Mathematical Programming 72 (1996) 177–194

An O($\sqrt{n}L$) iteration bound primal–dual cone affine scaling algorithm for linear programming

Jos F. Sturm [a], Shuzhong Zhang [b,*]

[a] *Tinbergen Institute, Rotterdam, The Netherlands*
[b] *Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

Received 1 January 1993; revised manuscript received 11 September 1995

Abstract

In this paper we introduce a primal–dual affine scaling method. The method uses a search-direction obtained by minimizing the duality gap over a linearly transformed conic section. This direction neither coincides with known primal–dual affine scaling directions (Jansen et al., 1993; Monteiro et al., 1990), nor does it fit in the generic primal–dual method (Kojima et al., 1989). The new method requires O($\sqrt{n}L$) main iterations. It is shown that the iterates follow the primal–dual central path in a neighbourhood larger than the conventional $\mathcal{N}_2$ neighbourhood. The proximity to the primal–dual central path is measured by trigonometric functions.

Keywords: Linear programming; Primal–dual interior point algorithm; Affine scaling; Polynomiality

1. Introduction

The original interior point algorithm for linear programming introduced by Karmarkar [14] is called a *projective scaling* algorithm [9], since at each iteration it maps the current iterate to the center of a simplex by means of a projective scaling transformation. This transformation is nonlinear due to the existence of a nonhomogeneous constraint. Very promising numerical results were reported later by Adler et al. [1] for a simplification of Karmarkar's algorithm. This simplification dispenses with the nonlinear projective transformation, thus the mapping becomes merely an affine transformation. Hence it was referred to as the *affine scaling* algorithm. (To our best knowledge, the earliest reference using the term *affine scaling* is [9].) Surprisingly, it turned out that this algorithm was already proposed in 1967 by Dikin [3]. Unlike

* Corresponding author.

SSDI 0025-5610(95)00042-9

Karmarkar's algorithm, the affine scaling algorithm has not been proved to be polynomial. In implementing the algorithm, it involves an important parameter: the step length. By taking a fixed partial step along the Dikin direction, the ordinary affine scaling algorithm is obtained. The convergence proof of that algorithm without any nondegeneracy condition is given by Tsuchiya [26]. Taking a step along the Dikin direction with a fixed fraction towards the boundary (maximal step-length), the algorithm obtained is called the *large step* affine scaling algorithm. The convergence of the large step affine scaling algorithm has been established by Tsuchiya and Muramutsu [27].

By Karmarkar's projective scaling transformation, the original linear objective function becomes a fractional linear function. The search direction used in Karmarkar's method (in the transformed space) is obtained by optimizing only the numerator of the transformed fractional linear function (thus a simplification) over a sphere inscribed in the solution space. As a result, this search direction is in general not a descent direction for the original linear objective; it is descent only for the potential function [14]. In contrast to this strategy, Padberg [22] derived a search direction by optimizing the *whole fractional objective* over the sphere. Similar algorithms were independently proposed and analysed by Goldfarb and Xiao [5] and Gonzaga [6]. In particular, Goldfarb and Xiao [5] gave a convergence proof for the algorithm under primal and dual nondegeneracy assumptions. Additionally, Goldfarb and Xiao [5] proposed a variant of the algorithm that is provably polynomial.

The radial projection of a nonnegative vector w on the simplex $\{y \geqslant 0 \mid e^{\mathrm{T}} y = 1\}$, i.e., $(1/e^{\mathrm{T}} w)w$, belongs to the sphere

$$\left\{ y \mid \| y - \frac{1}{n} e \| \leqslant 1/\sqrt{n(n-1)} \right\}$$

if and only if w belongs to the inscribed circular cone

$$\left\{ y \mid \cos(e, y) \geqslant \sqrt{\frac{n-1}{n}} \right\}.$$

In fact, one may obtain the search direction derived in Padberg [22] and Goldfarb and Xiao [5] by optimizing the original linear objective over a circular cone, using merely an affine transformation. In this sense, the direction may be called a *cone affine scaling* direction. Under this interpretation, Jan and Fang [11] rediscovered this direction, and provided computational results suggesting an improvement over Dikin's affine scaling in practice.

Path following is another important class of interior point methods. The idea of following the central path, or the trajectory of centers, stems from Huard [10] and Fiacco and McCormick [4]. The polynomiality of a variant of Huard's method, the short step center method, was proved by Renegar [24] in 1988. Renegar's method is the first $O(\sqrt{n}\,L)$ iteration method for linear programming. This $O(\sqrt{n}\,L)$ iteration bound was later obtained for other variants of interior point methods, including the potential reduction method [25,30], the primal–dual path-following method [20] and the predictor–corrector method [19].

As recent intensive research indicates, the primal–dual interior point algorithms are theoretically and practically superior to the primal-only variants. Moreover, primal–dual methods can be easily understood as a scheme to solve the approximated K–K–T system using Newton's method (cf. [8,15,16,19]). Monteiro et al. [21] introduced a primal–dual version of the affine scaling method. The interpretation of this method is to apply Newton's method to solve the K–K–T system aiming directly at the optimal point. In [21] it is proved that the method has an iteration bound of $O(nL^2)$ if very short steps are taken. The length of the steps can also be determined by incorporating a potential function, see [18,28]. Recently, Jansen et al. [12] introduced a different version of the primal–dual affine scaling method, in which the search direction is obtained by minimizing the duality gap under the feasibility restrictions with an additional sphere constraint (after scaling). Clearly, this direction-finding procedure resembles Dikin's affine scaling method for the primal-only case. Moreover, Jansen et al. [12] proved that the method requires at most $O(nL)$ iterations, which is a considerable improvement over Monteiro et al.'s $O(nL^2)$ bound.

In this paper, we introduce a new variant of primal-dual affine scaling method. The main feature of our method is that we obtain our search direction by minimizing the duality gap under the feasibility restrictions and a certain conic constraint. Following our discussion on the cone affine scaling by Padberg [22], Goldfard and Xiao [5] and Jan and Fang [11], the new method may be called the primal–dual cone affine scaling – an analog in the primal–dual case. As we will show later, its iterates follow the central path, and this enables us to prove an $O(\sqrt{n}\,L)$ iteration bound for the new method.

Before discussing our method, we outline the notation and the organization of this paper.

We consider the standard linear programming problem

$$(P) \qquad \min\{c^{\mathrm{T}}x \mid Ax=b,\ x\geqslant 0\},$$

where A is an $m\times n$ matrix ($n\geqslant 2$), b and c are m- and n-dimensional vectors respectively, and $x\in\mathbb{R}^n_+$ is the decision variable, and the dual problem

$$(D) \qquad \max\{b^{\mathrm{T}}y \mid A^{\mathrm{T}}y+s=c,\ s\geqslant 0\},$$

where $y\in\mathbb{R}^m$, $s\in\mathbb{R}^n$ are the decision variables.

In this paper we assume without loss of generality that the rank of A is m, and that there exist solutions $x>0$ and y, s with $s>0$ such that $Ax=b$ and $A^{\mathrm{T}}y+s=c$.

For a given vector denoted by w the corresponding upper case letter W denotes the diagonal matrix $\mathrm{diag}(w)$ defined by that vector w. We denote the L_p norm of w by $\| w\|_p$ and the Euclidean norm (the L_2-norm) simply by $\|w\|$. We will use trigonometric functions in defining the search direction. For two vectors f and w, define

$$\cos(f,w) := \frac{f^{\mathrm{T}}w}{\|f\|\,\|w\|}, \qquad \sin(f,w) := \sqrt{1-\cos(f,w)^2},$$

$$\tan(f,w) := \frac{\sin(f,w)}{\cos(f,w)}.$$

Subscripts i will denote the ith component of a vector, e.g., w_i, except for the all-one vector $e = [1, 1, \ldots, 1]^{\mathrm{T}}$ where e_i will denote the ith unity vector. We let I be the identity matrix $I = [e_1, e_2, \ldots, e_n]$. We will use superscript i to refer to quantities in the ith iteration of the algorithm, e.g., w^i.

The organization of this paper is as follows. In Section 2, we derive the primal–dual cone affine scaling direction. Based on this direction, a primal–dual cone affine scaling algorithm is introduced in Section 3. We show in Section 4 that this algorithm guarantees a reduction in duality gap of $1 - 1/\mathrm{O}(\sqrt{n})$ per step. Clearly this implies an $\mathrm{O}(\sqrt{n}\, L)$ iteration bound for the algorithm. In Section 5, we compare the primal–dual cone affine scaling direction with other primal–dual search directions, and we investigate how our neighbourhood of the primal–dual central path relates to a more commonly used neighbourhood.

2. The primal–dual cone affine scaling

2.1. The primal–dual interior-point method

The nonnegative primal variables x in (P) and the dual slacks s in (D), are restricted to orthogonal affine spaces, viz.

$$x \in A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1} b + \mathrm{Ker}(A), \qquad s \in c + \mathrm{Range}(A^{\mathrm{T}}).$$

Since the dimensions of these two orthogonal affine spaces sum up to n, they have a unique intersection point. Following Kojima et al. [16], per iteration the solution space is transformed linearly such that the current iterate coincides with this unique intersection point of the transformed space. Let the current solution be x and s. The transformation is:

$$\bar{x}_i \to \sqrt{\frac{s_i}{x_i}}\, \bar{x}_i \quad \text{for } i = 1, 2, \ldots, n, \qquad \bar{s}_i \to \sqrt{\frac{x_i}{s_i}}\, \bar{s}_i, \quad \text{for } i = 1, 2, \ldots, n.$$

Introducing the scaling vector d,

$$d_i := \sqrt{\frac{x_i}{s_i}}\,, \quad \text{for } i = 1, 2, \ldots, n,$$

and the primal–dual vector v,

$$v_i := \sqrt{x_i s_i}\,, \quad \text{for } i = 1, 2, \ldots, n,$$

it follows that

$$v = D^{-1} x = Ds.$$

Displacements in the original space will be denoted by Δx and Δs, and in the transformed space by $p_x := D^{-1}\Delta x$ and $p_s := D\Delta s$, respectively. Moreover, $p := p_x + p_s$.

The system

$$D^{-1}\Delta x + D\Delta s = p, \qquad A\Delta x = 0, \qquad A^{\mathrm{T}}\Delta y + \Delta s = 0$$

describes the one-to-one correspondence between a search direction p in the v-space and a pair $(\Delta x, \Delta s)$. This can be seen by the relations

$$p_x := P_{AD}\,p \quad \text{and} \quad p_s := p - p_x,$$

where P_{AD} denotes the orthogonal projection matrix on Ker(AD) given as

$$P_{AD} := I - DA^{\mathrm{T}}(AD^2A^{\mathrm{T}})^{-1}AD.$$

Different primal–dual interior point algorithms use different p in defining search directions. (See Section 5 for a discussion on search directions.)

2.2. *An inscribed circular cone*

In the introduction, it was discussed that the sphere constraint $\| x - e/n \| \leqslant 1/\sqrt{n(n-1)}$ used in Karmarkar's projective scaling algorithm corresponds to the circular cone constraint

$$\cos(e, x) \geqslant \sqrt{\frac{n-1}{n}}, \tag{1}$$

dispensing with projection. We will show that this condition is stronger than the inequality constraints $x \geqslant 0$, but weaker than Dikin's sphere constraint

$$\| x - e \| \leqslant 1. \tag{2}$$

We first mention the following lemma.

Lemma 2.1. *Let* $w \in \mathbb{R}^n$. *Then,*

$$\max_{1 \leqslant i \leqslant n} \left| w_i - \frac{e^{\mathrm{T}}w}{n} \right| \leqslant \sqrt{n-1}\,\tan(e, w)\frac{e^{\mathrm{T}}w}{n}.$$

Proof.

We will prove the lemma by showing

$$\left(e_i - \frac{e}{n}\right)\left(e_i - \frac{e}{n}\right)^{\mathrm{T}} \preceq (n-1)\left(\frac{1}{n}I - \frac{ee^{\mathrm{T}}}{n^2}\right) \tag{3}$$

for arbitrary $i \in \{1, 2, \ldots, n\}$.

The only nonzero eigenvalue of the rank-one matrix

$$\left(e_i - \frac{e}{n}\right)\left(e_i - \frac{e}{n}\right)^{\mathrm{T}}$$

is $(n-1)/n$ with corresponding eigenvector $(e_i - e/n)$. The positive semi-definite matrix

$$(n-1)\left(\frac{1}{n}I - \frac{ee^{\mathrm{T}}}{n^2}\right)$$

has also an eigenvalue $(n-1)/n$ corresponding to the eigenvector $(e_i - e/n)$. This proves (3). By pre-multiplying by w^{T} and post-multiplying by w in (3), it follows that

$$\begin{aligned}\max_{1\leqslant i\leqslant n}\left|w_i-\frac{e^{\mathrm{T}}w}{n}\right| &\leqslant \sqrt{n-1}\sqrt{\frac{\|w\|^2}{n}-\left(\frac{e^{\mathrm{T}}w}{n}\right)^2}\\ &=\sqrt{n-1}\sqrt{\frac{1}{\cos^2(e,w)}-1}\,\frac{|e^{\mathrm{T}}w|}{n}\\ &=\sqrt{n-1}\tan(e,w)\frac{e^{\mathrm{T}}w}{n},\end{aligned}$$

completing the proof. □

Lemma 2.2. *Let $w\in\mathbb{R}^n$. If $\cos(e,w)\geqslant\sqrt{(n-1)/n}$, then $w\geqslant 0$.*

Proof.
If $\cos(e,w)\geqslant\sqrt{(n-1)/n}$, then $e^{\mathrm{T}}w\geqslant 0$, $\sin(e,w)\leqslant\sqrt{1/n}$ and $\tan(e,w) \leqslant\sqrt{1/(n-1)}$. Using Lemma 2.1, we obtain

$$\max_{1\leqslant i\leqslant n}\left|w_i-\frac{e^{\mathrm{T}}w}{n}\right|\leqslant\frac{e^{\mathrm{T}}w}{n}$$

and so consequently $w\geqslant 0$. □

Actually we can further show that (1) implies either $x>0$ or $x=e^{\mathrm{T}}x(e-e_i)/(n-1)$ for some $i\in\{1,2,\ldots,n\}$.

What remains to be shown is that the cone constraint (1) is weaker than the sphere constraint (2). Clearly, if $\|x-e\|\leqslant 1$, then

$$\|x\|^2-2e^{\mathrm{T}}x+n\leqslant 1,$$

which implies

$$\cos(e,x)=\frac{e^{\mathrm{T}}x}{\sqrt{n}\,\|x\|}\geqslant\frac{\|x\|+(n-1)/\|x\|}{2\sqrt{n}}.$$

After noting that the minimum of the above right-hand side is obtained when $\|x\| =\sqrt{n-1}$, it follows that the sphere constraint is at least as strong as the cone constraint. Certainly, $x=10e$ satisfies the cone constraint but not the sphere constraint. This proves that the cone constraint is indeed weaker than the sphere constraint.

2.3. Derivation of the new direction

Similar to the way the affine scaling search direction is derived, we now require that $v+p_x\geqslant 0$ and $v+p_s\geqslant 0$. Notice that the largest inscribed circular cone in primal–dual scaled space is described by

$$\cos\left(\begin{bmatrix}e\\e\end{bmatrix},\begin{bmatrix}v+p_x\\v+p_s\end{bmatrix}\right)\geqslant\sqrt{\frac{2n-1}{2n}}.$$

According to Lemma 2.2, satisfaction of the above condition is sufficient for the nonnegativity of $v+p_x$ and $v+p_s$.

Since $e^{\mathrm{T}}(v+p_x+v+p_s)=e^{\mathrm{T}}(2v+p)$ and

$$\left\|\begin{bmatrix} v+p_x \\ v+p_s \end{bmatrix}\right\|^2 = \|v+p_x\|^2+\|v+p_s\|^2$$

$$=\|(I-P_{AD})v+P_{AD}(v+p)\|^2+\|P_{AD}v+(I-P_{AD})(v+p)\|^2$$

$$=\|v\|^2+\|v+p\|^2,$$

it follows that

$$\cos\left(\begin{bmatrix} e \\ e \end{bmatrix}, \begin{bmatrix} v+p_x \\ v+p_s \end{bmatrix}\right) = \cos\left(\begin{bmatrix} e \\ e \end{bmatrix}, \begin{bmatrix} v \\ v+p \end{bmatrix}\right).$$

Further notice that

$$(x+\Delta x)^{\mathrm{T}}(s+\Delta s)=\|v\|^2+v^{\mathrm{T}}p,$$

which shows that taking a full step along that direction will result in a reduction in duality gap of $v^{\mathrm{T}}p$.

The preceding discussion motivates to define the new primal–dual cone affine scaling direction p as the solution of the following program,

$$(\overline{\mathrm{PD}})\quad \text{minimize} \quad v^{\mathrm{T}}p$$
$$\text{subject to} \quad \cos\left(\begin{bmatrix} e \\ e \end{bmatrix}, \begin{bmatrix} v \\ v+p \end{bmatrix}\right) \geqslant \sqrt{\frac{2n-1}{2n}},$$

where p is the decision variable.

Notice that if the angle between v and e is large, $(\overline{\mathrm{PD}})$ will not have any feasible solution at all. In this respect, it is useful to recall the concept of the *primal–dual central path*.

Definition 2.1. (*Primal–dual central path*). A pair (x, s) lies on the *primal–dual central path* if and only if $v=(e^{\mathrm{T}}v/n)e$.

We introduce

$$\delta := \sqrt{n-1}\tan(e, v)$$

as a measure of proximity to the primal–dual central path, cf. Lemma 2.1. Based on this measure, a new neighbourhood of the central path is defined as

$$\mathscr{N}(\beta) := \{v \in \mathbb{R}^n_+ \mid \delta \leqslant \beta\}$$

for some $\beta \in (0, 1)$. In Section 5, it will be shown that the standard neighbourhood $\mathscr{N}_2(\beta)$ as used in, e.g., [19] is tighter than $\mathscr{N}(\beta)$.

The next theorem shows that the problem $(\overline{\mathrm{PD}})$ has an analytical solution.

Theorem 2.1. *If $v \in \mathcal{N}(\beta)$ for some $\beta \in (0, 1)$ and $v \neq 0$, then $(\overline{\mathrm{PD}})$ has an optimal solution p given by*

$$p := -\frac{\xi+1}{\xi}v + \frac{\xi-1}{\xi}\frac{e^{\mathrm{T}}v}{n-1}e,$$

where $\xi = \sqrt{2n/(1-\delta^2)-1}$.

Proof.

Rewrite $(\overline{\mathrm{PD}})$ as

$$\min_{p}\left\{v^{\mathrm{T}}p \mid \sqrt{2n-1}\left\|\begin{bmatrix} v \\ v+p \end{bmatrix}\right\| - e^{\mathrm{T}}(2v+p) \leqslant 0\right\},$$

for which the Lagrangian is

$$\mathscr{L}_{\lambda}(p) = v^{\mathrm{T}}p + \lambda\left(\sqrt{2n-1}\sqrt{\|v\|^2 + \|v+p\|^2} - e^{\mathrm{T}}(2v+p)\right).$$

The gradient is

$$\nabla\mathscr{L}_{\lambda}(p) = v + \eta(v+p) - \lambda e,$$

where we let

$$\eta := \frac{\lambda\sqrt{2n-1}}{\sqrt{\|v\|^2 + \|v+p\|^2}}.$$

The Karush–Kuhn–Tucker optimality conditions are

$$\nabla\mathscr{L}_{\lambda}(p) = 0, \tag{4}$$

$$\lambda\left(\sqrt{2n-1}\sqrt{\|v\|^2 + \|v+p\|^2} - e^{\mathrm{T}}(2v+p)\right) = 0, \tag{5}$$

$$\lambda \geqslant 0 \quad \text{and} \quad \sqrt{2n-1}\sqrt{\|v\|^2 + \|v+p\|^2} - e^{\mathrm{T}}(2v+p) \leqslant 0. \tag{6}$$

If $\lambda = 0$, then (4) implies $v = 0$ which contradicts the assumptions of the theorem. Therefore, $\lambda > 0$ and (4) yields

$$p = -\frac{\eta+1}{\eta}v + \frac{\lambda}{\eta}e. \tag{7}$$

Since $\lambda > 0$, using (5) and (7) and the definition of η, we obtain

$$\begin{aligned} 0 &= \sqrt{2n-1}\sqrt{\|v\|^2 + \|v+p\|^2} - e^{\mathrm{T}}(2v+p) \\ &= (2n-1)\frac{\lambda}{\eta} - e^{\mathrm{T}}\left(2v - \frac{\eta+1}{\eta}v + \frac{\lambda}{\eta}e\right) \\ &= \frac{(n-1)\lambda - (\eta-1)e^{\mathrm{T}}v}{\eta}, \end{aligned}$$

which implies that

$$\lambda = (\eta-1)\frac{e^{\mathrm{T}}v}{n-1}. \tag{8}$$

What remains is to solve for η. By definition of η we have

$$\lambda^2(2n-1) = \eta^2\left(\| v \|^2 + \| v+p \|^2\right) = \eta^2 \| v \|^2 + \| v - \lambda e \|^2,$$

where we used (7). Rewriting yields

$$\begin{aligned} 0 &= \eta^2 \| v \|^2 + \| v \|^2 - 2\lambda e^{\mathrm{T}} v + n\lambda^2 - (2n-1)\lambda^2 \\ &= (\eta^2+1)\| v \|^2 - \lambda(2e^{\mathrm{T}} v + (n-1)\lambda) \\ &= (\eta^2+1)\| v \|^2 - \lambda(\eta+1)e^{\mathrm{T}} v, \end{aligned}$$

where we used (8) in the last equality. Applying (8) once more we obtain

$$(\eta^2+1)\| v \|^2 = (\eta^2-1)\frac{(e^{\mathrm{T}} v)^2}{n-1}. \tag{9}$$

Because η is nonnegative, we obtain from (9) that

$$\eta = \sqrt{\frac{(n-1)\| v \|^2 + (e^{\mathrm{T}} v)^2}{(e^{\mathrm{T}} v)^2 - (n-1)\| v \|^2}} = \sqrt{\frac{2(e^{\mathrm{T}} v)^2}{(e^{\mathrm{T}} v)^2 - (n-1)\| v \|^2} - 1}\ . \tag{10}$$

Using

$$\tan^2(e,v) = \frac{1}{\cos^2(e,v)} - 1 = \frac{n\| v \|^2}{(e^{\mathrm{T}} v)^2} - 1,$$

it follows from (10) that

$$\eta = \sqrt{\frac{2n}{1-(n-1)\tan^2(e,v)} - 1} = \sqrt{\frac{2n}{1-\delta^2} - 1} = \xi.$$

Together with (7) and (8) this completes the proof. □

The search directions Δx and Δs in the original primal and dual spaces are as follows:

$$\Delta x := Dp_x = DP_{AD}p = -\frac{\xi+1}{\xi}DP_{AD}v + \frac{\xi-1}{\xi}\frac{e^{\mathrm{T}} v}{n-1}DP_{AD}e \tag{11}$$

and

$$\Delta s := D^{-1}(p-p_x) = \frac{\xi+1}{\xi}D^{-1}(I-P_{AD})v - \frac{\xi-1}{\xi}\frac{e^{\mathrm{T}} v}{n-1}D^{-1}(I-P_{AD})e. \tag{12}$$

For notational reason we introduce a new vector:

$$q := \frac{\| v \|^2}{e^{\mathrm{T}} v}e - v.$$

Clearly, q is orthogonal to v and therefore,

$$\delta = \sqrt{n-1}\tan(e,v) = \sqrt{n-1}\,\frac{\| q \|}{\| v \|}.$$

In addition, we obtain the relation

$$\frac{\| v \|^2}{e^{\mathrm{T}} v} e^{\mathrm{T}} q = \| q \|^2 = \frac{\delta^2}{n-1} \| v \|^2. \tag{13}$$

Because of (9) and $\eta = \xi$ we have

$$(\xi^2 + 1) \| v \|^2 = (\xi^2 - 1) \frac{(e^{\mathrm{T}} v)^2}{n-1}.$$

Hence, letting $r := (\xi^2 + 1)/(2\xi)$,

$$\begin{aligned} p &= -\frac{\xi+1}{\xi} v + \frac{\xi-1}{\xi} \frac{e^{\mathrm{T}} v}{n-1} e \\ &= -\frac{\xi+1}{\xi} v + \frac{2}{\xi+1} \frac{\xi^2-1}{2\xi} \frac{(e^{\mathrm{T}} v)^2}{(n-1)\| v \|^2} \frac{\| v \|^2}{e^{\mathrm{T}} v} e \\ &= -\frac{\xi+1}{\xi} v + \frac{2r}{\xi+1} (q + v). \end{aligned}$$

In order to simplify this expression, we rewrite

$$\frac{\xi+1}{\xi} - \frac{2r}{\xi+1} = \frac{(\xi+1)^2 - (\xi^2+1)}{\xi(\xi+1)} = \frac{2}{\xi+1},$$

to obtain

$$p = \frac{2}{\xi+1} (-v + rq).$$

By the orthogonality of v and q, this shows immediately that

$$\| p \| = \frac{2}{\xi+1} \sqrt{1 + r^2 \delta^2/(n-1)} \, \| v \|. \tag{14}$$

and

$$v^{\mathrm{T}} p = -\frac{2}{\xi+1} \| v \|^2,$$

so that for any t,

$$(v + tp_x)^{\mathrm{T}} (v + tp_s) = \| v \|^2 + t v^{\mathrm{T}} p = \left(1 - \frac{2t}{\xi+1} \right) \| v \|^2. \tag{15}$$

3. The algorithm

Now we are ready to present the cone affine scaling algorithm. The iterative solutions in the algorithm are generated by steps along the primal–dual cone affine scaling direction p obtained in the previous section. In order for p to be well defined, the step length is to be controlled such that all iterates are contained in $\mathcal{N}(\beta)$ for some fixed $\beta \in (0, 1)$. Without loss of generality initial primal and dual solutions in $\mathcal{N}(\beta)$ are assumed to be known.

Algorithm 1. (A, b, c, x^0, s^0)

Input: Initial feasible solution (x^0, s^0) such that $v^0 = (X^0 S^0)^{1/2} e \in \mathcal{N}(\beta)$.

Output: Feasible solution (x^i, s^i) such that $(x^i)^{\mathrm{T}} s^i < 2^{-2L}$

and $v^i = (X^i S^i)^{1/2} e \in \mathcal{N}(\beta)$.

Step 0 (Initialization). Set $i = 0$.

Step 1 (Check stopping criterion). If $(x^i)^{\mathrm{T}} s^i < 2^{-2L}$, then stop

Step 2 (Compute search directions). Compute Δx^i and Δs^i according to Eqs. (11) and (12).

Step 3 (Compute step length). Compute the largest t such that

$$(X^i + \bar{t}\Delta X^i)^{1/2} (S^i + \bar{t}\Delta S^i)^{1/2} e \in \mathcal{N}(\beta) \quad \text{for all } 0 \leqslant \bar{t} \leqslant t.$$

Step 4 (Take steps). Set $x^{i+1} := x^i + t\Delta x^i$ and $s^{i+1} := s^i + t\Delta s^i$.

Step 5 (Set $i := i + 1$ and return to Step 1).

It is well known that when the duality gap is at most 2^{-2L}, the corresponding primal and dual solutions can be purified to optimal solutions of (P) and (D) in $\mathrm{O}(n^3)$ operations [23].

4. Convergence analysis

In this section we will derive the $\mathrm{O}(\sqrt{n}\, L)$ iteration bound for Algorithm 1. To this end, we will first derive a lower bound on the step length.

4.1. A lower bound on the step length

Denoting the current and next iterate by v and v^+ respectively, we have for any $i \in \{1, 2, \dots, n\}$ and feasible step length t that

$$\begin{aligned} v_i^+ &= \sqrt{(v_i + t(p_x)_i)(v_i + t(p_s)_i)} \\ &= (v_i + tp_i/2)\sqrt{1 + \frac{(p_x)_i(p_s)_i - \frac{1}{4}p_i^2}{(v_i + tp_i/2)^2}\, t^2}, \end{aligned}$$

where the nonnegativity of $v + \frac{1}{2}tp$ follows from $v + tp_x \geqslant 0$ and $v + tp_s \geqslant 0$.

Remark here that for any $\rho \geqslant -1$,

$$\sqrt{1+\rho} - 1 = \frac{\rho}{1 + \sqrt{1+\rho}} \geqslant -|\rho|,$$

so that

$$v_i^+ \geqslant v_i + \tfrac{1}{2}tp_i - \frac{|(p_x)_i(p_s)_i - \frac{1}{4}p_i^2|}{v_i + \frac{1}{2}tp_i}t^2. \tag{16}$$

In order to estimate $\cos(e, v^+)$, we will derive a lower bound on $e^{\mathrm{T}}v^+$. Using $p = 2(-v + rq)/(\xi + 1)$, we obtain

$$e^{\mathrm{T}}\left(v + \tfrac{1}{2}tp\right) = \left(1 - \frac{t}{\xi + 1}\right)e^{\mathrm{T}}v + \frac{rt}{\xi + 1}e^{\mathrm{T}}q.$$

Together with (13) this implies

$$\frac{e^{\mathrm{T}}\left(v + \frac{1}{2}tp\right)}{e^{\mathrm{T}}v} = 1 - \frac{t}{\xi + 1} + \frac{r}{\xi + 1}\frac{\delta^2}{n - 1}t. \tag{17}$$

Again using $p = 2(-v + rq)/(\xi + 1)$, we rewrite

$$\begin{aligned} v + \tfrac{1}{2}tp &= \left(1 - \frac{t}{\xi + 1}\right)v + \frac{rt}{\xi + 1}q \\ &= \left(1 - \frac{(r + 1)t}{\xi + 1}\right)v + \frac{rt}{\xi + 1}\frac{1}{\cos^2(e, v)}\frac{e^{\mathrm{T}}v}{n}e. \end{aligned} \tag{18}$$

Remark here that by definition of r and ξ, there holds $r = (\xi^2 + 1)/2\xi < \xi$ so that

$$1 - \frac{r + 1}{\xi + 1}t > 0 \text{ for } 0 \leqslant t \leqslant 1.$$

From Lemma 2.1 we have

$$v \geqslant (1 - \delta)\frac{e^{\mathrm{T}}v}{n}e,$$

which implies for $0 \leqslant t \leqslant 1$ that

$$\left(1 - \frac{(r + 1)t}{\xi + 1}\right)v \geqslant \left(1 - \frac{t}{\xi + 1}\right)(1 - \delta)\frac{e^{\mathrm{T}}v}{n}e - \frac{rt}{\xi + 1}(1 - \delta)\frac{e^{\mathrm{T}}v}{n}e.$$

Combination with (18) yields

$$v + \tfrac{1}{2}tp \geqslant \left(1 - \frac{t}{\xi + 1}\right)(1 - \delta)\frac{e^{\mathrm{T}}v}{n}e \geqslant \frac{\xi}{\xi + 1}(1 - \delta)\frac{e^{\mathrm{T}}v}{n}e \tag{19}$$

for $0 \leqslant t \leqslant 1$.

From the Cauchy–Schwartz inequality, it follows that

$$\sum_{i=1}^{n}|(p_x)_i(p_s)_i| \leqslant \|p_x\|\,\|p_s\| = \|p_x\|\sqrt{\|p\|^2 - \|p_x\|^2} \leqslant \tfrac{1}{2}\|p\|^2 \tag{20}$$

where, from (14),

$$\|p\|^2 = 4\frac{1 + r^2\delta^2/(n - 1)}{(\xi + 1)^2}\|v\|^2. \tag{21}$$

Summarizing (16), (17), (19)–(21), we have

$$\frac{e^{\mathrm{T}}(v^{+})}{e^{\mathrm{T}}v} \geqslant 1 - \frac{t}{\xi+1} + \frac{r}{\xi+1}\frac{\delta^2}{n-1}t - 3\frac{\xi+1}{(1-\delta)\xi}\frac{1+r^2\delta^2/(n-1)}{(\xi+1)^2}\frac{n\|v\|^2}{(e^{\mathrm{T}}v)^2}t^2.$$

Using (15), we estimate

$$\|v^{+}\| = \sqrt{1-2t/(\xi+1)}\,\|v\| \leqslant \left(1 - \frac{t}{\xi+1}\right)\|v\|, \tag{22}$$

from which we obtain

$$\cos(e, v^{+}) \geqslant \cos(e, v) + \frac{r}{\xi+1}\frac{\delta^2}{n-1}\frac{e^{\mathrm{T}}v}{\sqrt{n}\,\|v^{+}\|}t - 3\frac{n-1+r^2\delta^2}{(1-\delta)(n-1)\xi(\xi+1)}\left(1+\frac{\delta^2}{n-1}\right)\frac{e^{\mathrm{T}}v}{\sqrt{n}\,\|v^{+}\|}t^2. \tag{23}$$

Now we are in a position to prove the following result.

Lemma 4.1. *After the first main iteration of Algorithm* 1, *the step length is at least*

$$t \geqslant 0.1\beta^2(1-\beta).$$

Proof. Notice that after the initial iteration, there always holds $\delta = \beta$. Consider the case that

$$0 \leqslant t < 0.1\beta^2(1-\beta).$$

We know that

$$\xi^2 + 1 = \frac{2n}{1-\delta^2},$$

so, $\xi^2 > 3$, as $n \geqslant 2$. Combining this with the definition of r, we obtain

$$\frac{n-1+r^2\beta^2}{\xi^2+1} \leqslant \frac{n-1}{\xi^2+1} + \frac{r^2}{\xi^2+1} \leqslant (1-\delta^2)\frac{n-1}{2n} + \frac{\xi^2+1}{4\xi^2} \leqslant \frac{1}{2} + \frac{4}{12} = \frac{5}{6}.$$

By definition of r,

$$\frac{\xi^2+1}{r\xi} = 2,$$

and using $n \geqslant 2$,

$$\left(1 + \frac{\beta^2}{n-1}\right) \leqslant 2,$$

so that we can further estimate (23) as follows:

$$\begin{aligned}\cos(e, v^+) &\geqslant \cos(e, v) + \frac{r}{\xi+1}\frac{\beta^2}{n-1}\frac{e^{\mathrm{T}}v}{\sqrt{n}\,\|v^+\|}t \\ &\quad - 3\frac{2rt}{(1-\beta)(\xi+1)(n-1)}\frac{5}{6}\left(1+\frac{\beta^2}{n-1}\right)\frac{e^{\mathrm{T}}v}{\sqrt{n}\,\|v^+\|}t \\ &\geqslant \cos(e, v) \\ &\quad + \frac{r}{\xi+1}\frac{\beta^2}{n-1}\frac{e^{\mathrm{T}}v}{\sqrt{n}\,\|v^+\|}t\left(1-\frac{10t}{\beta^2(1-\beta)}\right) \\ &> \cos(e, v).\end{aligned}$$

Therefore, we conclude that the length of the step towards the boundary of $\mathcal{N}(\beta)$ is at least $0.1\beta^2(1-\beta)$. □

4.2. *The convergence result*

Combining Lemma 4.1 with (15), it follows that only $O(\sqrt{n}\,L)$ main iterations are needed by Algorithm 1 to solve (P) and (D) simultaneously. This result is included in the following theorem.

Theorem 4.1. *Choose a parameter β independent of n and L, e.g., $\beta = \frac{1}{2}$. Suppose x^0 and s^0 are feasible interior solutions to* (P) *and* (D) *respectively with* $\log((x^0)^{\mathrm{T}}s^0) = O(L)$ *and $\delta^{(0)} \leqslant \beta$. Then, Algorithm* 1 *yields a pair of primal and dual feasible solutions (x, s) with $x^{\mathrm{T}}s < 2^{-2L}$ in at most $O(\sqrt{n}\,L)$ main iterations.*

Proof. By definition of the step length in Algorithm 1, we have $\delta^{(i)} \leqslant \beta$ for all i, so that

$$\xi^i = \sqrt{\frac{2n}{1-(\delta^i)^2} - 1} = O(\sqrt{n}).$$

By Lemma 4.1 we know that for any $i \geqslant 1$ there holds $t \geqslant 0.1\ \beta^2(1-\beta)$. From (15), we thus have for $i \geqslant 1$,

$$(x^{i+1})^{\mathrm{T}}s^{i+1} = \left(1 - \frac{2t}{\xi^{(i)}+1}\right)(x^i)^{\mathrm{T}}s^i = \left(1 - \frac{1}{O(\sqrt{n})}\right)(x^i)^{\mathrm{T}}s^i.$$

The theorem follows immediately from the above inequality. □

5. Comparisons

It was observed by Yamashita [29], Gonzaga [7] and Den Hertog and Roos [2] that many search directions used in primal interior point methods can be expressed somehow as a linear combination of two directions given by $-XP_{AX}Xc$ and $XP_{AX}e$, where P_{AX} denotes the orthogonal projection matrix on Ker(AX). These two directions are called the primal affine scaling direction and the primal centering direction respectively. Similarly, dual interior point methods use linear combinations of the vectors $-S(I-P_{AS^{-1}})Sb$ and $S(I-P_{AS^{-1}})e$, which are known as the dual affine scaling direction and the dual centering direction respectively. The more recent primal–dual methods, however, appear to be more versatile in their choice of the search direction. First, the primal–dual path following method uses a linear combination of

$$\left(DP_{AD}v,\ D^{-1}(I-P_{AD})v\right)$$

and

$$\left(DP_{AD}V^{-1}e,\ D^{-1}(I-P_{AD})V^{-1}e\right)$$

for the primal–dual search direction $(\Delta x, \Delta s)$.

The primal–dual affine scaling direction of Monteiro et al. [21] is simply

$$(\Delta x, \Delta s)=\left(DP_{AD}v, D^{-1}(I-P_{AD})v\right),$$

whereas Jansen et al. [12] proposed

$$(\Delta x, \Delta s)=\left(DP_{AD}V^{2}v,\ D^{-1}(I-P_{AD})V^{2}v\right) \tag{24}$$

as their Dikin-type search direction.

Finally, the primal–dual cone affine scaling direction introduced in this paper is a linear combination of

$$\left(DP_{AD}v,\ D^{-1}(I-P_{AD})v\right) \quad\text{and}\quad \left(DP_{AD}e,\ D^{-1}(I-P_{AD})e\right),$$

i.e., it is a combination of the primal–dual affine scaling direction of Monteiro et al. [21] and a new centering direction. Hence it is different from both the primal–dual path-following direction and the primal–dual Dikin-type affine scaling direction (24). On the primal–dual central path, however, v is a multiple of e so that all primal–dual directions coincide with the direction $(XP_{AX}e, S(I-P_{AX})e)$.

Another interesting issue in polynomial interior point methods is the neighbourhood of the central path used in obtaining a step length. Many $O(\sqrt{n}L)$ iteration methods use the $\mathcal{N}_2(\beta)$ neighbourhood where

$$\mathcal{N}_2(\beta):=\left\{(x, s)\mid \| Xs-\mu e\| \leqslant \beta\mu,\ \mu=\frac{x^{\mathrm{T}}s}{n}\right\},$$

see, e.g., [19] and [8].

As $e\perp Xs-\mu e$, we have

$$\| Xs-\mu e\|/\mu=\sqrt{n}\tan(e, Xs),$$

so that the $\mathcal{N}_2(\beta)$ neighbourhood can be written in terms of the primal–dual vector v as

$$\mathcal{N}_2(\beta)=\left\{v\mid \sqrt{n}\tan(e, V^{2}e)\leqslant\beta\right\}.$$

We will use the following lemma to show that $\mathcal{N}_2(\sqrt{1+1/(n-1)}\,\beta)\subset\mathcal{N}(\beta)$.

Lemma 5.1. *Let $x \in \mathbb{R}^n_+$. Assume that x is positive and non-multiple of e. Then,*

$$\| x \|_1 \| x \|_4^2 > \| x \|_2^3.$$

Proof. It is well known that the function

$$f(t) := \log\left(\sum_{i=1}^n x_i^t \right)$$

is a strictly convex function in t because

$$f\left(\tfrac{1}{2}\{t_1 + t_2\}\right) < \tfrac{1}{2}\{f(t_1) + f(t_2)\}$$

for any $t_1 \neq t_2$ due to the Cauchy–Schwartz inequality. Therefore,

$$\tfrac{2}{3}f(1) + \tfrac{1}{3}f(4) > f(2).$$

Taking the natural exponential we obtain

$$\left(\sum_{i=1}^n x_i \right)^{2/3} \left(\sum_{i=1}^n x_i^4 \right)^{1/3} > \sum_{i=1}^n x_i^2,$$

completing the proof. □

From the above lemma, it follows for $v \in \mathbb{R}^n_+$ and v non-multiple of e that

$$\cos(e, v) > \cos(e, V^2 e),$$

and therefore

$$\mathscr{N}_2(\beta) \subset \mathscr{N}_2\left(\sqrt{1 + \frac{1}{n-1}}\, \beta \right) \subset \mathscr{N}(\beta),$$

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

Notice that the above containing relations are all strict. This shows that the new neighbourhood is indeed wider than the standard $\mathscr{N}_2$ neighbourhood. The iterates of Algorithm 1, which are always on the boundary of $\mathscr{N}(\beta)$, are therefore outside the standard $\mathscr{N}_2(\beta)$ neighbourhood.

Acknowledgements

Two anonymous referees are kindly acknowledged. Their comments greatly helped to improve the paper. We are grateful to Dr. Jan Brinkhuis for providing us an elegant proof for Lemma 5.1.

References

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

[2] D. Den Hertog and C. Roos, "A survey of search directions in interior point methods for linear programming," *Mathematical Programming* 52 (1991) 481–509.

[3] I.I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Soviet Mathematics Doklady* 8 (1967) 674–675.

[4] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (Wiley, New York, 1968).

[5] D. Goldfarb and D. Xiao, "A primal projective interior point method for linear programming," *Mathematical Programming* 51 (1991) 17–43.

[6] C.C. Gonzaga, "Conical projection algorithms for linear programming," *Mathematical Programming* 43 (1989) 151–173.

[7] C.C. Gonzaga, "Search directions for interior linear programming methods," *Algorithmica* 6 (1991) 153–181.

[8] C.C. Gonzaga, "The largest step path following algorithm for monotone linear complementarity problems," Report 94-07, Delft University of Technology, (1994).

[9] J.N. Hooker, "Karmarkar's linear programming algorithm," *Interfaces* 16 (1986) 75–90.

[10] P. Huard, "Resolution of mathematical programming with nonlinear constraints by the method of centres," in: J. Abadie, ed., *Nonlinear Programming* (North Holland, Amsterdam, 1967) pp. 207–219.

[11] G-M. Jan and S-C Fang, "A new variant of the primal affine scaling algorithm for linear programs", *Optimization* 22 (1991) 681–715.

[12] B. Jansen, C. Roos and T. Terlaky, "A polynomial primal–dual Dikin-type algorithm for linear programming," Report 93-36, Delft University of Technology, (1993).

[13] B. Jansen, C. Roos, T. Terlaky and J.-Ph. Vial, "Interior-point methodology for linear programming: Duality, sensitivity analysis and computational aspects," Report 93-28, Delft University of Technology, (1993).

[14] N.K. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica* 4 (1984) 373–395.

[15] 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, Vol. 538 (Springer, Berlin, 1991).

[16] 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, New York, 1989) pp. 29–48.

[17] N. Megiddo, "On the complexity of linear programming," in: T. Bewley, ed., *Advances in Economic Theory* (Cambridge University Press, Cambridge, 1987) pp. 225–268.

[18] S. Mizuno and A. Nagasawa, "A primal–dual affine scaling potential reduction algorithm for linear programming," *Mathematical Programming* 62 (1993) 119–131.

[19] S. Mizuno, M.J. Todd and Y. Ye, "On adaptive-step primal–dual interior-point algorithms for linear programming," *Mathematics of Operations Research* (1993) 964–981.

[20] R.D.C. Monteiro and I. Adler, "Interior path following primal–dual algorithm," *Mathematical Programming* 44 (1989) 27–42.

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

[22] M.W. Padberg, "Solution of a nonlinear programming problem arising in the projective method for linear programming," Manuscript, School of Business and Administration, New York University (New York, 1985).

[23] C.H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization, Algorithms and Complexity* (Prentice-Hall, Englewood Cliffs, NJ, 1982).

[24] J. Renegar, "A polynomial-time algorithm, based on Newton's method, for linear programming," *Mathematical Programming* 40 (1988) 59–93.

[25] J.F. Sturm and S. Zhang, "New complexity results for the Iri–Imai method," Research Memorandum 530, University of Groningen (Groningen, 1993). To appear in *Annals of Operations Research*.

[26] T. Tsuchiya, "Global convergence of the affine scaling methods for degenerate linear programming problems," *Mathematical Programming* 52 (1991) 377–404.

[27] T. Tsuchiya and M. Muramatsu, "Global convergence of the long-step affine scaling algorithm for degenerate linear programming problems," *SIAM Journal on Optimization* 5 (1995) 525–551.

[28] L. Tunçel, "Constant potential primal-dual algorithm: A framework," *Mathematical Programming* 66 (1994) 145–159.
[29] H. Yamashita, *A Polynomially and Quadratically Convergent Method for Linear Programming* (Mathematical Systems Inc., Tokyo, 1986).
[30] Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programming," *Mathematical Programming* 50 (1991) 239–258.