteration, and let $\mu_k := x_k^T s_k/n$. By (3.4) we have

$$\begin{aligned}
\mu_i &= \prod_{k=0}^{i-1} (1-\theta_k)\mu_0 \\
&\leq \prod_{k \in \mathcal{K}_i^+} (1-\theta_k)\mu_0 \\
&\leq (1-\theta_*)^{K_i^+} \mu_0.
\end{aligned}$$

Then $\mu_i \leq \epsilon$ will be achieved if

$$K_i^+ \geq \frac{\ln(\epsilon/\mu_0)}{\ln(1-\theta_*)} \approx \theta_*^{-1} \ln \frac{\mu_0}{\epsilon} = \frac{15n}{\beta} \ln \frac{\mu_0}{\epsilon},$$

here we use $\ln(1-\theta_*) \approx -\theta_*$ for small θ_*. The above inequality and (3.14) imply that μ_i will not be greater than ϵ if

$$i \geq \frac{30n}{\beta} \ln \frac{\mu_0}{\epsilon}.$$

Finally, we note that the number of corrector steps is not greater than the number of predictor steps, and each corrector step needs only $O(n^2)$ operations which is less than $O(n^3)$ needed by each predictor step. Hence the corrector steps do not affect the overall complexity. #

References

[1] Adler, I.; N. Karmarkar, M.G.C. Resende and G. Veiga: "An implementation of Karmarkar's algorithm for linear programming", *Mathematical Programming* 44 (1989) 297–335.

[2] Bayer, D.A. and J.C. lagarias: "The nonlinear geometry of linear programming, I, II, III", *Transaction of the American Mathematical Society*, 314 (1989) 499–526, 527–581, and 320 (1990) 193–225.

[3] Jansen, B.; C. Roos and T. Terlaky: "A polynomial primal-dual Dikin-type algorithm for linear programming", Report 93-36, Faculty of Technical mathematics and Informatics, TU Delft, The Netherlands, (1993).

[4] Jarre, Fl.; G. Sonnevend and J. Stoer: "An implementation of the method of analytic centers", in: *Lect. Notes in Control and Inf. Sci.*, Vol. 111, Springer, Berlin (1988) 297–307.

[5] Kojima, M.; N. Megiddo and S. Mizuno: "Theoretical convergence of large-step primal-dual interior point algorithms for linear programming", *Mathematical Programming* 59 (1993) 1–21.

[6] Megiddo, N. and M Shub: "Boundary behavior of interior point algorithms in linear programming", *Mathematics of Operations Research*, 14 (1989) 97–146.

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

[8] Mizuno, S.; M.J. Todd and Y. Ye: "On adaptive-step primal-dual interior-point algorithm for linear programming", Technical Report No. 944, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY (1990), to appear in *Mathematics of Operations Research.*

[9] Monteiro, R.D.C.; 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.

[10] Nesterov, Y. and A. Nemirovskii: *Interior Point Polynomial Algorithms in Convex Programming*, SIAM Publications, Philadelphia, PA (1994).

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

[12] Todd, M.: "Recent developments and new directions in linear programming", *Mathematical Programming — Recent Developments and Applications* (M.Iri and K. Tanabe, eds.), Kluwer Academic Publishers (1989) 109–157.

[13] Witzgall, C.; P.T. Boggs and P.D. Domich: "On the convergence behavior of trajectories for linear programming", In: Mathematical Developments Arising from Linear Programming Algorithms (J.C. Lagarias, M.J. Todd, eds.), *Contemporary Mathematics*, Vol. 114, American Mathematical Society, Providence, RI (1991) 161–187.

[14] Zhao, G. and J. Stoer: "Estimating the complexity of a class of path-following methods for solving linear programs by curvature integrals", *Appl. Math. & Optimization* 27 (1993) 85–103.