Properties of the central points in linear programming problems

Clovis C. Gonzaga* Marli Cardia†

Abstract

In this paper we list several useful properties of central points in linear programming problems. We study the logarithmic barrier function, the analytic center and the central path, relating the proximity measures and scaled Euclidean distances defined for the primal and primal-dual problems. We study the Newton centering steps, and show how large the short steps used in path following algorithms can actually be, and what variation can be ensured for the barrier function in each iteration of such methods. We relate the primal and primal-dual Newton centering steps and propose a primal-only path following algorithm for linear programming.

1 Introduction

The central path associated with linear programs and linear complementarity problems has been described in many papers, and its properties are exploited by all interior point algorithms. Many of these properties are dispersed in several references, and ad hoc proofs are needed whenever they are used. In this paper we collect several properties of central points, analytic centers, proximity measures and Newton steps, providing proofs for most of them.

Much is known about the asymptotic behavior of the central path, both in primal and in primal-dual settings. First order derivatives of the central path were studied by Adler and Monteiro [1], and are described in [16], Ch. 16. High order derivatives are systematically derived in Güler [10]. The asymptotic behavior of the derivatives is described in these references, and Güler studies both ends of the central path.

Most studies in the last decade were concentrated on primal-dual methods, due to their superior performance in solving linear programming problems. Nowadays there is a renewed interest in primal-only methods, due to the extensions of the approach to nonlinear problems.

*Department of Mathematics, Federal University of Santa Catarina, Florianópolis, Brazil.
Research partially sponsored by CAPES-Brazil.
†Department of Mathematics, Federal University of Paraná, Curitiba, Brazil

It is known that Iri and Imai's [11, 19] multiplicative barrier function method has quadratic convergence properties. The affine-scaling algorithm can be made superlinear with careful choices of steplengths [18, 17], and Karmarkar's algorithm becomes quadratically convergent if each iteration makes a bidirectional search [7]. But we are unaware of any primal-only predictor-corrector algorithm. In our last section we shall address this question, propose a new primal predictor-corrector algorithm and discuss its expected behavior.

In some occasions one is interested in the behavior of the central path near the analytic center of a bounded polyhedron: this is the case when studying cutting plane methods based on centers. Each iteration of such methods starts from a point near the analytic center of a polytope Ω and adds a new constraint, say $a^T x \leq b$, which cuts the previously known point. Such methods were introduced by Goffin and Vial [3], and are surveyed in [4]. The complexity of these methods when used for solving convex feasibility problems is studied in Nesterov [15] and in Goffin, Luo and Ye [2], and depends on the variation of the logarithmic barrier function between successive centers.

The analytic center of the new polytope will belong to the central path of a linear programming problem in which the new constraint plays the role of cost, and the feasible set is Ω. A clear understanding of this problem and the development of path following algorithms for introducing the new cut depends on a clear understanding of the behavior of the central path near the analytic center, and also of the variations of the barrier function along the path: this is the main scope of this paper. The use of primal-dual path following algorithms for introducing deep cuts will be described in [9].

Lower bounds to the steps taken by primal or primal-dual path following algorithms are well known (see [5]), as well as the asymptotic behavior of primal-dual methods. The relationship between primal and primal-dual steps is not so well understood, and there seems to be no predictor-corrector primal algorithm. We try to shed some light on these questions.

Our tasks are:

- Define primal and primal-dual proximity measures and relate them to the scaled Euclidean distances.
- Show that the central path begins at the analytic center θ of the feasible set, quickly moves away from it, and then stays forever out of a ball (in scaled norm) centered at θ.
- Show that the variation of the primal barrier function along the central path depends on the proximity to the analytic center.
- Study the short step path following algorithm and show how large the short steps can be: the possible parameter variation depends on the proximity to the analytic center.
- Show that the barrier function variation in each iteration of the short step algorithm has a lower bound which does not depend on the problem.

- Establish a relationship between primal and primal-dual Newton steps in path following methods, and propose a primal predictor-corrector algorithm.

Our notation is quite standard: $I\!R^n$ is the Euclidean space, $I\!R^n_+$ is the non negative orthant and $I\!R^n_{++}$ is the positive orthant. The letter e denotes the vector of ones, with dimension given by the context. Given a vector x, the upper case X denotes the matrix $\text{diag}(x_1, x_2, \ldots, x_n)$. Given vectors $x, s \in I\!R^n$ x^{-1}, xs, $\frac{x}{s}$ denote the vectors with components x_i^{-1}, $x_i s_i$ and $\frac{x_i}{s_i}$, $i = 1, ..., n$. $\|\cdot\|$ denotes the Euclidean norm, and $x^T s$ is the inner product of the vectors x and s. The null space and the range space of a matrix A are respectively $\mathcal{N}(A)$ and $\mathcal{R}(A)$. The projection matrix into $\mathcal{N}(A)$ is denoted P_A, and if A has full row rank then it can be computed by $P_A = I - A^T(AA^T)^{-1}A$. The orthogonal complement of P_A is $\tilde{P}_A = I - P_A$.

2 Definitions and known properties

In this section we present definitions and some well known properties, most of which can be found in reference [6]. We shall study the primal and dual linear programming problems and their optimality conditions:

$$
\begin{array}{lllllllllll}
 & (P) & & & & & (D) & & & & (PD) \\
\min & c^T x & & & \max & b^T y & & & xs & = & 0 \\
\text{s. t.} & Ax & = & b & \text{s. t.} & A^T y + s & = & c & Ax & = & b \\
 & x & \geq & 0 & & s & \geq & 0 & A^T y + s & = & c \\
 & & & & & & & & x, s & \geq & 0
\end{array}
$$

where $c \in I\!R^n$, $b \in I\!R^m$ and $A \in I\!R^{m \times n}$ $(m < n)$ has full row rank.

The primal feasible set and the set of primal interior points are given by

$$\mathcal{X} = \{x \in I\!R^n; Ax = b, x \geq 0\} \quad , \quad \overset{\circ}{\mathcal{X}} = \{x \in \mathcal{X}; x > 0\}.$$

The dual feasible set and the set of dual interior points are given by

$$\Omega = \{(y, s) \in I\!R^{n+m}; A^T y + s = c, \ s \geq 0\} \quad , \quad \overset{\circ}{\Omega} = \{(y, s) \in \Omega; \ s > 0\}.$$

We shall assume that $\overset{\circ}{\mathcal{X}} \neq \emptyset$ and $\overset{\circ}{\Omega} \neq \emptyset$. This is the main hypothesis for interior point methods, and it is equivalent to the hypothesis that the primal-dual optimal face is bounded. It is known that the primal and dual feasible sets cannot be simultaneously bounded (and this will be clear in a moment, when we define central points). We shall assume that the primal feasible set is bounded.

The dual problem can be stated in terms of either y or s alone, because A has full row rank. In terms of y, (D) is

$$\underset{y \in \Omega_y}{\text{maximize}}\, b^T y, \qquad \Omega_y = \{y \in I\!R^m; \ A^T y \leq c\}.$$

To state (D) in terms of s, choose an arbitrary $z \in \mathcal{X}$ and let B be a matrix whose rows span $\mathcal{N}(A)$, for instance $B = P_A$, the projection matrix into $\mathcal{N}(A)$. The problem

$$
(D') \qquad \begin{array}{lrl} \text{minimize} & z^t s & \\ \text{subject to} & Bs & = Bc \\ & s & \geq 0 \end{array}
$$

is equivalent to (D) in the sense that its feasible set is $\Omega_s = \{s;\ (y,s) \in \Omega\}$, and its optimal set coincides with the set of optimal slacks for (D). The symmetry between (P) and (D') allows us to study many properties of the primal problem and then adapt them for the dual problem. We shall do this in this paper: we study the primal and primal-dual problems, and in the last section rephrase some properties for the dual problem.

Central points. The logarithmic barrier function is defined by

$$
x \in I\!R^n_{++} \mapsto p(x) = -\sum_{i=1}^{n} \ln x_i.
$$

The central paths associated with (P) and (D) are respectively

$$
\alpha > 0 \mapsto x(\alpha) = \operatorname*{argmin}_{x \in \overset{\circ}{\mathcal{X}}} \alpha c^T x + p(x).
$$

$$
\alpha > 0 \mapsto (y(\alpha), s(\alpha)) = \operatorname*{argmax}_{(y,s) \in \overset{\circ}{\Omega}} \alpha b^T y - p(s).
$$

Only $\mathcal{X}$ is bounded, by hypothesis. Then the primal central path starts at $\alpha = 0$, and the starting point is the analytic center of $\mathcal{X}$, denoted

$$
\theta = x(0) = \operatorname*{argmin}_{x \in \overset{\circ}{\mathcal{X}}} p(x).
$$

The analytic center and the central points are well defined, since $p(\cdot)$ is strictly convex and unbounded near the boundary of $\mathcal{X}$.

The following lemma proved for instance in [6] relates the primal and dual central paths to the so called primal-dual central path:

Lemma 2.1 *Given $\alpha > 0, (x(\alpha), s(\alpha))$ are the unique solutions of the relaxed primal-dual equations $(PD)_\alpha$ given by*

$$
\alpha x s = e, \quad x \in \mathcal{X},\ s \in \Omega_s.
$$

From this lemma we can see that for α approaching zero and $x(\alpha)$ approaching the analytic center, $s(\alpha)$ will be unbounded.

Now we introduce some basic concepts and some notation.

The following function will be useful when studying proximity measures:

$$
\lambda \in [0, \sqrt{2}/2] \mapsto R(\lambda) = \sqrt{1 - \sqrt{1 - 2\lambda^2}} \in [\lambda, 1]. \tag{1}
$$

This function is monotone increasing in its domain, and stisfies the following properties: $R(0.1) \leq 0.1004$; $R(0.4) \leq 0.42$; if $\lambda \leq 0.4$, then $R(\lambda) \leq 1.05\,\lambda$.

Scaled norms. Given a point $z \in I\!R^n_{++}$, we define the following *scaled* norm associated with z:

$$x \in I\!R^n \mapsto \|x\|_z = \|x/z\|.$$

The scaled norm of a vector x corresponds to its Euclidean norm after a scaling which transfers z to the vector $e = [1\,1\ldots 1]^T$. The corresponding l_∞ norm will be denoted by $\|x\|_z^\infty$.

This norm is studied in Gonzaga [5], and the following lemma is a slight extension of a property proved there.

Lemma 2.2 *Given $a, b \in I\!R^n_{++}$ and $\delta \in (0,1)$, if $\|a-b\|_a^\infty \le \delta$ then for $h \in I\!R^n$,*

i) $\dfrac{1}{(1+\delta)}\|h\|_a \le \|h\|_b \le \dfrac{1}{1-\delta}\|h\|_a$

ii) $(1-\delta)\|h\|_b \le \|h\|_a \le (1+\delta)\|h\|_b$.

Proof. Assume that $\|a-b\|_a^\infty \le \delta$. Then for $i = 1,\ldots,n$, $1-\delta \le b_i/a_i \le 1+\delta$. Hence

$$\frac{1}{1+\delta} \le \frac{a_i}{b_i} \le \frac{1}{1-\delta}.$$

We can write $\|h\|_b = \left\| \left[\dfrac{h_i}{a_i} \dfrac{a_i}{b_i} \right] \right\|$, and it follows that

$$\|h\|_a \min_i\{a_i/b_i\} \le \|h\|_b \le \|h\|_a \max_i\{a_i/b_i\},$$

which implies (i). Also, $\|h\|_a = \left\| \left[\dfrac{h_i}{b_i} \dfrac{b_i}{a_i} \right] \right\|$, and by a similar construction we deduce (ii), completing the proof. $\Box$

Shifted scalings. Now we study the effect of a change of scale on the projection of a vector.

Lemma 2.3 *Consider a vector $c \in I\!R^n$ and a matrix $A \in I\!R^{m\times n}$. Let $x, z \in I\!R^n_{++}$ be such that $\|z-x\|_x^\infty \le \beta \in (0,1)$. Then*

$$1-\beta \le \frac{\|P_{AZ}Z\,c\|}{\|P_{AX}X c\|} \le 1+\beta.$$

Proof. By definition of projection, there exist $w_1, w_2 \in I\!R^m$ such that

$$\begin{array}{rcl} Zc &=& P_{AZ}Zc + ZA^T w_1 \\ Xc &=& P_{AX}Xc + XA^T w_2. \end{array}$$

Multiplying the second equation by ZX^{-1} and subtracting,

$$P_{AZ}Zc = ZX^{-1}P_{AX}Xc + ZA^T w,$$

where $w \in \mathbb{R}^m$. Projecting into $\mathcal{N}(AZ)$,

$$P_{AZ}Zc = P_{AZ}ZX^{-1}P_{AX}Xc,$$

and since orthogonal projections are contractions,

$$\|P_{AZ}Zc\| \leq \left\|ZX^{-1}P_{AX}Xc\right\| \leq \max_i\{z_i/x_i\}\,\|P_{AX}Xc\|\,.$$

By hypothesis, $\max_i\{|z_i/x_i - 1|\} = \beta$, which implies $\max_i\{z_i/x_i\} \leq 1+\beta$, proving the right inequality. From this same equality we also deduce that $\min_i\{z_i/x_i\} \geq 1-\beta$, which implies $\max_i\{x_i/z_i\} \leq 1/(1-\beta)$, which we shall use in a moment.

The other side is proved similarly: multiply the first equation above by XZ^{-1}, subtract and project into $\mathcal{N}(AX)$ to obtain

$$P_{AX}Xc = P_{AX}XZ^{-1}P_{AZ}Zc,$$

and taking norms as before,

$$\|P_{AX}Xc\| \leq \left\|XZ^{-1}P_{AZ}Zc\right\| \leq \max_i\{x_i/z_i\}\,\|P_{AZ}Zc\| \leq \|P_{AZ}Zc\|\,/(1-\beta),$$

completing the proof. □

3 Proximity measures

The primal penalized function for a given parameter value $\alpha \geq 0$ is defined as

$$x \in \mathbb{R}^n_{++} \mapsto f_\alpha(x) = \alpha c^t x + p(x).$$

The Newton centering step. Let $x \in \overset{\circ}{\mathcal{X}}$ be a given point, and consider a penalty parameter value $\alpha \geq 0$. The Newton step from x for minimizing $f_\alpha(\cdot)$ in $\overset{\circ}{\mathcal{X}}$ is well known, and studied in detail in [6]. It is computed as follows:

$$h(x,\alpha) = X\bar{h}(x,\alpha), \tag{2}$$

where $\bar{h}(x,\alpha)$ is the scaled centering step, which corresponds to the Newton centering step from $\bar{x} = e$ after a change of coordinates (scaling) which maps x to e. It is given by

$$\bar{h}(x,\alpha) = -P_{AX}X\nabla f_\alpha(x) = -P_{AX}(\alpha Xc - e). \tag{3}$$

When dealing with proximity measures, the following notation will be helpful. Consider a point $x \in \overset{\circ}{\mathcal{X}}$ (and also the special case $x = \theta$). The size of the projected cost after a scaling which transfers x to e is denoted

$$\begin{array}{rcl} r_x &=& \|P_{AX}Xc\| \\ r &=& \|P_{A\Theta}\Theta c\|\,. \end{array} \tag{4}$$

Proximity. Let $\alpha \geq 0$, $x \in \overset{\circ}{\mathcal{X}}$ and $s \in \overset{\circ}{\Omega}_s$ be given. Proximity measures are estimates of the distance from x to $x(\alpha)$, or from (x,s) to $(x(\alpha), s(\alpha))$. The following measures will be relevant:
The primal-dual proximity:

$$\delta(x,s,\alpha) = \|\alpha x s - e\|.$$

The primal proximity:

$$\delta(x,\alpha) = \left\|\bar{h}(x,\alpha)\right\|.$$

The primal proximity to the analytic center:

$$\delta(x) = \delta(x,0) = \left\|\bar{h}(x,0)\right\| = \|P_{AX} e\|.$$

The primal scaled distance:

$$\mathrm{dist}(x,\alpha) = \|x - x(\alpha)\|_{x(\alpha)} \equiv \left\|\frac{x}{x(\alpha)} - e\right\|.$$

The notation for the proximities differs in the number of arguments, which should not cause confusion.

The primal scaled distance cannot be computed without the knowledge of the central point. It plays an obviously important role, and was first related to the primal proximity by Megiddo and Shub [13]. The tightest relationship between these quantities is given by the following lemma, proved in Roos, Terlaky and Vial [16]:

Lemma 3.1 *Consider $\alpha \geq 0$, the central point $x(\alpha)$ and a point $x \in \mathcal{X}$. If $\delta(x,\alpha) \leq \sqrt{2}/2$, then*

$$\delta(x,\alpha) \leq \mathrm{dist}(x,\alpha) \leq R(\delta(x,\alpha)).$$

Note that if $\delta(x,\alpha)$ is small, then it is an excellent approximation to the scaled distance. If $\delta(x,\alpha) \leq 0.4$, then the error is smaller than 5%.

The primal and primal-dual proximities are related by the following fact: for $\alpha \geq 0$ and $x \in \overset{\circ}{\mathcal{X}}$,

$$\delta(x,\alpha) = \min\{\delta(x,s,\alpha);\ A^T y + s = c, y \in I\!R^m\}. \tag{5}$$

For given $\alpha \geq 0$ and $x \in \overset{\circ}{\mathcal{X}}$, the dual slack for which this minimum is assumed is given by the following important theorem:

Theorem 3.2 *Let $\alpha \geq 0$ and $x \in \overset{\circ}{\mathcal{X}}$ be such that $\delta(x,\alpha) < 1$. Then*

$$s = \frac{x^{-1}}{\alpha}(e - \bar{h}(x,\alpha)) \in \overset{\circ}{\Omega}_s$$

and $\delta(x,\alpha) = \delta(x,s,\alpha)$. Besides this,

$$\delta(x + h(x,\alpha), \alpha) \leq \delta(x + h(x,\alpha), s, \alpha) \leq (\delta(x,\alpha))^2.$$

This theorem is proved in [6]. The last statement of the theorem is possibly the most important fact in interior point methods, ensuring the quadratic convergence of the Newton centering algorithm, and proved in many references.

The following two lemmas describe the proximities related to the analytic center.

Lemma 3.3 *The proximity from the analytic center θ to the central point $x(\alpha)$ for $\alpha > 0$ is given by*

$$\delta(\theta, \alpha) = r\alpha.$$

The proximity from a point $x \in \overset{\circ}{\mathcal{X}}$ to the analytic center θ is given by

$$\delta(x) \equiv \delta(x, 0) = \|P_{AX}e\| = \min\{\|xs - e\|\,;\ A^T y + s = 0, y \in I\!R^m\}.$$

Proof. The first statement and the first equality of the second follow directly from (3) because $P_{A\Theta}e = 0$. Let us prove the last equality. From the definition of projection,

$$\begin{aligned} \|P_{AX}e\| &= \min\{\|e - \bar{s}\|\,;\ \bar{s} \in \mathcal{R}(XA^T)\} \\ &= \min\{\|e - \bar{s}\|\,;\ \bar{s} = XA^T y, y \in I\!R^m\}. \end{aligned}$$

Setting $s = X^{-1}\bar{s}$, the result follows by substitution, completing the proof. $\Box$

Lemma 3.4 *Consider the analytic center θ of $\mathcal{X}$ and a parameter $\alpha > 0$ such that $r\alpha \in [0, \sqrt{2}/2]$ and $\delta(x(\alpha)) \in [0, \sqrt{2}/2]$. Then*

$$\delta(\theta, \alpha) = r\alpha \leq \operatorname{dist}(\theta, \alpha) \leq R(r\alpha), \tag{6}$$

$$\frac{r\alpha}{1 + r\alpha} \leq \operatorname{dist}(x(\alpha), 0) \leq \frac{R(r\alpha)}{1 - R(r\alpha)}. \tag{7}$$

Proof. (6) follows immediately from Lemmas 2.2 and 3.4. Let us prove (7). By definition, $\operatorname{dist}(x(\alpha), 0) = \|x(\alpha) - \theta\|_\theta$. Let us use 0 2.2(i): the right inequality of (7) follows immediately because $\|x(\alpha) - \theta\|_{x(\alpha)} \leq R(r\alpha)$ by (6); with $\delta = \|\theta - x(\alpha)\|_{x(\alpha)}$, we obtain

$$\frac{\|\theta - x(\alpha)\|_{x(\alpha)}}{1 + \|\theta - x(\alpha)\|_{x(\alpha)}} \leq \|x(\alpha) - \theta\|_\theta .$$

By (6), $\|\theta - x(\alpha)\|_{x(\alpha)} = \operatorname{dist}(\theta, \alpha) \geq r\alpha$. The map $y \in [0, 1] \mapsto y/(1 + y)$ increases monotonically, proving (7) and completing the proof. $\Box$

The next lemma shows that near the analytic center, the scaled projected cost does not change much.

Lemma 3.5 *Let $x \in \overset{\circ}{\mathcal{X}}$ be such that $\delta(x) \in (0, \sqrt{2}/2)$. Then for r and r_x defined in (4),*

$$(1 - R(\delta(x)))r \leq r_x \leq (1 + R(\delta(x)))r.$$

Proof. By (5), $\|x-\theta\|_\theta = \mathrm{dist}(x, 0) \le R(\delta(x))$. Let $\beta = R(\delta(x))$: from Lemma 2.3,

$$1-\beta \le \frac{P_{AX}Xc}{P_{A\theta}\theta c} \le 1+\beta,$$

which coincides with the expression in the lemma and completes the proof. □

Now we shall state two lemmas showing geometrical features of the central path as it leaves the analytic center. The picture becomes very clear if we assume that the analytic center is $\theta = e$, the cost vector belongs to $\mathcal{N}(A)$ and has unit norm. Then Figure 1 represents the results: the central path very quickly reaches the point $x(\alpha)$ for $\alpha = 0.4$ (Lemma 3.6 will explain why). Then, for all $\alpha \ge 0.4$, all nearly central points (points x such that $\delta(x, \alpha) \le 0.1$) stay out of a ball centered in e with radius 0.23.

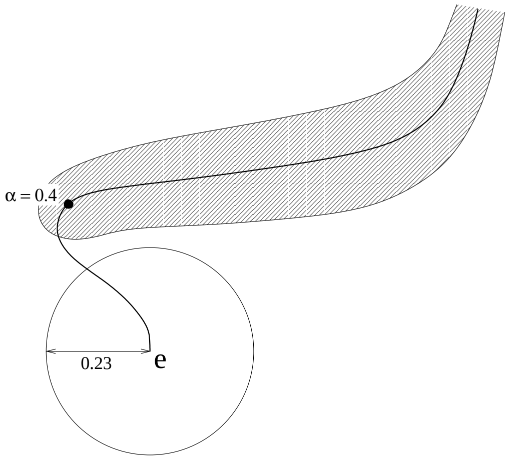

Figure 1: Nearly central points for $\alpha \ge 0.4$.

For $\theta \ne e$, the same is true using the scaled norm $\|\cdot\|_\theta$, and instead of $\alpha = 0.4$, the first piece of the central path uses $r\alpha = 0.4$.

First we show why $r\alpha \le 0.4$ implies that $x(\alpha)$ is near the analytic center: we show that such point can be reached from a point near θ in two Newton steps.

Lemma 3.6 *Let $\alpha > 0$ and $x \in \overset{\circ}{\mathcal{X}}$ be given, with $\delta(x) \le 0.1$ and $r\alpha = 0.4$. Then a point $x^+ \in \overset{\circ}{\mathcal{X}}$ such that $\delta(x^+, \alpha) \le 0.1$ can be obtained from x in two primal Newton iterations.*

Proof. Assume that $\delta(x) \le 0.1$ and $r\alpha = 0.4$. By definition,

$$\delta(x, \alpha) = \|P_{AX}(\alpha Xc - e)\| \le \alpha \|P_{AX}Xc\| + \|P_{AX}e\| = \alpha r_x + \delta(x).$$

By Lemma 3.5, $r_x \le (1 + R(0.1))r$. With $R(0.1) \le 0.1004$ and $r\alpha \le 0.4$, we obtain $r_x\alpha \le 0.45$ and hence $\delta(x, \alpha) \le 0.55$. Using Theorem 3.2 twice, in two Newton iterations we obtain x^+ such that $\delta(x^+, \alpha) \le 0.55^4 < 0.1$, completing the proof. □

So, in the analysis of primal path following algorithms we can always assume that an initial point x^0 is available near $x(\alpha)$ for $r\alpha = 0.4$, as an alternative to assuming that x^0 is near the analytic center. The next lemma completes our geometrical features.

Lemma 3.7 *Consider α such that $r\alpha \geq 0.4$ and $x \in \overset{\circ}{\mathcal{X}}$ such that $\delta(x,\alpha) \leq 0.1$. Then*

i) $\mathrm{dist}(x(\alpha), 0) \geq 0.28$.

ii) $\mathrm{dist}(x, 0) \geq 0.21$.

iii) $0.2 \leq \delta(x) \leq 0.1004 + 1.1004\,\delta(x(\alpha)) \leq 1.5\,\delta(x(\alpha))$.

Proof. Assume that $r\alpha \geq 0.4$ and $\delta(x,\alpha) \leq 0.1$.
i) From Lemma 3.4, $\mathrm{dist}(x(\alpha), 0) \geq r\alpha/(1 + r\alpha) \geq 0.4/1.4 \geq 0.28$.
ii) From Lemma 3.3 with $r\alpha \geq 0.4$, $\delta(\theta, \alpha) = \|\theta - x(\alpha)\|_{x(\alpha)} \geq 0.4$.
From Lemma 3.1, $\|x - x(\alpha)\|_{x(\alpha)} \leq R(0.1) \leq 0.1004$.
By the triangle inequality,

$$\|x - \theta\|_{x(\alpha)} \geq \|x(\alpha) - \theta\|_{x(\alpha)} - \|x - x(\alpha)\|_{x(\alpha)} \geq r\alpha - 0.1004 \geq 0.2996.$$

From Lemma 2.2 and $\|x - x(\alpha)\|_{x(\alpha)} \leq 0.1$ (with $a = x(\alpha)$ and $b = x$),

$$\|\theta - x\|_x \geq \frac{1}{1.1}\|\theta - x\|_{x(\alpha)} \geq \frac{0.2996}{1.1}.$$

Assume by contradiction that $\mathrm{dist}(x, 0) = \|x - \theta\|_\theta < 0.21$.
Again by Lemma 2.2 (with $a = \theta$ and $b = x$),

$$0.21 \geq \|x - \theta\|_\theta \geq \|x - \theta\|_x (1 - 0.21).$$

Merging with the inequality above,

$$0.21 \geq \frac{0.2996}{1.1}(1 - 0.21) = 0.2152,$$

a contradiction. This proves item (ii).
The following fact will be useful below in the proof: repeating the same computation for $x = x(\alpha)$ instead of $\delta(x,\alpha) \leq 0.1$, the bound is higher:

$$\mathrm{dist}(x(\alpha), 0) \geq 0.285.$$

iii) From (ii) and the properties of the function $R(\cdot)$, Lemma 3.1 implies $\delta(x) \geq 0.21/1.05 = 0.2$, proving the left inequality. By the same reasoning, for $x = x(\alpha)$ we get

$$\delta(x(\alpha)) \geq 0.285/1.05 > 0.27. \tag{8}$$

We must prove the right inequality. Assume without loss of generality that $x(\alpha) = e$. Then by Lemma 3.1, $\mathrm{dist}(x, \alpha) = \|x - e\| \leq R(0.1)$. Let $\bar{\delta} = R(0.1) \leq 0.1004$. We have:

$$\delta(x(\alpha)) = \delta(e) = \|P_A e\| \quad , \quad \delta(x) = \|P_{AX} e\|.$$

By definition of projection, there exist $w, \bar{w} \in I\!R^m$ such that

$$e = P_A e + A^T w \quad , \quad e = P_{AX} e + X A^T \bar{w}.$$

Multiplying the first equation by X and subtracting,

$$x - e = X P_A e - P_{AX} e + X A^T (w - \bar{w}).$$

This expression corresponds to the definition of projection of $x - e$:

$$X P_A e - P_{AX} e = P_{AX}(x - e)$$

and since orthogonal projections are contractions,

$$\|X P_A e - P_{AX} e\| \leq \|x - e\| \leq \bar{\delta},$$

and hence

$$\|P_{AX} e\| \leq \|X P_A e\| + \bar{\delta} \leq \|x\|_\infty \|P_A e\| + \bar{\delta}.$$

But $\|x\|_\infty \leq 1 + \bar{\delta}$, because $\|x - e\| \leq \bar{\delta}$, and finally

$$\delta(x) \leq (1 + \bar{\delta})\delta(x(\alpha)) + \bar{\delta} \leq 1.1004\, \delta(x(\alpha)) + 0.1004,$$

proving the second inequality.
From (8), $\delta(x(\alpha)) > 0.27$, which implies $\bar{\delta} \leq \bar{\delta}\, \delta(x(\alpha))/0.27$ and hence

$$\delta(x) \leq (1 + \bar{\delta} + \bar{\delta}/0.27)\delta(x(\alpha)) < 1.5\, \delta(x(\alpha)),$$

completing the proof. □

We also calculated the last result for a point x such that $\delta(x, \alpha) \leq 0.3$, obtaining for the worst case $\delta(x(\alpha)) \leq \delta(x)/3$.

Variation of the barrier function. Let us study the variation of the barrier function along the central path.

The first lemma shows differential properties of the central path.

Lemma 3.8 *Consider the functions $\alpha > 0 \mapsto x(\alpha)$ and $\alpha > 0 \mapsto p(x(\alpha))$. Then*

$$\begin{aligned} \frac{dx(\alpha)}{d\alpha} &= -\frac{1}{\alpha} X P_{AX} e, \\ \frac{dp(x(\alpha))}{d\alpha} &= -\frac{1}{\alpha} \|X P_{AX} e\|^2 = \frac{1}{\alpha}\delta(x(\alpha))^2 \end{aligned}$$

Proof. The first relation is shown for instance in [5]. Let us prove the second one.

Differentiating $p(x(\alpha)) = -\sum_{i=1}^n \ln\ x_i(\alpha)$,

$$\frac{d\,p(x(\alpha))}{d\,\alpha} = \frac{d\,p(x(\alpha))}{d\,x}\frac{d\,x}{d\alpha} = (-x^{-1})^T(-\frac{X}{\alpha} P_{AX} e) = \frac{1}{\alpha} e^T P_{AX} e = \frac{1}{\alpha}\|P_{AX} e\|^2,$$

completing the proof. □

Lemma 3.9 *Consider a point $x \in \mathcal{X}$ and α such that $r\alpha \geq 0.4$ and $\delta(x,\alpha) \leq 0.1$. Then $p(x)-p(\theta) \geq 0.018$. If x^0 is such that $\delta(x^0) \leq 0.1$, then $p(x)-p(x^0) \geq 0.012$.*

Proof. Assume without loss of generality that $\theta = e$. Then $p(\theta) = 0$. By Lemma 3.7, $\|x-e\| \geq 0.21$. Since $p(\cdot)$ is continuous, there exists a point $\bar{x} = \lambda e + (1-\lambda)x$ with $\lambda \in [0,1]$ such that $\|\bar{x}-e\| = 0.21$. Let $h = \bar{x} - e$.
The barrier function satisfies the following property proved in [5]:

$$p(e+h) = -e^T h + \frac{\|h\|^2}{2} + o(h), \quad |o(h)| \leq \frac{\|h\|^3}{3(1-\|h\|)}.$$

For $\theta = e$ we have, by definition of analytic center, $\nabla p(e)^T h = e^T h = 0$. With $\|h\| = 0.21$ in this expression, we get $p(\bar{x}) \geq 0.018$. Since $p(\cdot)$ is convex with a minimum at e and $\bar{x}$ is a point in the segment from e to x, $p(x) \geq p(\bar{x}) \geq 0.018$.

Now assume that $\delta(x^0) \leq 0.1$. Then $\left\|x^0 - e\right\| \leq 0.1004$ and using again the property of the logarithmic barrier we obtain $p(x^0) \leq 0.006$. Hence $p(x) - p(x^0) \geq 0.012$, completing the proof. □

Results for the dual problem.

Consider the dual problem (D) and assume now that its feasible set Ω is bounded with $\overset{\circ}{\Omega} \neq \emptyset$. The central path is given by

$$\alpha \geq 0 \mapsto (y(\alpha), s(\alpha)) = \text{argmin}\{\alpha b^T y + p(s);\ (y,s) \in \overset{\circ}{\Omega}\}.$$

As we saw above, with a full row rank matrix A, only one of y and s is needed to characterize the dual feasible pair (y,s). Due to the symmetry between (D') and (P), all the results derived above for (P) can be extended to (D). All we have to do here is to present the formulae for the computation of centering steps and proximities. The dual centering direction is computed in [6]. The computation is done directly for problem (D) using the Taylor expansion of the penalized function.

Given $\alpha \geq 0$ and $(y,s) \in \overset{\circ}{\Omega}$, define $\bar{A} = AS^{-1}$. Then:
In terms of y, the Newton centering direction is

$$h(y,\alpha) = (\bar{A}\bar{A}^T)^{-1}(\alpha b - \bar{A}e).$$

In terms of s, the direction is

$$h(s,\alpha) = S\bar{h}(s,\alpha), \quad \bar{h}(s,\alpha) = -\bar{A}^T(\bar{A}\bar{A}^T)^{-1}(\alpha b - \bar{A}e).$$

Note that $h(s,\alpha) = -A^T h(y,\alpha)$, as can be deduced from the equation $A^T y + s = c$.

The expression for $h(s,\alpha)$ could alternatively be obtained directly from (D') using the expression (3) for the primal Newton step, $\bar{h}(s,\alpha) = -P_{BS}(\alpha S z - e)$, with $P_{BS} = \tilde{P}_{AS^{-1}}$, and simplifying (see [6]).

We now gather some results from this section rephrased for the dual problem:

Lemma 3.10 *Given* $(y, s) \in \overset{\circ}{\Omega}$ *and* $\alpha \geq 0$, *and assuming that* Ω *is bounded,*

i) $\delta(y, \alpha) = \delta(s, \alpha) = \|\bar{h}(s, \alpha)\|$.

ii) *If* $\delta(s, \alpha) < 1$, *then* $\delta(s + h(s, \alpha), \alpha) \leq \delta(s, \alpha)^2$.

iii) *For any* $x \in \overset{\circ}{\mathcal{X}}$, $\delta(s, \alpha) \leq \delta(x, s, \alpha)$.

iv) *If* $\delta(s, \alpha) < 1$, *then* $x(s, \alpha) = \frac{s^{-1}}{\alpha}(e - \bar{h}(s, \alpha))$ *is a primal feasible solution and* $\delta(s, x(s, \alpha), \alpha) = \delta(s, \alpha)$.

v) *For* $\alpha = 0$, $y(0)$ *or* $s(0)$ *characterize the analytic center of* Ω, *and*

$$\delta(y) \equiv \delta(s) = \left\|\bar{h}(s, 0)\right\| = \left\|\tilde{P}_{AS^{-1}} e\right\| = \min\{\|xs - e\| \,;\, Ax = 0\}.$$

4 The short step path following algorithm

Now we describe the primal short step path following algorithm as presented in [6], and present some new results on how large these short steps can be. This algorithm generates a sequence (α^k) of penalty parameter values and a sequence (x^k) of points in $\overset{\circ}{\mathcal{X}}$. A sequence of unknown central points $(x(\alpha^k))$ is associated with (α^k). A proximity bound $\beta < 1$ is given and the sequences satisfy the following:

x^k is very near $x(\alpha^k)$, with $\delta(x^k, \alpha^k) \leq \beta^2$.

x^k is near $x(\alpha^{k+1})$, with $\delta(x^k, \alpha^{k+1}) \leq \beta$.

The proximity bound is usually chosen around 0.5. In our complexity study we shall use $\beta = 0.1$. The algorithm below is stated with no stopping rule.

Algorithm 4.1 *Short step path following. Given* $\beta \in (0, 1)$, $x^0 \in \overset{\circ}{\mathcal{X}}$ *and* $\alpha^0 > 0$ *such that* $\delta(x^0, \alpha^0) \leq \beta^2$.

$k := 0$.

REPEAT

Compute α^{k+1} such that $\delta(x^k, \alpha^{k+1}) = \beta$.

$x^{k+1} = x^k + h(x^k, \alpha^{k+1})$.

$k := k + 1$.

The computation of α^{k+1} is trivial. We must solve $\left\|\bar{h}(x^k, \alpha^{k+1})\right\| = \beta$, using the definition (3) with a fixed value of x^k: this amounts to solving a second degree equation. Now we study the variation of α in each step of the algorithm. The lower bound for this variation that appears in the literature is $\alpha^{k+1} \geq (1 + \frac{\nu}{\sqrt{n}})\alpha^k$, where $\nu < 1$. We will show that this estimate is usually too conservative.

Lemma 4.2 *Consider an iteration k of the short step algorithm above, and assume that $\delta(x^k) = \|P_{AX^k}e\| > \beta^2$. Then*

$$\alpha^{k+1} \geq \alpha^k(1 + \frac{\nu}{\delta(x^k)}),$$

where $\nu \geq \dfrac{\beta - \beta^2}{2}$.

Proof. Without loss of generality, we shall assume that $x^k = e$. Then $h(x^k, \alpha^k) = \bar{h}(x^k, \alpha^k)$, and $\delta(x^k, \alpha^k) = \left\|h(x^k, \alpha^k)\right\| = \left\|\alpha^k c_p - e_p\right\|$, where $c_p = P_A c$ and $e_p = P_A e$.

Let us express α^{k+1} as

$$\alpha^{k+1} = \alpha^k + \frac{\nu\alpha^k}{\|e_p\|}.$$

By hypothesis, $\delta(x^k) = \|e_p\| \geq \beta^2$. Let us study the value of ν in the expression above such that $\left\|\alpha^{k+1}c_p - e_p\right\| = \beta$.

$$\begin{aligned}
\beta &= \left\|(\alpha^k + \frac{\nu\alpha^k}{\|e_p\|})c_p - e_p\right\| \\
&= \left\|\alpha^k c_p - e_p + \frac{\nu\alpha^k c_p}{\|e_p\|}\right\| \\
&\leq \delta(x^k, \alpha^k) + \nu\alpha^k \frac{\|c_p\|}{\|e_p\|}.
\end{aligned}$$

The algorithm ensures that $\delta(x^k, \alpha^k) \leq \beta^2$. Since $\left\|\alpha^k c_p - e_p\right\| \leq \beta^2$, we have $\left\|\alpha^k c_p\right\| \leq \|e_p\| + \beta^2$. It follows that

$$\begin{aligned}
\beta &\leq \beta^2 + \nu\frac{(\beta^2 + \|e_p\|)}{\|e_p\|} \\
&\leq \beta^2 + 2\nu,
\end{aligned}$$

because $\beta^2 \leq \|e_p\|$, completing the proof. □

Variation of $p(\cdot)$ along the central path. We can finally study the variation of the barrier function due to a short step along the central path. We study the variation of $p(x(\cdot))$ between two points $x(\alpha^k)$ and $x(\alpha^{k+1})$ with α^{k+1} generated by the algorithm.

Lemma 4.3 *Consider an iteration k of the short step algorithm using $\beta = 0.1$, starting at $x^k \in \overset{\circ}{\mathcal{X}}$, $r\alpha^k \geq 0.4$ such that $\delta(x^k, \alpha^k) \leq \beta^2$. Then*

$$p(x(\alpha^{k+1})) \geq p(x(\alpha^k)) + 0.016\,\delta(x^k).$$

Proof. Assume that $r\alpha \geq 0.4$ and $\delta(x^k, \alpha^k) \leq 0.01$. Define $\tilde{\alpha} = \alpha^k(1 + \nu/\delta(x^k))$, with $\nu = 0.045$. By Lemma 3.8, $p(x(\cdot))$ increases monotonically. Using the mean value theorem and the expression in that Lemma,

$$\begin{aligned}\Delta &= p(x(\alpha^{k+1})) - p(x(\alpha^k)) \\ &\geq p(x(\tilde{\alpha})) - p(x(\alpha^k)) \\ &= (\tilde{\alpha} - \alpha^k)\frac{dp(x(\alpha))}{d\alpha} \\ &= (\tilde{\alpha} - \alpha^k)\frac{\delta(x(\alpha))^2}{\alpha},\end{aligned}$$

for some $\alpha \in [\alpha^k, \tilde{\alpha}]$. It follows that

$$\Delta \geq \frac{\tilde{\alpha} - \alpha^k}{\tilde{\alpha}}\delta(x(\alpha))^2 = \frac{\nu/\delta(x^k)}{1 + \nu/\delta(x^k)}\delta(x(\alpha))^2.$$

By construction, $\delta(x^k, \alpha) \leq \beta = 0.1$, and hence by Lemma 3.7, $\delta(x(\alpha)) \geq \delta(x^k)/1.5$. Hence

$$\Delta \geq \frac{\nu}{1 + \nu/\delta(x^k)}\frac{\delta(x^k)}{1.5^2} \geq \frac{\nu}{1 + \nu/0.2}\frac{\delta(x^k)}{1.5^2}, \tag{9}$$

because by Lemma 3.7, $\delta(x^k) \geq 0.2$. Substituting $\nu = 0.045$, we get the result in the lemma, completing the proof. □

It is interesting to compare the variation of the primal barrier function $p(x)$ with the variation of the primal-dual barrier $p(x, s) = -\sum_{i=1}^n \log(x_i s_i) = p(x) + p(s)$. Let us do this for two central points associated with α^k and $\alpha^{k+1} = \alpha^k(1 + \eta)$, with $\eta = \nu/\delta(x(\alpha^k))$. We have

$$x^k s^k = \frac{1}{\alpha^k}e, \quad x^{k+1}s^{k+1} = \frac{1}{\alpha^{k+1}}e,$$

and hence $p(x^k, s^k) = n\log(\alpha^k)$ and $p(x^{k+1}, s^{k+1}) = n\log(\alpha^{k+1})$. It follows that $p(x^{k+1}, s^{k+1}) = n\log(\alpha^{k+1}/\alpha^k) = n\log(1 + \eta) \approx n\eta$. For $\eta = \nu/\delta(x^k) \geq \nu/\sqrt{n}$, $p(x^{k+1}, s^{k+1}) - p(x^k, s^k) \geq \nu\sqrt{n}$.

So, the primal-dual barrier function varies by more than $\nu\sqrt{n}$ at each iteration, while for the variation of $p(x)$ our lower bound is $0.016\delta(x^k)$. Far from θ, $\delta(x^k)$ will in general be of the order of $\sqrt{n}$, and the bounds are similar, but near θ this indicates that the variations of $p(x)$ and $p(s)$ may be very uneven.

5 Relationship between primal and primal-dual models

In this section we compare the Newton centering directions obtained for the primal-only and for the primal-dual models. Let us define the problem: we assume without loss of generality that the given primal point is $x = e$.

Assume that $x = e \in \overset{\circ}{\mathcal{X}}$, and $\delta(x, \alpha) \le \beta < 0.25$. We shall use the following notation: $c_p \equiv P_A c$, $e_p \equiv P_A e$, $\tilde{e}_p = e - e_p$, $\mu = 1/\alpha$. Consider also a parameter value $\mu^+ = \gamma\mu$, with $\gamma \in (0, 1]$, and define $\alpha^+ = 1/\mu^+$.

Let s be a dual interior solution such that $\delta(x, s, \alpha) \le \beta$. We shall see that primal methods do not need the knowledge of s: either it is irrelevant or it can be trivially constructed.

A centering step computes from (x, s) a point (x^+, s^+) which should approach the solution of the equivalent equations below:

$$xs - \mu^+ e = 0, \qquad (x, s) \text{ interior primal-dual solution,} \tag{10}$$

$$s - \mu^+ x^{-1} = 0, \qquad (x, s) \text{ interior primal-dual solution.} \tag{11}$$

The Newton steps for (10) and (11) are respectively the centering steps for the primal-dual and primal models. This paper has studied the primal model up to now. Now we state the centering steps for both models and compare them (keeping in mind that $x = e$, as an unnecessary but convenient simplification).

Primal-dual model. The Newton step, denoted $(\tilde{u}, \tilde{v})$ is the solution of

$$\tilde{v} + s\tilde{u} = -s + \mu^+ e, \quad \tilde{u} \in \mathcal{N}(A),\ \tilde{v} \in \mathcal{R}(A^T). \tag{12}$$

Defining the diagonal scaling matrix $D = \sqrt{\mu S^{-1}}$, we obtain (see for instance [6])

$$\tilde{u} = DP_{AD}D(-\alpha s + \gamma e), \quad \tilde{v} = D^{-1}P_{AD}D(-s + \mu^+ e), \tag{13}$$

Primal model. The Newton step will be denoted (h, v):

$$v + \mu^+ h = -s + \mu^+ e, \quad h \in \mathcal{N}(A),\ v \in \mathcal{R}(A^T), \tag{14}$$

with solution

$$\begin{array}{rcl} h & = & P_A(-\alpha^+ s + e) = -\alpha^+ c_p + e_p, \\ v & = & \tilde{P}_A(-s + \mu^+ e). \end{array} \tag{15}$$

Here we used the fact that for any feasible dual solution s, $P_A s = P_A c$, because $s = c - A^T y$. The dual component of the step is not needed, and we can see directly from (14) that

$$s^+ = s + v = \mu^+(e - h) = c_p + \mu^+ \tilde{e}_p, \tag{16}$$

as we already knew from Theorem 3.2. If we want to define s and v for the given x and α, then we can use this last expression with $\mu^+ = \mu$ to obtain

$$s = \mu(e - h) = c_p + \mu\tilde{e}_p, \tag{17}$$

and consequently

$$v = s^+ - s = -(1 - \gamma)\mu\tilde{e}_p. \tag{18}$$

Short step algorithms. Let us compare the steps for both models in short step algorithms. We are not interested in a precise comparison, and so we shall simplify the treatment by assuming that $x = e$ is the central point $x(\alpha)$.
Primal short step: we know from Lemma 4.2 that

$$\frac{\alpha^+}{\alpha} \geq 1 + \frac{\beta - \beta^2}{2\delta(x)}.$$

Primal-dual model: we have $\alpha xs = e$ and need $\|\alpha^+ xs - e\| = \beta$. Substituting $xs = e/\alpha$, we obtain $(\alpha^+/\alpha - 1)\,\|e\| = \beta$ and thus

$$\frac{\alpha^+}{\alpha} = 1 + \frac{\beta}{\sqrt{n}}.$$

Comparing these results, we see that whenever $\delta(x) << \sqrt{n}$, which happens near the analytic center, the primal short step is more efficient than the primal-dual short step. Far from the analytic center, $\delta(x)$ is of the order of $\sqrt{n}$, and the short steps become comparable.

Large step path following algorithms. We know that the primal-dual Newton step is very efficient, and has nice asymptotic properties, leading to quadratically convergent predictor-corrector algorithms like the one by Mizuno, Todd and Ye [14]. The resulting point $(x^+, s^+) = (x, s) + (\tilde{u}, \tilde{v})$ will be feasible for small values of γ when (x, s) approaches the optimal face.

Let us compare the directions obtained by the primal and primal-dual models:
Dual directions: if $x = e$ is nearly central, then $s \approx \mu e$, $D \approx I$ and $D^{-1}\tilde{P}_{AD}D \approx \tilde{P}_A$, by Lemma 2.3. It follows that

$$\tilde{v} \approx \tilde{P}_A(-s + \mu^+ e) = v.$$

The dual directions do not differ much in both models from well centered points.
Primal directions: here the situation will be different. Using the same argument, we immediately see from (13) that for well centered points

$$\tilde{u} \approx P_A(-\alpha s + \gamma e) = \gamma h.$$

The step $x^+ = x + h$ overshoots badly for $\gamma < 1$, and it is easy to show that from central points the result will be infeasible for $\gamma \leq 0.5$. This suggests that the best step for primal-only algorithms is given by

$$u = \gamma h = -\alpha c_p + \gamma e_p. \tag{19}$$

Now both models give the same results from central points, and we can construct a large step primal-only path following algorithm. The algorithm will follow the scheme of the largest step method studied by McShane [12] and in [8]. We shall only describe a typical iteration, starting from $x \in \overset{\circ}{\mathcal{X}}$ and $\alpha > 0$ such that $\delta(x, \alpha) \leq \beta < 0.5$. Note that dual variables are used but not stored, as usually happens in primal-only methods.

Algorithm 5.1 *Primal path following iteration.*

Compute $c_p = P_{AX}Xc$, $e_p = P_{AX}e$, $\tilde{e}_p = e - e_p$.

Solve the following one-dimensional problem: find $\gamma \in (0,1)$ such that

$$\begin{aligned} \left\| \frac{\bar{x}^+ \bar{s}^+}{\gamma\mu} - e \right\| &= \beta \\ \bar{x}^+ &= e - \alpha c_p + \gamma e_p \\ \bar{s}^+ &= c_p + \gamma\mu\tilde{e}_p. \end{aligned}$$

Set $x^+ = X\bar{x}^+$, $s^+ = X^{-1}\bar{s}^+$, $\mu^+ = \gamma\mu$, $\alpha^+ = 1/\mu^+$.

Corrector step: $x^{++} = x^+ + h(x^+, \alpha^+)$.

The corrector step at the end of the iteration may or may not be used. If it is used, then $\delta(x^{++}, \alpha^+) \leq \beta^2$. The main question now is whether this algorithm can be superlinearly convergent, or maybe quadratically convergent.

Let us address this question. Assume as above that $x = e$, and use the modified Newton step $u = \gamma h$ as in the algorithm. Let us derive an expression for the primal-dual proximity $\delta(x+, s^+, \alpha^+)$ at the resulting point. From (14), $v + s = -\mu u + \mu^+ e$. We get

$$\begin{aligned} x^+ s^+ = (e+u)(s+v) &= s + v + us + uv \\ &= \mu^+ e - \mu u + us + uv. \end{aligned}$$

Dividing by $\mu^+ = \gamma\mu$,

$$\frac{x^+ s^+}{\mu^+} - e = \frac{1}{\gamma}(-u + \frac{us}{\mu}) + \frac{uv}{\gamma\mu}.$$

Taking norms and setting $\delta(x^+, s^+, \alpha^+) = \beta$ we get

$$\beta = \left\| \frac{x^+ s^+}{\mu^+} - e \right\| \leq \frac{\|u\|}{\gamma} \left\| \frac{s}{\mu} - e \right\| + \left\| \frac{uv}{\gamma\mu} \right\|,$$

and finally

$$\left\| \frac{uv}{\gamma\mu} \right\| + \frac{\|u\|}{\gamma}\delta(x, s, \alpha) \geq \beta.$$

Doing the same computation for the primal-dual model, only the first term of the left hand side appears, and then each iteration will make $\left\| u^k v^k / \gamma^k \mu^k \right\| = \beta$. It is known that $\left\| u^k v^k / \mu^k \right\| \to 0$, and this implies in $\gamma^k \to 0$. Hence $\mu^k \to 0$ superlinearly (see McShane [12]). If corrector steps are used (even very sparingly), then quadratic convergence is obtained (see [8]).

In the primal setting it is reasonable to expect that the first term will behave similarly, but the other term poses problems. To achieve superlinear convergence (i. e. $\gamma^k \to 0$) we need $\left\| u^k \right\| \delta(x^k, s^k, \alpha^k) \to 0$, and $\left\| u^k \right\|$ is of the order of $\sqrt{n}$. So we need $\delta(x^k, s^k, \alpha^k) \to 0$. This is not true if corrector steps are not used, because by construction $\delta(x^k, s^k, \alpha^k) = \beta$. Hence superlinear convergence cannot be proved.

In fact, superlinear convergence should never happen in this case, due to the following argument: near the central path, Newton large steps approach the primal afine-scaling step, which has an exaggerated centering discussed by Tsuchiya and Monteiro [18]. It approaches the central path near halfway to the boundary and then starts moving away from it. We can hope for a steplength of about 2/3 of the maximum step, like in their affine-scaling algorithm.

If each iteration does one centering, then $\delta(x^k, s^k, \alpha^k) \leq \beta^2$, and the steps may be much longer.

In a forthcoming paper we intend to show that as $\mu^k \to 0$, the efficiency of the centering steps increases, so that using a corrector step causes $\delta(x^k, s^k, \alpha^k) \to 0$. At this time this is a conjecture: if proved true, then the primal-only predictor-corrector algorithm is superlinearly convergent.

References

[1] I. Adler and R. D. C. Monteiro, *Limiting behavior of the affine scaling continuous trajectories for linear programming problems*, Mathematical Programming, 50 (1991), pp. 29–51.

[2] J. L. Goffin, Z. Q. Luo, and Y. Ye, *Complexity analysis of an interior cutting plane method for convex feasibility problems*, SIAM Journal on Optimization, 6 (1996), pp. 638–652.

[3] J. L. Goffin and J. P. Vial, *Cutting planes and column generation techniques with the projective algorithm*, Journal of Optimization Theory and Applications, 65 (1990), pp. 409–429.

[4] J. L. Goffin and J. P. Vial, *Shallow, deep and very deep cuts in the analytic center cutting plane method*, Technical Report 96–3, Dept. of Management Studies, Univertity of Geneva, Geneve, Switzerland, 1996.

[5] C. C. Gonzaga, *An algorithm for solving linear programming problems in $O(n^3L)$ operations*, in Progress in Mathematical Programming : Interior Point and Related Methods, N. Megiddo, ed., Springer Verlag, New York, 1989, pp. 1–28.

[6] C. C. Gonzaga, *Path following methods for linear programming*, SIAM Review, 34 (1992), pp. 167–227.

[7] C. C. Gonzaga, *Search directions for interior linear programming methods*, Algorithmica, 6, pp. 153–185.

[8] C. C. Gonzaga, *The largest step path following algorithm for monotone linear complementarity problems*, Mathematical Programming, 76 (1997), pp. 309–332.

[9] C. C. Gonzaga and M. Cardia, *Using the central path for computing the analytic center of a polytope*, technical report, Dept. of Mathematics, Federal University of Santa Catarina, Florianópolis, Brazil, 2000.

[10] O. Güler, *Limiting behavior of the weighted central paths in linear programming*, Mathematical Programming, 65 (1994), pp. 247–382.

[11] M. Iri and H. Imai, *A multiplicative barrier function method for linear programming*, Algorithmica 1 (1986), 455–482.

[12] K. A. McShane, *A superlinearly convergent $O(\sqrt{n}L)$ iteration primal–dual linear programming algorithm*, SIAM Journal on Optimization, 4 (1994), pp. 247–261.

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

[14] S. Mizuno, M. J. Todd, and Y. Ye, *On adaptive step primal–dual interior–point algorithms for linear programming*, Mathematics of Operations Research, 18 (1993), pp. 964–981.

[15] Y. Nesterov, *Cutting plane algorithms from analytic centers: efficiency estimates*, Mathematical Programming, 69 (1995), pp. 149–176.

[16] C. Roos, T. Terlaky and J. P. Vial, *Theory and Algorithms for Linear Optimization: An Interior Point Approach*, John Wiley & Sons, Chichester, 1997.

[17] R. Saigal, *A three step quadratically convergent version of primal affine scaling method*, Journal of Operations Research Society of Japan, 40 (1997), pp. 310–328.

[18] T. Tsuchiya and R. D. C. Monteiro, *Superlinear convergence of the affine-scaling algorithm*, Mathematical Programming, 75 (1996), pp. 77–110.

[19] T. Tsuchiya, *Quadratic convergence of the Iri-Imai algorithm for degenerate linear programming problems*, J. Optim. Tho. Appl., 87 (1995), pp. 703–726.