CORVINUS
ECONOMICS
WORKING
PAPERS

05/2024

Marianna E. Nagy, Tibor Illés, Yurii Nesterov and Petra Renáta Rigó

New Interior-Point Algorithm for Linear Optimization Based on a Universal Tangent Direction

http://unipub.lib.uni-corvinus.hu/10470

CORVINUS UNIVERSITY OF BUDAPEST

NEW INTERIOR-POINT ALGORITHM FOR LINEAR OPTIMIZATION BASED ON A UNIVERSAL TANGENT DIRECTION

MARIANNA E.-NAGY*, TIBOR ILLÉS†, YURII NESTEROV‡, AND PETRA RENÁTA RIGÓ§

Abstract. In this paper, we suggest a new interior-point method for linear optimization, based on the idea of Parabolic Target Space. Our method can start at any strictly feasible primal-dual pair and go directly towards a solution by a predictor-corrector scheme. Each iteration needs inversion of a matrix in small dimension. The worst-case upper bound for the number of matrix factorizations is $O\left(\sqrt{n}\ln\frac{1}{\varepsilon}\right)$, where n is the dimension of the problem. Our preliminary computational experiments show that the actual number of iterations depends logarithmically on the dimensions. In the end, the size of the predictor step typically goes above 99.97% of the maximal possible step to the boundary. The method demonstrates a very fast local convergence. One of the main differences between our approach and the standard framework is that the method is based on a parabolic barrier function.

Key words. Linear optimization, interior-point algorithms, parabolic target space, universal tangent direction.

JEL codes. C61

1. Introduction. The development of the polynomial-time interior-point algorithms (IPAs) for linear optimization (LO) started from the seminal paper by Karmarkar [9], where he proposed a new projective algorithm, based on the description of the positive orthant by an analytic barrier. As compared with the previously known polynomial-time ellipsoid method [10, 14], the theoretical progress was not very significant. However, the new algorithm had an important advantage of possible acceleration on problems with an appropriate internal structure.

These ideas were transformed into a more conventional form by Renegar, who introduced the methods of centers [21], and Gonzaga, who proposed the polynomial-time path-following barrier methods [7]. These contributions were based on explicit use of the *central path*, the object which has been already known by Fiacco and McCormick [6], and which was actively popularized by Sonnevend [23]. All new methods were computationally expensive since at each iteration it was necessary to invert a matrix, formed as a sum of n rank-one $(m \times m)$-matrices $(m < n)$.

Despite the attractive theoretical bound $O\left(\sqrt{n}\ln\frac{n}{\varepsilon}\right)$ on the total number of iterations, the first versions of the new methods were not very efficient since they followed the central path by small steps. This disadvantage was eliminated in the primal-dual settings by Mehrotra [12], who introduced a large infinity-neighbourhood of the central path, allowing much bigger steps. From the analysis of one of its versions (e.g. Section 14.1 in [18]), we know its worst-case complexity bound $O\left(n\ln\frac{n}{\varepsilon}\right)$. However, its practical behavior is usually much better.

Another predictor-corrector algorithm, working with the Euclidean neighbour-

*Corvinus Centre for Operations Research, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Fővám tér 8., H-1093 Budapest, Hungary (marianna.eisenberg-nagy@uni-corvinus.hu)

†Corvinus Centre for Operations Research, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Fővám tér 8., H-1093 Budapest, Hungary (tibor.illes@uni-corvinus.hu)

‡Corvinus Centre for Operations Research, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Fővám tér 8., H-1093 Budapest, Hungary (yurii.nesterov@uni-corvinus.hu)

§Corvinus Centre for Operations Research, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Fővám tér 8., H-1093 Budapest, Hungary (petra.rigo@uni-corvinus.hu)

hood of the central path, was proposed by Mizuno et al. in [13]. A short-step version of this scheme shares the best-known complexity bound of $O\left(\sqrt{n}\ln\frac{n}{\varepsilon}\right)$ iterations. Today, the most efficient practical IPAs, implement one of these two strategies, trying to enlarge somehow the size of the predictor step and keeping the number of corrector steps at some reasonable level. As a paradox, we can confirm that the most efficient practical methods today do not share the best theoretical complexity bounds.

All these methods, working in the primal-dual space, require two stages of the process even if we know a strictly feasible pair of starting points. Indeed, in the first stage, we need to find a point in a small neighbourhood of the central path, and in the second stage, we follow this curve.

This inconvenience served as the main motivation for the development of alternative approaches. One of them is related to potential-reduction methods (see [24, 25]). However, the potential functions in this approach explicitly depend on the dimension and this reduces the adaptivity of such schemes to problems with simple internal structure.

The second approach is called *target-following* framework [22], which will be discussed in detail later on. Nesterov [16] cused a new target-following approach for LO, a long-step greedy target-following algorithm, which can start at an arbitrary strictly feasible solution. In this method, he embedded the problem into a higher-dimensional one introducing control variables according to parabolic target space (PTS). On the extended space, he defined a new parabolic barrier function, which plays a key role in the new algorithm.

In this paper, we analyse a new version of the long-step PTS greedy algorithm from [16], based on the *universal tangent direction* defined in (4.5). This direction is obtained by using the above-mentioned parabolic barrier function (3.3). In the algorithm proposed in [16] the search directions are obtained by Newton's method applied to the barrier function, which needs invertation of the Hessians. On the other hand, in this paper, we define the search directions by solving system (2.4) with different right-hand sides, which is less expensive than the computation of the search directions in [16]. We prove that the complexity of the new long-step IPA coincides with the known best ones in the theory of IPAs. Furthermore, we provide preliminary computational results on randomly generated test set problems. The method demonstrates a very fast local convergence.

The paper is organized as follows. In Section 2, we give the problem formulation and the system for determining search directions. In Section 3 we present the parabolic target-following approach and we introduce the main definitions of the new barrier function and functional proximity measures. Section 4 presents the new IPA in PTS based on a universal tangent direction, which alternates predictor and corrector steps. Section 5 is devoted to the worst-case complexity analysis of the algorithm. First, we prove that the length of the corrector stage is bounded by an absolute constant. After that, we show that the progress achieved by the predictor step is sufficient for bounding the total number of iterations by the value $O\left(\sqrt{n}\ln\frac{\mu^*(w^0)}{\varepsilon}\right)$, where $\mu^*(w^0)$ is the merit function of the process, evaluated at the starting point, and ε is the required accuracy of the solution. In Section 6, we present the results of preliminary testing on a large set of random problems (1500 instances) and we show that our results are very promising on this generated test set problems.

Our notations are standard. We work with column vectors in $\mathbb{R}^n$ and denote by $\mathbb{R}^n_\oplus$ the nonnegative orthant and by $\mathbb{R}^n_+$ its interior. For two vectors $x, y \in \mathbb{R}^n$, we use notation xy an x/y for the vector formed by componentwise operations with

the components of x and y. For the norms, we denote by $\|x\| = \sqrt{x^T x}$ and $\|x\|_1 = \sum_{i=1}^{n} |x^{(i)}|,\ x \in \mathbb{R}^n$. Notation $e \in \mathbb{R}^n$ is used for the vector of all ones, and $\bar{e}$ is the vector of all ones in $\mathbb{R}^{n+1}$. Note that X denotes the diagonal matrix containing the components of the vector x on the main diagonal. Let $G : \mathbb{R}^k \to \mathbb{R}^n$ be an operator and $z, \Delta z \in \mathbb{R}^k$. Then, we use the notation $\mathcal{D}G(z)[\Delta z]$ for the directional derivative of G in z along the direction Δz. Let us consider the functions defined in [15]

$$\omega : \mathbb{R}_+ \to \mathbb{R}, \qquad \omega(t) = t - \ln(1+t) \tag{1.1}$$

and

$$\omega_* : [0,1) \to \mathbb{R}, \qquad \omega_*(t) = -t - \ln(1-t), \tag{1.2}$$

that will be used in the analysis of the algorithm.

2. Problem formulation. In this paper, we propose a new efficient long-step method based on a universal tangent direction for solving the primal-dual pair of linear programming problems. The primal problem is the following:

$$\min \left\{ c^T x : \ Ax = b, \ x \geq 0 \right\}, \tag{P}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Without loss of generality, we may assume that $\mathrm{rank}(A) = m$. The dual of this problem is the following:

$$\max \left\{ b^T y : \ A^T y + s = c, \ s \geq 0 \right\}. \tag{D}$$

The *optimality criteria* of the primal-dual pair (P) and (D) can be formulated as:

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= 0. \end{aligned} \tag{2.1}$$

We denote by

$$\mathcal{F} = \left\{ u = (x, y, s) \in \mathbb{R}^n_{\oplus} \times \mathbb{R}^m \times \mathbb{R}^n_{\oplus} : Ax = b, A^T y + s = c \right\}$$

the feasible set of the primal-dual problem (P) and (D). We assume the existence of an interior point, namely that the set of interior points

$$\mathcal{F}^+ = \{ u = (x, y, s) \in \mathcal{F} : x > 0, s > 0 \}$$

is not empty. We define the duality gap

$$c^T x - b^T y = (A^T y + s)^T x - b^T y = s^T x, \tag{2.2}$$

where $(x, y, s) \in \mathcal{F}$. We define the *central path* for LO problems in the standard way [22, 23], as

$$Ax = b, \ A^T y + s = c, \quad x, s > 0, \quad xs = \mu\, e, \tag{2.3}$$

where $\mu > 0$ and e is the all one vector.

In order to define the search directions for classical IPAs for LO, we apply Newton's method on the barrier function, which results in the following linearized system:

$$\begin{aligned} A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \\ S\Delta x + X\Delta s &= a, \end{aligned} \tag{2.4}$$

where $a \in \mathbb{R}^n$ is properly selected, see for example in [1, 2, 3, 5, 8, 11, 19, 20, 22]. Note that system (2.4) has a unique solution for any $a \in \mathbb{R}^n$ vector, see [22]. Hence, $\Delta y = -[AS^{-1}XA^T]^{-1}S^{-1}a$ and $\Delta x = S^{-1}(a + XA^T\Delta y)$. Moreover, due to the structure of the linear system in (2.4), independent on the right-hand side a, we have

$$\Delta x^T \Delta s = 0. \tag{2.5}$$

In order to estimate the norms of $\Delta x \Delta s$, we use the classical method from the literature, see [22]:

$$\|\Delta x \Delta s\| \leq \left\| \sqrt{\frac{s}{x}}\Delta x \sqrt{\frac{x}{s}}\Delta s \right\|_1 \leq \frac{1}{2}\left\| \sqrt{\frac{s}{x}}\Delta x + \sqrt{\frac{x}{s}}\Delta s \right\|^2 = \frac{1}{2}\left\| \frac{a}{\sqrt{xs}} \right\|^2. \tag{2.6}$$

Usually, the direction defined by (2.4) is applied to the classical logarithmic barrier function for positive orthant. However, we show that it is applicable to a parabolic barrier function, too.

In the following section, we present the parabolic target-following framework.

3. Target-following framework for LO and its modification. The main difference between target-following approaches and other IPAs lies in the fact that IPAs follow the central path, while the target-following approaches follow the so-called *weighted central path* defined as

$$Ax = b,\ A^T y + s = c, \quad x, s > 0, \quad xs = \bar{v}, \tag{3.1}$$

where $\bar{v} \in \mathbb{R}^n$ is a positive vector of control parameters, see [22]. The weighted central path has a unique solution [16], namely $(x(\bar{v}), y(\bar{v}), s(\bar{v}))$. The main advantage of this approach is that any strictly feasible starting point $u^0 = (x^0, y^0, s^0)$ defines a starting point of this path by choosing $\bar{v}_i = x_i^0 s_i^0$, $i = 1, \ldots, n$. Thus, we can immediately start the main stage of the process with $\bar{v} > 0$.

However, the target-following approach suffers from a serious drawback. The point $x(\bar{v})$ is characterized there as a solution of the following *weighted center problem*:

$$\min\ c^T x - \sum_{i=1}^{n} \bar{v}_i \ln x_i, \quad \text{such that}\ \ Ax = b. \tag{3.2}$$

The complexity of finding an approximate solution to this problem depends on the condition number of the weights

$$\kappa(\bar{v}) \quad = \quad \max_{1 \leq i,j \leq n} \frac{\bar{v}_i}{\bar{v}_j}.$$

If the weights are well balanced ($\kappa(\bar{v}) \approx 1$), then the problem is easy. However, then the point $x(\bar{v})$ is close to the classical central path. If $\kappa(\bar{v})$ is big, then this operation

is expensive. At the same time, during the minimization process, it is difficult to keep the condition number reasonably small, especially at the beginning.

The above difficulty was eliminated in [16] by replacing in (3.1) the nonlinear equations by *convex inequalities* and introducing the *parabolic target space (PTS)* defined as

$$\mathcal{F}_w \quad = \quad \Big\{ w = (v_0, v) \in \mathbb{R}_+ \times \mathbb{R}^n : \; v_0 \geq \|v\|^2 \Big\},$$

which is a convex set and denote its interior by $\mathcal{F}_w^+$. Following the steps discussed in [16] for LO and in [4] for weighted LCP, we can relax the third equation of (2.1) for the original variable $u = (x, y, s)$ of (P)-(D) from two points of view using the variable $w = (v_0, v)$ of PTS. In this way, we can define the convex set for the joint variable $z = (u, w)$ as follows:

$$\mathcal{F}_z = \Big\{ z = (u, w) \in \mathcal{F} \times \mathcal{F}_w : \quad xs \geq v^2, \quad v_0 \geq c^T x - b^T y \Big\}.$$

Similarly to the set $\mathcal{F}$, due to the affine constraints, we will work with the relative interior of the set, which for $\mathcal{F}_z$ is

$$\mathcal{F}_z^+ = \Big\{ z = (u, w) \in \mathcal{F}_z : \quad xs > v^2, \quad v_0 > c^T x - b^T y \Big\}.$$

Note that $\mathcal{F}_z^+$ is nonempty, because $\mathcal{F}^+$ is nonempty. It admits a $(2n+1)$-self-concordant barrier

$$F(z) = -\sum_{i=1}^{n} \ln \left(x_i s_i - v_i^2 \right) - \ln \left(v_0 - c^T x + b^T y \right), \quad z \in \mathcal{F}_z^+. \tag{3.3}$$

For $z = (u = (x, y, s), w = (v_0, v)) \in \mathcal{F}_z$, we can define the following residuals

$$r_0(z) = v_0 - s^T x, \quad r_i(z) \; = \; x_i s_i - v_i^2, \quad i = 1, \ldots, n, \tag{3.4}$$

hence $r(z) \in \mathbb{R}^{n+1}_{\oplus}$. Note that the average of the residuals does not depend on u. Indeed,

$$\frac{\bar{e}^T r(z)}{n+1} = \frac{v_0 - \|v\|^2}{n+1} =: \rho(w). \tag{3.5}$$

Using (3.4), the self-concordant barrier given in (3.3) can be written as

$$F(z) \quad = \quad -\sum_{i=0}^{n} \ln(r_i(z)), \quad z \in \mathcal{F}_z^+. \tag{3.6}$$

We define the control barrier function, $\phi : \mathbb{R}^{1+n} \to \mathbb{R}$ as

$$\phi(w) = \min_u \left\{ F(u, w) : (u, w) \in \mathcal{F}_z^+ \right\}, \quad w \in \mathcal{F}_w^+, \tag{3.7}$$

which is a self-concordant barrier for its domain. Following the ideas used in Section 2 of [16], Lemma 3.1 shows that any $w \in \mathcal{F}_w^+$ determines a unique point u in $\mathcal{F}^+$. Furthermore, it gives a closed-form representation for the function ϕ.

LEMMA 3.1. *Let $w \in \mathcal{F}_w^+$. Then, $u(w) = (x(w), s(w), y(w)) := \operatorname{argmin}_u \{F(u, w) : (u, w) \in \mathcal{F}_z^+\}$ is the unique solution of the following system:*

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ xs &= v^2 + \rho(w) e, \end{aligned} \tag{3.8}$$

where $\rho(w)$ is given in (3.5). Moreover, we have

$$\phi(w) = -(n+1)\ln\rho(w). \tag{3.9}$$

Based on Lemma 3.1, any sequence $\{w^k\} \subset \mathcal{F}_w^+$ uniquely defines a sequence $\{u(w^k)\} \subset \mathcal{F}^+$. Moreover, if w^k goes to 0, then $u(w^k)$ tends to a solution of the problem (P)-(D). We can consider the line segment $w(t) = tw^0 = t(v_0^0, v^0)$, $t \in (0,1]$, i.e. the greedy scheme used in [16], which defines the following parameterized system:

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ xs &= t\frac{v_0^0}{n+1}e + t^2\left(\left(v^0\right)^2 - \frac{\|v^0\|^2}{n+1}e\right), \end{aligned} \tag{3.10}$$

which can be considered as a new type of central path problem. In our algorithm, we follow this new central path.

If v^0 is the zero vector, then the second term in the last equation of the system (3.10) disappears, hence we get back the classical central path. However, our path-following algorithm is new even in this case since it is based on the functional proximity measure defined by (3.19).

Consider now the general system (3.8). The last equation of system (3.8) is equivalent with $r_i(z) = \rho(w)$, for all $1 \le i \le n$. Summing up the coordinates of the last equation of system (3.8), we get

$$s(w)^T x(w) \;=\; \|v\|^2 + n\rho(w) \;=\; \tfrac{nv_0+\|v\|^2}{n+1}, \tag{3.11}$$

thus

$$r_0(u(w), w) = v_0 - s(w)^T x(w) = \rho(w). \tag{3.12}$$

For $z \in \mathcal{F}_z^+$, let $\hat{r}(z) \in \mathbb{R}_+^{n+1}$ be defined as

$$\hat{r}(z) = \sqrt{\frac{r(z)}{\rho(w)}}. \tag{3.13}$$

We have

$$\bar{e}^T\hat{r}^2(z) = \|\hat{r}(z)\|^2 = n+1. \tag{3.14}$$

It should be mentioned that $\hat{r}(z) = \bar{e}$ if and only if $u = u(w)$. Note that we cannot efficiently compute $u(w)$. However, using (3.13) we can define the following measures of closeness of point u to $u(w)$:

$$\zeta_0^2(z) \;=\; \left\|\hat{r}(z) - \tfrac{1}{\hat{r}(z)}\right\|^2 = \bar{e}^T\left(\tfrac{1}{\hat{r}^2(z)} - \bar{e}\right), \quad \zeta_1(z) = \left\|\tfrac{1}{\hat{r}^2(z)} - \bar{e}\right\| \tag{3.15}$$

and

$$\delta(z) \;=\; \frac{\zeta_0^2(z)}{\zeta_1(z)} \quad (\text{for } \zeta_1(z) = 0, \text{ we define } \delta(z) = 0). \tag{3.16}$$

Using $\delta(z)$, we can define the following neighbourhood with parameter $\beta > 0$:

$$\mathcal{N}_\delta(\beta) = \{z = (u, w) \in \mathcal{F}_z^+ : \delta(z) \le \beta\}. \tag{3.17}$$

Another possibility to measure the closeness of u to $u(w)$ is based on the definition of the control barrier function ϕ and Lemma 3.1. For the unique optimal solution $u(w)$ of problem (3.7), we have $F(u(w), w) = \phi(w)$ and $F(z) \geq \phi(w)$ for all $z = (u, w) \in \mathcal{F}_z^+$. Therefore, let us define the *functional proximity measure*

$$\Psi(z) = F(z) - \phi(w) = -\sum_{i=0}^{n} \ln r_i(z) + (n+1) \ln \rho(w) = -\sum_{i=0}^{n} \ln \hat{r}_i^2(z) \geq 0. \tag{3.18}$$

We can introduce the following neighbourhood using the functional proximity measure:

$$\mathcal{N}_\Psi(\tau) = \{z = (u, w) \in \mathcal{F}_z^+ : \Psi(z) \leq \tau\}, \tag{3.19}$$

where $\tau > 0$.

4. Interior-point algorithm for linear optimization based on a universal tangent direction. Let us present the algorithmic scheme. Our method starts from an interior point $u^0 = (x^0, y^0, s^0) \in \mathcal{F}^+$. In order to initiate a scheme in the PTS, we need to point out a starting control variable $w(u) = (v_0(u), v(u))$ such that $u = u(w(u))$. Let us present the rules for computing such a point. Let

$$\begin{aligned} \xi(u) &\stackrel{\text{def}}{=} \min_{1 \leq i \leq n} x_i s_i, \quad v_0(u) = s^T x + \xi(u), \\ v(u)_i &= \sqrt{x_i s_i - \xi(u)}, \ i = 1, \ldots, n. \end{aligned} \tag{4.1}$$

In this case, $\rho(w) = \xi(u)$, so the point u satisfies the last equation of system (3.8) with $w = w(u)$. Note that in [16], the search directions were obtained by applying Newton's method on the parabolic barrier function. This involved the invertation of the Hessian, which is $(3n + m + 1) \times (3n + m + 1)$-dimensional. In order to avoid the Hessian and reduce the size of the coefficient matrix of the corresponding system, in this paper we determine the search directions by considering systems of the form (2.4), where we get the right-hand side a based on the last equation of system (3.8). In this way, let us introduce the following quadratic operator $G : \mathcal{F}_z^+ \to \mathbb{R}^n$:

$$G(z) = xs - v^2 - \rho(w)e. \tag{4.2}$$

From Lemma 3.1 we have $G(u(w), w) = 0$. Let

$$z(\alpha) = z + \alpha \Delta z. \tag{4.3}$$

Then,

$$G(z(\alpha)) = G(z) + \alpha \mathcal{D}G(z)[\Delta z] + \frac{\alpha^2}{2} \mathcal{D}^2 G[\Delta z]^2. \tag{4.4}$$

We will work with the *universal tangent direction* Δz for the operator G defined as

$$\mathcal{D}G(z)[\Delta z] = 0. \tag{4.5}$$

This means that

$$\mathcal{D}G(z)[\Delta z] = S\Delta x + X\Delta s - 2v\Delta v - \frac{1}{n+1}(\Delta v_0 - 2v^T \Delta v)e = 0. \tag{4.6}$$

Furthermore,

$$\mathcal{D}^2 G[\Delta z]^2 = 2\left(\Delta x \Delta s - (\Delta v)^2 + \frac{1}{n+1}\|\Delta v\|^2 e\right). \tag{4.7}$$

Thus, we have

$$G(z(\alpha)) = G(z) + \alpha^2\left(\Delta x \Delta s - (\Delta v)^2 + \frac{1}{n+1}\|\Delta v\|^2 e\right). \tag{4.8}$$

We will deal with the greedy direction $\Delta_g z = (\Delta x, \Delta s, -v, -v_0)$ in the PTS, hence our aim is to take a step towards the origin.

Using the greedy direction in the PTS and (4.6), we obtain

$$s\Delta x + x\Delta s + 2v^2 + \frac{1}{n+1}\left(v_0 - 2\|v\|^2\right) e = 0. \tag{4.9}$$

Hence, we obtain the predictor search direction by solving system (2.4) with

$$a = \left(\frac{\|v\|^2}{n+1} - \rho(w)\right) e - 2v^2 = \frac{e}{n+1} v_0 - 2x(w)s(w), \tag{4.10}$$

where the second equality follows from the last equation of system (3.8). This means that we do not apply the Newton's method on the parabolic barrier function F, therefore we simply call system (2.4) as linearized system instead of Newton-system.

After a predictor step, it might happen that $\delta(z) > \beta$. Therefore, we need to perform corrector steps towards $u(w)$. Namely, using (3.8) our aim is to have $(x + \Delta x)(s + \Delta s) = v^2 + \rho(w)e$. Neglecting the quadratic term, the corrector search direction is obtained by solving system (2.4) with

$$a_i := \rho(w) - r_i(z), \quad i = 1, \ldots, n. \tag{4.11}$$

Algorithm 4.1 IPA for LO based on a universal tangent direction

Require: the initial point $u_0 \in \mathcal{F}^+$ and the corresponding $w_0 = w(u_0)$,
$\beta \in (0, \frac{1}{3}]$, $\tau > \omega_*\left(\frac{\beta}{1-\beta}\right)$,
the accuracy parameter $\varepsilon > 0$.

$u := u^0$; $w := w^0$.
while $v_0 > \varepsilon$ **do**
  Predictor step:
    Compute Δu by solving system (2.4) with $a := \left(\frac{\|v\|^2}{n+1} - \rho(w)\right) e - 2v^2$.
    Compute $\alpha_p = \max\{\alpha \in (0,1) : (u + \alpha\Delta u, (1-\alpha)w) \in \mathcal{N}_\Psi(\tau)\}$.
    $u := u + \alpha_p \Delta u$; $w := (1 - \alpha_p)w$.
  while $z \notin \mathcal{N}_\delta(\beta)$ **do**
    Corrector step:
      Compute Δu by solving system (2.4) with $a_i := \rho(w) - r_i(z)$, $i = 1, \ldots, n$.
      $\alpha_c := \operatorname{argmin}\{F(u + \alpha\Delta u, w) : (u + \alpha\Delta u) \in \mathcal{F}^+,\ \alpha \in (0,1)\}$.
      $u := u + \alpha_c \Delta u$.
  end while
end while

In what follows, we make some observations regarding the algorithm. In the predictor step, we take a step in the tangent direction, but we stay in the neighbourhood $\mathcal{N}_\Psi(\tau)$. We take a greedy step, so the new point will be on the boundary of the neighbourhood $\mathcal{N}_\Psi(\tau)$, namely $\Psi(u+\alpha_p\Delta u,(1-\alpha_p)w)=\tau$. In the corrector stage, we may take several corrector steps until we enter the neighbourhood $\mathcal{N}_\delta(\beta)$. The determination of the corrector step is a one-dimensional optimization problem, which can be solved by the damped Newton method. It is enough to take only a few damped Newton steps in order to get a good enough step length. In Lemma 5.1, we will show that the corrector step length obtained by only one damped Newton step will provide an enough big decrease of the measure $\delta(z)$.

5. Complexity analysis. Algorithm 4.1 performs prediction steps. Each of the predictor steps is followed by a sequence of corrector steps with w-component of points z being unchanged. First of all, let us prove that such a sequence cannot be too long, by providing an upper bound on the number of the corrector steps.

5.1. Corrector stage. In the following lemma, we give an upper bound on the number of corrector steps inside a main iteration. Although the proximity measures are different than in [4], we follow the main steps of the proof given in Lemma 4.2 in [4].

LEMMA 5.1. *Let $z\in\mathcal{F}_z^+$ such that $\Psi(z)=\tau$, where $\tau>0$. Then, we need in each main iteration at most*

$$\left\lfloor \frac{\tau}{\omega\left(\frac{\beta}{\sqrt{1+\beta}}\right)}\right\rfloor+1 \tag{5.1}$$

corrector steps in order to have $z\in\mathcal{N}_\delta(\beta)$, where ω is defined in (1.1).

Proof. The aim of the corrector stage is to enter into the neighbourhood $\mathcal{N}_\delta(\beta)$, i.e. $\delta(z)\le\beta$. Hence, we want to give a lower bound on the decrease of $\delta(z)$ after a corrector step. For this, we consider only an approximate corrector step length. This is calculated by taking a single damped Newton step applied on the function $f(\alpha)=F(u+\alpha\Delta u,w)=F(z+\alpha\Delta z)$, where $\Delta z=(\Delta u,0)$. Using (2.4) with the right hand side given in (4.11), we get

$$\begin{aligned} r_0(z+\alpha\Delta z)&=r_0(z)+\alpha\rho(w),\\ r_i(z+\alpha\Delta z)&=(1-\alpha)r_i(z)+\alpha\rho(w)+\alpha^2\Delta x_i\Delta s_i,\quad i=1,\dots,n.\end{aligned}$$

Using (3.6) and (3.13), we have

$$\begin{aligned} f(\alpha)=&-(n+1)\ln\rho(w)-\ln\left[\hat r_0^2(z)+\alpha\left(1-\hat r_0^2(z)\right)\right]\\ &-\sum_{i=1}^n\ln\left[\hat r_i^2(z)+\alpha\left(1-\hat r_i^2(z)\right)+\frac{\alpha^2\Delta x_i\Delta s_i}{\rho(w)}\right]\end{aligned}$$

and

$$f'(\alpha)=-\frac{1-\hat r_0^2(z)}{\hat r_0^2(z)+\alpha\left(1-\hat r_0^2(z)\right)}-\sum_{i=1}^n\frac{1-\hat r_i^2(z)+2\alpha\frac{\Delta x_i\Delta s_i}{\rho(w)}}{\hat r_i^2(z)+\alpha\left(1-\hat r_i^2(z)\right)+\frac{\alpha^2\Delta x_i\Delta s_i}{\rho(w)}}.$$

Then, using (3.15), we have

$$f'(0)=-\sum_{i=0}^n\frac{1-\hat r_i^2(z)}{\hat r_i^2(z)}=e^T\left(e-\frac{1}{\hat r^2(z)}\right)=-\zeta_0^2(z). \tag{5.2}$$

Furthermore,

$$f''(\alpha) = \left(\frac{1-\hat{r}_0^2(z)}{\hat{r}_0^2(z)+\alpha\,(1-\hat{r}_0^2(z))}\right)^2 - \sum_{i=1}^{n} \frac{\frac{2\Delta x_i \Delta s_i}{\rho(w)}}{\hat{r}_i^2(z)+\alpha\,(1-\hat{r}_i^2(z)) + \frac{\alpha^2 \Delta x_i \Delta s_i}{\rho(w)}}$$
$$+ \sum_{i=1}^{n} \left(\frac{1-\hat{r}_i^2(z)+2\alpha\frac{\Delta x_i \Delta s_i}{\rho(w)}}{\hat{r}_i^2(z)+\alpha\,(1-\hat{r}_i^2(z)) + \frac{\alpha^2 \Delta x_i \Delta s_i}{\rho(w)}}\right)^2$$

and from (3.13) and (3.15), we have

$$f''(0) = \sum_{i=0}^{n} \left(\frac{1}{\hat{r}_i^2(z)} - 1\right)^2 - \sum_{i=1}^{n} \frac{2\Delta x_i \Delta s_i}{\rho(w)\hat{r}_i^2(z)} = \zeta_1^2(z) - \sum_{i=1}^{n} \frac{2\Delta x_i \Delta s_i}{r_i(z)}. \tag{5.3}$$

From (2.6) and (4.11), we derive

$$\frac{2}{\rho(w)} \sum_{i=1}^{n} |\Delta x_i \Delta s_i| \le \sum_{i=1}^{n} \frac{(r_i(z)-\rho(w))^2}{\rho(w) x_i s_i} \;\le\; \sum_{i=1}^{n} \frac{(r_i(z)-\rho(w))^2}{\rho(w) r_i(z)} \;\le\; \zeta_0^2(z). \tag{5.4}$$

Moreover, using $\Delta x^T \Delta s = 0$, (3.15) and (5.4), we obtain

$$-\sum_{i=1}^{n} \frac{2\Delta x_i \Delta s_i}{r_i(z)} = \frac{2}{\rho(w)} \sum_{i=1}^{n} \frac{r_i(z)-\rho(w)}{r_i(z)} \Delta x_i \Delta s_i$$
$$\le \zeta_0^2(z) \max_{1\le i\le n} \left|1 - \frac{\rho(w)}{r_i(z)}\right| \le \zeta_0^2(z)\zeta_1(z). \tag{5.5}$$

From (5.3) and (5.5), we derive the following upper bound on $f''(\alpha)$:

$$f''(0) \le \zeta_1^2(z) + \zeta_0^2(z)\zeta_1(z). \tag{5.6}$$

Thus, we can formulate the following lower bound for the Newton decrement of the function $f(\cdot)$:

$$\lambda := \sqrt{\frac{(f'(0))^2}{f''(0)}} \;\ge\; \frac{\zeta_0^2(z)}{\sqrt{\zeta_1^2(z)+\zeta_0^2(z)\zeta_1(z)}} \;=\; \frac{\delta(z)}{\sqrt{1+\delta(z)}} \;\ge\; \frac{\beta}{\sqrt{1+\beta}}, \tag{5.7}$$

since we are at the corrector stage, hence $\delta(z) \ge \beta$.

Note that the function $f(\cdot)$ is self-concordant. Therefore, from the general theory (e.g. Chapter 4 in [15]), we know that taking a single damped Newton step with $\alpha = -\frac{f'(0)}{(1+\lambda)f''(0)}$ ensures the following progress in the function value:

$$F(u,w) - F(u_+,w) \;\;=\;\; f(0) - f(\alpha) \ge \omega(\lambda) \;\ge\; \omega\left(\tfrac{\beta}{\sqrt{1+\beta}}\right),$$

where $z = (u,w)$ and $z_+ = (u_+,w)$ are two consecutive iterates of the inner loop.

In view of the assumption of the lemma, we have $\Psi(z) = \tau$. Hence, if we take k steps in the corrector stage, then we have

$$\tau \ge (k-1)\omega\left(\frac{\beta}{\sqrt{1+\beta}}\right),$$

which proves the lemma. □

In the next subsection, we deal with the analysis of the predictor stage.

5.2. Predictor step. The first lemma is a technical one which will be used several times in the next part of the analysis.

LEMMA 5.2. *Consider the function* $\Phi(\eta) := -\sum_{i=0}^{n} \ln(1+\eta_i)$. *If* $\|\eta\| < 1$ *and* $\bar{e}^T\eta = 0$, *then*

$$\Phi(\eta) \leq \omega_*(\|\eta\|), \tag{5.8}$$

where ω_* *is defined in (1.2).*

Proof. Using that the function Φ is self-concordant and the assumption $\|\eta\| < 1$, we can use Theorem 5.1.9 of [17]. Since $\nabla\Phi(0) = \bar{e}$ and $\nabla^2\Phi(0) = I$, where I is the $(n+1)$-dimensional identity matrix and by the assumption of the lemma $\bar{e}^T\eta = 0$, we obtain

$$\Phi(\eta) \leq \Phi(0) + \nabla\Phi(0)^T\eta + \omega_*(\|\eta\|) \;=\; \omega_*(\|\eta\|). \tag{5.9}$$

Hence, we get the desired inequality. ∎

The condition for entering the predictor step in Algorithm 4.1 is $z \in \mathcal{N}_\delta(\beta)$. Let us derive some of its consequences.

LEMMA 5.3. *Let* $z \in \mathcal{N}_\delta(\beta)$, *with* $\beta < \frac{1}{2}$. *Then,* $\zeta_0(z) \leq \frac{\beta}{\sqrt{1-\beta}}$ *and*

$$1-\beta \;\leq\; \hat{r}_i^2(z) \;\leq\; \frac{1}{1-\beta}, \quad i = 0, \ldots, n, \tag{5.10}$$

$$1-\beta \;\leq \frac{x_i s_i}{x_i(w)s_i(w)} \leq\; \frac{1}{1-\beta}, \quad i = 1, \ldots, n. \tag{5.11}$$

Proof. Denote by $\hat{r}_{\min} = \min_{0\leq i\leq n} \hat{r}_i(z)$ and $\hat{r}_{\max} = \max_{0\leq i\leq n} \hat{r}_i(z)$. In view of equality (3.5) and using the definition of $\hat{r}(z)$ in (3.13), we have $\hat{r}_{\min} \leq 1 \leq \hat{r}_{\max}$. By the definition of δ (3.16) and the assumption of the lemma $\zeta_0^2(z) \leq \beta\zeta_1(z)$, we have

$$\zeta_0^2(z) = \sum_{i=0}^{n} \left(\hat{r}_i(z) - \frac{1}{\hat{r}_i(z)}\right)^2 \;\leq\; \beta \left[\sum_{i=0}^{n} \left(1 - \frac{1}{\hat{r}_i(z)}\right)^2\right]^{1/2} \;\leq\; \beta\frac{1}{\hat{r}_{\min}}\zeta_0(z). \tag{5.12}$$

Hence, $0 \leq \frac{1}{\hat{r}_{\min}} - \hat{r}_{\min} \leq \frac{\beta}{\hat{r}_{\min}}$, and we get $\hat{r}_i^2(z) \geq 1-\beta$, for all $i = 0, \ldots, n$. Thus, using this and (5.12), we have $\zeta_0(z) \leq \frac{\beta}{\sqrt{1-\beta}}$. This means that $0 \leq \hat{r}_{\max} - \frac{1}{\hat{r}_{\max}} \leq \frac{\beta}{\sqrt{1-\beta}}$ and we get $\hat{r}_i^2(z) \leq \frac{1}{1-\beta}$.

For proving the inequalities (5.11), note that from (5.10), we have

$$x_i s_i - v_i^2 \;=\; r_i(z) \;\geq\; (1-\beta)\rho(w) \;=\; (1-\beta)\left[x_i(w)s_i(w) - v_i^2\right], \quad i = 1, \ldots, n.$$

Thus, $x_i s_i \geq (1-\beta)x_i(w)s_i(w) + \beta v_i^2 \geq (1-\beta)x_i(w)s_i(w)$. The remaining inequalities can be proven in the same way. ∎

In the following lemma, we show that with a predictor step, we reach the boundary of the neighbourhood $\mathcal{N}_\Psi(\tau)$.

LEMMA 5.4. *Let* $z \in \mathcal{N}_\delta(\beta)$ *with* $\beta < \frac{1}{2}$ *and let* $\tau > \omega_*\left(\frac{\beta}{1-\beta}\right)$. *Then, the predictor step length* $\alpha_p \in (0,1)$ *is well defined and* $\Psi(z(\alpha_p)) = \tau$ *holds.*

Proof. Note that the initial value $\Psi(z)$ can be represented as follows:

$$\Psi(z) \;\; = \;\; -\sum_{i=0}^{n} \ln\left(\hat{r}_i^2(z)\right), \; i = 0, \ldots, n.$$

We use Lemma 5.2 for $\eta = \hat{r}^2(z) - \bar{e}$. For this, we want to verify whether the assumptions of the lemma hold. First, denote by $\zeta_2(z) = \|\hat{r}^2(z) - \bar{e}\|$. Using $\delta(z) \leq \beta < \frac{1}{2}$, from Lemma 5.3, we have

$$\zeta_2(z) \;\; \leq \;\; \tfrac{\beta}{1-\beta} \; < \; 1. \tag{5.13}$$

On the other hand, from (3.14) we have $\bar{e}^T\left(\hat{r}^2(z) - \bar{e}\right) = 0$. Then, using Lemma 5.2, we obtain

$$\Psi(z) = \Phi\left(\hat{r}^2(z) - \bar{e}\right) \; \leq \omega_*(\zeta_2(z)) \; \leq \; \omega_*\left(\frac{\beta}{1-\beta}\right) \; < \; \tau, \tag{5.14}$$

where we used the assumption of the lemma and that the function ω^* is monotone increasing. Thus, z is in the interior of $\mathcal{N}_\Psi(\tau)$. Hence, there exists a positive predictor step length with which we stay in the neighbourhood $\mathcal{N}_\Psi(\tau)$. On the other hand, using the definition of the functional proximity measure given in (3.18) and the barrier property of F, we have $\Psi(z(\alpha_p)) = \tau$. ∎

In the following lemma, we derive an exact expression for the proximity measure $\Psi(\cdot)$ along the predictor direction given in (4.10). Consider

$$d(\alpha) = r(z(\alpha)) - \rho(w(\alpha))\bar{e} \in \mathbb{R}^{n+1}. \tag{5.15}$$

Lemma 5.5. *Consider the predictor search direction $\Delta z = (\Delta u, -w)$, i.e., $z(\alpha) = (u(\alpha), w(\alpha)) = (u + \alpha\Delta u, (1-\alpha)w)$, where $\alpha \in (0,1)$. Then,*

$$\Psi(z(\alpha)) \; = \; -\sum_{i=0}^{n} \ln\left(1 + \frac{1}{\rho(w(\alpha))} d_i(\alpha)\right), \tag{5.16}$$

where $d(\alpha) = r(z) - \rho(w)\bar{e} + \alpha^2 g(z)$, with $\bar{e} = (1, \ldots, 1)^T \in \mathbb{R}^{n+1}$ and

$$\begin{aligned} g_0(z) \;\; &= \;\; \tfrac{1}{n+1}\|v\|^2, \\ g_i(z) \;\; &= \;\; \Delta x_i \Delta s_i - v_i^2 + \tfrac{1}{n+1}\|v\|^2, \quad i = 1, \ldots, n. \end{aligned}$$

Proof. Using (3.18) we have

$$\Psi(z(\alpha)) = -\sum_{i=0}^{n} \ln r_i(z(\alpha)) + (n+1)\ln\rho(w(\alpha)) = -\sum_{i=0}^{n} \ln\left(1 + \frac{d_i(\alpha)}{\rho(w(\alpha))}\right). \tag{5.17}$$

We analyse $d(\alpha)$. Firstly, we have

$$\begin{aligned} \rho(w(\alpha)) &= \frac{1}{n+1}\left((1-\alpha)v_0 - (1-\alpha)^2\|v\|^2\right) = \frac{1-\alpha}{n+1}\left(v_0 - \|v\|^2 + \alpha\|v\|^2\right) \\ &= (1-\alpha)\left(\rho(w) + \frac{\alpha}{n+1}\|v\|^2\right). \end{aligned} \tag{5.18}$$

On the other hand, using (4.10), we get for $i = 1, \ldots, n$

$$\begin{aligned} r_i(z(\alpha)) = x_i(\alpha)s_i(\alpha) - (v_i(\alpha))^2 &= (x_i + \alpha\Delta x_i)(s_i + \alpha\Delta s_i) - (1-\alpha)^2 v_i^2 \\ &= r_i(z) - \alpha\left(2v_i^2 + \rho(w) - \frac{1}{n+1}\|v\|^2\right) + \alpha(2-\alpha)v_i^2 + \alpha^2\Delta x_i\Delta s_i \\ &= r_i(z) - \alpha\left(\rho(w) - \frac{1}{n+1}\|v\|^2\right) + \alpha^2\left(\Delta x_i\Delta s_i - v_i^2\right). \end{aligned} \tag{5.19}$$

Thus, combining (5.18) with (5.19), we have for $i = 1, \ldots, n$

$$\begin{aligned} d_i(\alpha) &= r_i(z(\alpha)) - \rho(w(\alpha)) \\ &= r_i(z) - \rho(w) + \frac{\alpha}{n+1}\|v\|^2 + \alpha^2\left(\left(\Delta x_i\Delta s_i - v_i^2\right) - \frac{\alpha(1-\alpha)}{n+1}\|v\|^2\right. \\ &= r_i(z) - \rho(w) + \alpha^2\left(\left(\Delta x_i\Delta s_i - v_i^2 + \frac{1}{n+1}\|v\|^2\right)\right). \end{aligned} \tag{5.20}$$

Similarly,

$$\begin{aligned} r_0(z(\alpha)) &= (1-\alpha)v_0 - (s+\alpha\Delta s)^T(x+\alpha\Delta x) \\ &= r_0(z) - \alpha\left(v_0 + \Delta s^T x + s^T\Delta x\right) \\ &= r_0(z) - \alpha\Big[v_0 - 2\|v\|^2 - n\left(\rho(w) - \frac{1}{n+1}\|v\|^2\right)\Big] \\ &= r_0(z) - \alpha\left(\rho(w) - \frac{1}{n+1}\|v\|^2\right). \end{aligned} \tag{5.21}$$

Hence, from (5.18) and (5.21), we get

$$d_0(\alpha) = r_0(z(\alpha)) - \rho(w(\alpha)) = r_0(z) - \rho(w) + \frac{\alpha^2}{n+1}\|v\|^2. \tag{5.22}$$

From (5.17), (5.20) and (5.22), we obtain the result of the lemma. ∎

In the following lemma we give a bound on the predictor step length α_p. Following the idea presented in [4], for $w \in \mathcal{F}_w^+$ we define

$$\underline{\alpha}(w) = \frac{v_0}{\|v\|^2} > 1. \tag{5.23}$$

The geometric interpretation of $\underline{\alpha}(w)$ can be found in Section 5 of [4].

LEMMA 5.6. *Let $z \in \mathcal{N}_\delta(\beta)$ with $\beta \le \frac{1}{3}$ and $\tau = \omega_*(\nu)$ for some $\nu \in \left(\frac{\beta}{1-\beta}, 1\right)$. Then, the predictor step length α_p satisfies the following conditions:*

$$0 \;<\; \nu - \tfrac{\beta}{1-\beta} \;\le\; \nu\alpha_p + (n+1)\left(\tfrac{\underline{\alpha}(w)}{\underline{\alpha}(w)-1}\right)^2\alpha_p^2. \tag{5.24}$$

Proof. We give an indirect proof. For the sake of simplicity, in the proof we use the notation $\alpha = \alpha_p$. Assume that

$$\nu\alpha + (n+1)\left(\frac{\underline{\alpha}(w)}{\underline{\alpha}(w)-1}\right)^2\alpha^2 < \nu - \frac{\beta}{1-\beta}. \tag{5.25}$$

We want to give an upper bound on $\Psi(z(\alpha))$ by using Lemma 5.2 for $\eta = \frac{d(\alpha)}{\rho(w(\alpha))}$. For this, we check the assumptions of the lemma. We consider the vector $g(z)$ defined in Lemma 5.5. Using (3.15), we get $\left(r(z) - \rho(w)\bar{e}\right)^T \bar{e} = 0$. Moreover,

$$g(z)^T \bar{e} \;=\; \tfrac{1}{n+1}\|v\|^2 + \sum_{i=1}^{n} \Delta x_i \Delta s_i - \|v\|^2 + \tfrac{n}{n+1}\|v\|^2 \;=\; 0.$$

Using these equations and Lemma 5.5, we obtain $\bar{e}^T d(\alpha) = 0$. Furthermore, we have to prove that $q(\alpha) := \frac{1}{\rho(w(\alpha))}\|d(\alpha)\| < 1$. Let us derive an upper bound for this value. From (5.18) we have $\rho(w(\alpha)) \geq (1-\alpha)\rho(w)$. Using this and (5.13) we have

$$\begin{aligned} q(\alpha) &\leq \frac{1}{(1-\alpha)\rho(w)}\left(\|r(z) - \rho(w)\bar{e}\| + \alpha^2\|g(z)\|\right) \;=\; \frac{1}{1-\alpha}\left(\zeta_2(z) + \frac{\alpha^2}{\rho(w)}\|g(z)\|\right) \\ (5.26) &\leq \frac{1}{1-\alpha}\left(\frac{\beta}{1-\beta} + \frac{\alpha^2}{\rho(w)}\|g(z)\|\right). \end{aligned}$$

It remains to bound $\|g(z)\|$. Note that $g(z) = g^{\Delta}(z) + g^{v}(z)$ with

$$\begin{aligned} g_0^{\Delta}(z) \;&=\; 0, \quad g_i^{\Delta}(z) \;=\; \Delta x_i \Delta s_i,\; i = 1, \ldots, n, \\ g_0^{v}(z) \;&=\; \tfrac{1}{n+1}\|v\|^2, \quad g_i^{v}(z) \;=\; \tfrac{1}{n+1}\|v\|^2 - v_i^2,\; i = 1, \ldots, n. \end{aligned}$$

Therefore, $\|g(z)\| \leq \|g^v(z)\| + \|g^{\Delta}(z)\|$, with

$$\|g^v(z)\|^2 \;=\; \tfrac{1}{n+1}\|v\|^4 - \tfrac{2}{n+1}\|v\|^4 + \sum_{i=1}^{n}(v_i)^4 \;\leq\; \|v\|^4.$$

At the same time, from (2.6), (4.10) and Lemma 5.3 we derive

$$\begin{aligned} 2\|g^{\Delta}(z)\| \;&\leq\; \sum_{i=1}^{n} \frac{1}{x_i s_i}\left[2x_i(w)s_i(w) - \frac{1}{n+1}v_0\right]^2 \\ &\leq\; \frac{1}{1-\beta}\sum_{i=1}^{n} \frac{1}{x_i(w)s_i(w)}\left[2x_i(w)s_i(w) - \frac{1}{n+1}v_0\right]^2 \\ &= \frac{1}{1-\beta}\left[4s(w)^T x(w) - 4\frac{n}{n+1}v_0 + \frac{1}{(n+1)^2}(v_0)^2 \sum_{i=1}^{n}\frac{1}{x_i(w)s_i(w)}\right] \\ &\leq\; \frac{1}{1-\beta}\left[4(v_0 - \rho(w)) - 4\frac{n}{n+1}v_0 + \frac{n}{(n+1)^2\rho(w)}(v_0)^2\right] \\ &=\; \frac{1}{1-\beta}\left[\frac{4\|v\|^2}{n+1} + \frac{n(v_0)^2}{(n+1)^2\rho(w)}\right] =\; \frac{\|v\|^2}{(1-\beta)(n+1)}\left[4 + \frac{n\underline{\alpha}(w)^2}{\underline{\alpha}(w) - 1}\right], \end{aligned}$$

where in the last inequality we used (3.12) and $x_i(w)s_i(w) = v_i^2 + \rho(w) \leq \rho(w)$. Thus, we conclude that

$$\begin{aligned} \|g(z)\| &\leq \|v\|^2 + \frac{\|v\|^2}{(1-\beta)(n+1)}\left[2 + \frac{n\underline{\alpha}(w)^2}{2(\underline{\alpha}(w) - 1)}\right] \\ &\leq \|v\|^2\left[1 + \frac{\underline{\alpha}(w)^2}{2(1-\beta)(\underline{\alpha}(w) - 1)}\right] \;\leq\; \frac{\underline{\alpha}(w)^2}{\underline{\alpha}(w) - 1}\|v\|^2. \end{aligned} \tag{5.27}$$

Note that the last inequality follows from the condition $\beta \leq \frac{1}{3}$. From (5.26) and (5.27), using that $\frac{1}{\rho(w)}\|v\|^2 = \frac{n+1}{\underline{\alpha}(w)-1}$, we come to the following bound:

$$q(\alpha) \;\leq\; \tfrac{1}{1-\alpha}\left[\tfrac{\beta}{1-\beta} + \alpha^2(n+1)\left(\tfrac{\underline{\alpha}(w)}{\underline{\alpha}(w)-1}\right)^2\right]. \tag{5.28}$$

Then, by the indirect assumption (5.25) we have

$$q(\alpha) < \nu < 1. \tag{5.29}$$

Hence, we can use Lemma 5.2 for $\eta = \frac{d(\alpha)}{\rho(w(\alpha))}$ and we get

$$\Psi(z(\alpha)) \;=\; \Phi\left(\tfrac{d(\alpha)}{\rho(w(\alpha))}\right) \;\leq\; \omega_*(q(\alpha)). \tag{5.30}$$

From Lemma 5.4 we have $\Psi(z(\alpha)) = \tau$. Using the assumption of the lemma and (5.30) we have $\omega_*(\nu) = \tau = \Psi(z(\alpha)) \leq \omega_*(q(\alpha))$. From the monotone increasing property of the function ω_*, we get $q(\alpha) \geq \nu$, which contradicts to (5.29). Therefore, (5.25) is not true, which proves the lemma. ∎

In accordance with the general scheme for analyzing PTS methods proposed in [4], we can establish a global convergence rate of the process of Algorithm 4.1 in terms of the merit function

$$\mu^*(w) = \frac{\underline{\alpha}(w)}{\underline{\alpha}(w)-1}v_0 = \frac{v_0^2}{v_0 - \|v\|^2} \;\geq\; v_0 \;\geq\; \|v\|^2. \tag{5.31}$$

In the following lemma, we give an upper bound on the new value of the merit function after a predictor step.

LEMMA 5.7. *Let $\alpha_p \in (0,1)$ be a feasible step length in the predictor step of the PTS IPA. If $\alpha_p \geq \gamma\,\frac{\underline{\alpha}(w)-1}{\underline{\alpha}(w)}$ with some $\gamma \in (0,1)$, then $\mu^*(w(\alpha_p)) < \frac{1}{1+\gamma}\,\mu^*(w)$.*

Proof. For the sake of simplicity, we use the notation $\alpha = \alpha_p$. By the assumption of the lemma and using that $\underline{\alpha}(w) > 1$,

$$\frac{1}{1-\alpha} = \frac{1+\alpha}{1-\alpha^2} > 1 + \alpha \geq 1 + \gamma\,\frac{\underline{\alpha}(w)-1}{\underline{\alpha}(w)}.$$

Since $\underline{\alpha}(w(\alpha)) = \frac{1}{1-\alpha}\,\underline{\alpha}(w)$,

$$\begin{aligned}\frac{\mu^*(w(\alpha))}{\mu^*(w)} &= \frac{\frac{1}{1-\alpha}\,\underline{\alpha}(w)}{\frac{1}{1-\alpha}\,\underline{\alpha}(w)-1}\,(1-\alpha)\,v_0\,\frac{\underline{\alpha}(w)-1}{\underline{\alpha}(w)v_0} = \frac{\underline{\alpha}(w)-1}{\frac{1}{1-\alpha}\,\underline{\alpha}(w)-1} \\ &< \frac{\underline{\alpha}(w)-1}{\left(1+\gamma\,\frac{\underline{\alpha}(w)-1}{\underline{\alpha}(w)}\right)\underline{\alpha}(w)-1} = \frac{1}{1+\gamma},\end{aligned}$$

hence the lemma is proven. ∎

Now we are ready to give the complexity result of the PTS IPA.

THEOREM 5.8. *Let $z \in \mathcal{N}_\delta(\beta)$ with $\beta \leq \frac{1}{3}$ and $\tau > \omega_*\left(\frac{\beta}{1-\beta}\right)$ and let $z_0 = (u_0, w_0) \in \mathcal{F}_z^+$ be the starting point. Then, the PTS IPA gives a feasible solution of the problem (P)-(D) with $x^T s \leq \varepsilon$ after at most*

$$O\left(\sqrt{n}\ln\tfrac{\mu^*(w^0)}{\varepsilon}\right) \tag{5.32}$$

numbers of iterations.

Proof. Based on Lemma 5.7, if $\pi = \frac{\underline{\alpha}(w)}{\underline{\alpha}(w)-1}\alpha_p \geq \gamma$, then we have $\mu^*(w(\alpha_p)) < \frac{1}{1+\gamma}\mu^*(w)$. In order to apply this result to our case, note that $\frac{\underline{\alpha}(w)}{\underline{\alpha}(w)-1} > 1$. Hence, using (5.24), for the value π we get the following lower bound:

$$\nu - \tfrac{\beta}{1-\beta} \quad \leq \quad \nu\pi + (n+1)\pi^2. \tag{5.33}$$

Hence, $\pi = \Omega\left(\frac{1}{\sqrt{n}}\right)$, thus for Algorithm 4.1 we can take $\gamma = O\left(\frac{1}{\sqrt{n}}\right)$. Since the lengths of corrector stages are uniformly bounded by inequality (5.1), we conclude that the total number of steps of this method for obtaining a feasible solution of the problem (P)-(D) with $x^T s \leq \varepsilon$, cannot be bigger than (5.32). □

In the following section the obtained numerical results are presented.

6. Preliminary computational experiments. In this section, we present the preliminary numerical results obtained by using our new IPA in PTS. An advantage of this method is that it can start from an arbitrary strictly feasible primal-dual point. Using (4.1) we can compute the starting value of w.

From the computational point of view, in the predictor step, we do not determine the best step length. However, starting from the maximum feasible step length α, we used the bisection technique to get an α for which $|\Psi(z(\alpha)) - \tau| \leq 0.1\tau$.

By theory, in the corrector stage, we might need several corrector steps. In the complexity analysis, in the proof of Lemma 5.1 we showed that with only a single damped Newton step applied to the univariate function $f(\cdot)$ we obtained a good enough corrector step length. However, in the implementation, we performed several damped Newton steps applied to $f(\cdot)$ in order to determine the corrector step length. In this way, for all considered test problems we needed only one corrector step after each predictor step.

In order to compute the search directions, we used Cholesky decomposition for the Newton system (2.4) with $\frac{1}{2}nm(m+1) + \frac{1}{6}m(m+1)(m+2)$ multiplications. Note that we need to compute different right-hand sides a for the predictor and corrector steps. This computation needs only $O(n)$ operations.

For our computational experiments, we used a simple random generator working as follows.

- Generate a strictly feasible primal-dual pair of points $(\hat{x}, \hat{s})$. Their entries are uniformly distributed in the interval $(0, 1)$.
- Generate matrix $A \in \mathbb{R}^{m\times n}$ with entries uniformly distributed in $(-1, 1)$.
- Set $b = A\hat{x}$ and $c = \hat{s}$.
- The starting point u^0 for Algorithm 4.1 is defined as $(\hat{x}, 0, \hat{s})$.

In Table 1 we present preliminary computational results for random problems of small and medium dimensions with $32 \leq m \leq \frac{n}{2}$ and $64 \leq n \leq 1024$. In our experiments with Algorithm 4.1, we used $\beta = \frac{1}{4}$ and $\tau = 1$.

In each cell, we put the average number of predictor steps of Algorithm 4.1 required for reaching the accuracy $\varepsilon = 10^{-8}$ in the duality gap. Our results correspond to the series of random test problems of length one hundred. The second value in the cell corresponds to the relative standard deviation.

In our opinion, these results are very promising. As we have already mentioned, in all our experiments, any predictor step is followed only by one corrector step. Hence, we do not display the number of corrector steps. The growth of the number of predictor steps is clearly proportional to $\ln(nm)$. A quite accurate prediction of this

m \ n	64	128	256	512	1024
32	13.6 ± 9.9%	15.4 ± 8.5%	17.0 ± 8.9%	18.8 ± 7.0%	21.2 ± 7.2%
64		17.0± 9.1%	18.8 ± 7.2%	21.0 ± 6.9%	23.0 ± 6.3%
128			20.7 ± 6.3%	22.9 ± 5.6%	25.2 ± 5.7%
256				25.1 ± 5.9%	27.9 ± 4.7%
512					30.1 ± 4.6%

TABLE 1
Average number of iterations for random LO problems

number is given by the model

$$k \;\approx\; 1 + 2\log_2 \tfrac{mn}{32}. \tag{6.1}$$

For data in Table 1, the average absolute deviation in this forecast is 0.29 iterations.

From Lemma 5.7 we have

$$\mu^*(w^k) \le \left(\frac{1}{1+\gamma}\right)^k \mu^*(w^0) \le \mu^*(w^0)\exp\left(-\frac{\gamma\, k}{1+\gamma}\right).$$

Therefore, the rate of convergence of our method is $\exp\left(-\frac{\gamma\, k}{1+\gamma}\right)$, where $\gamma = O\left(\frac{1}{\sqrt{n}}\right)$ and k is the number of predictor steps. But of course, its practical performance is much better. It depends logarithmically on the dimension and has very weak dependence on the required accuracy.

Indeed, with our test problems, we observed a very fast local convergence. In Table 2, we present the progress in the objective function v_0 in terms of the predictor steps for a random problem with $n = 512$, $m = 256$, and $\varepsilon = 10^{-8}$. The third column of Table 2 shows the size of the predictor step relative to the maximal possible step up to the boundary of the set $\mathcal{F}^+$. Thus, our definition (3.18) of the functional neighbourhood of greedy trajectory allows really large steps.

Iteration	v_0	% of max step
1	$1.2 \cdot 10^{2}$	31.25
2	$1.1 \cdot 10^{2}$	46.88
3	$1.0 \cdot 10^{2}$	53.13
...	...	...
14	$1.0 \cdot 10^{0}$	78.13
...	...	...
20	$6.7 \cdot 10^{-3}$	84.37
21	$2.3 \cdot 10^{-3}$	93.75
22	$5.7 \cdot 10^{-4}$	95.31
23	$8.1 \cdot 10^{-5}$	95.31
24	$2.3 \cdot 10^{-6}$	99.02
25	$1.9 \cdot 10^{-9}$	99.97

TABLE 2
Progress in v_0 and the size of the predictor step relative to the maximal feasible step length

A similar explosion of the convergence rate in the end of the process can be seen in all our experiments. It would be very interesting to support our observations by a convincing theoretical explanation.

7. Conclusion. In this paper, we introduced an IPA for LO in PTS based on a universal tangent direction. The IPA can start at any strictly feasible primal-dual pair and go directly towards a solution by a predictor-corrector scheme. The concept of universal tangent direction yields to an easier computation of the search directions than in the initial algorithm [16]. We proved that the complexity of the proposed method coincides with the currently known best complexity results for IPAs.

Our preliminary computational results are very promising. At the end of the process, the size of the predictor step typically goes above 99.97% of the maximal possible step to the boundary. The method demonstrates a very fast local convergence.

As future research, it would be interesting to extend this new IPA in PTS to Linear Complementarity Problems and to Semidefinite Programming. Furthermore, it would be interesting to analyse the effect of considering other schemes instead of the greedy one in $\mathcal{F}_w^+$. This would yield to different types of central paths.

Acknowledgments. The research was partially supported by the Austrian-Hungarian Action Foundation, project number 116öu8. Petra Renáta Rigó is thankful for the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the PD_22 funding scheme, project no. 142154. The research of Marianna E.-Nagy was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES

[1] W. Ai and S. Zhang, *An $O(\sqrt{n}L)$ iteration primal-dual path-following method, based on wide neighborhoods and large updates, for monotone LCP*, SIAM J. Optim., 16 (2005), pp. 400–417.

[2] Y. Bai, M. El Ghami, and C. Roos, *A new efficient large-update primal-dual interior-point method based on a finite barrier*, SIAM J. Optim., 13 (2003), pp. 766–782.

[3] Zs. Darvay, *New interior point algorithms in linear programming*, Adv. Model. Optim., 5 (2003), pp. 51–92.

[4] M. E.-Nagy, T. Illés, Yu. Nesterov, and P. Rigó, *Parabolic target space interior-point algorithms for weighted monotone linear complementarity problem*, Tech. Report 4, Corvinus Econ. Work. Paper, 2024.

[5] M. E.-Nagy and A. Varga, *A long-step interior point framework and a related function class for linear optimization*, Tech. Report 1, Corvinus Econ. Work. Paper, 2024.

[6] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, 1969.

[7] 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.

[8] T. Illés, P. Rigó, and R. Török, *Unified approach of interior-point algorithms for $P_*(\kappa)$-LCPs using a new class of algebraically equivalent transformations*, J. Optim. Theory Appl., 202 (2024), pp. 27–49.

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

[10] L. Khachiyan, *A polynomial algorithm for linear programming*, Soviet Math. Dokl., 20 (1979), pp. 191–194.

[11] G. Lešaja and C. Roos, *Unified analysis of kernel-based interior-point methods for $P_*(\kappa)$-linear complementarity problems*, SIAM J. Optim., 20 (2010), pp. 3014–3039.

[12] S. Mehrotra, *On the implementation of a primal-dual interior point method*, SIAM J. Optim., 2 (1992), pp. 575–601.

[13] S. Mizuno, M. Todd, and Y. Ye, *On adaptive-step primal-dual interior-point algorithms for linear programming*, Math. Oper. Res., 18 (1993), pp. 964–981.

[14] A. Nemirovski and D. Yudin, *Optimization methods adapting to the significant dimension of the problem*, Automatika i Telemekhanika, 38 (1979), pp. 75–87. translated in Automation and Remote Control 38(4), 513-524, 1983.

[15] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Springer, New York, 2004.

[16] Yu. Nesterov, *Parabolic target space and primal–dual interior-point methods*, Discret. Appl. Math., 156 (2008), pp. 2079–2100. In Memory of Leonid Khachiyan (1952 - 2005).
[17] Yu. Nesterov, *Lectures on Convex Optimization*, Springer Optimization and Its Applications, Springer, Switzerland, 2018.
[18] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 2006.
[19] J. Peng, C. Roos, and T. Terlaky, *A new class of polynomial primal-dual methods for linear and semidefinite optimization*, Eur. J. Oper. Res., 143 (2002), pp. 234–256.
[20] F. Potra, *Corrector-predictor methods for monotone linear complementarity problems in a wide neighborhood of the central path*, Math. Program., 111 (2008), pp. 243–272.
[21] J. Renegar, *A polynomial-time algorithm, based on Newton's method, for linear programming*, Math. Program., 40 (1988), pp. 59–93.
[22] C. Roos, T. Terlaky, and J.-P. Vial, *Interior Point Methods for Linear Optimization*, Springer, New York, USA, 2005.
[23] Gy. Sonnevend, *An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming*, in System Modelling and Optimization: Proceedings of the 12th IFIP-Conference held in Budapest, Hungary, September 1985, A. Prékopa, J. Szelezsán, and B. Strazicky, eds., vol. 84 of Lecture Notes in Control and Information Sciences, Springer Verlag, Berlin, West-Germany, 1986, pp. 866–876.
[24] M. Todd, *Potential-reduction methods in mathematical programming*, Math. Program., 76 (1997), pp. 3–45.
[25] L. Tuncel and M. Todd, *On the interplay among entropy, variable metrics and potential functions in interior-point algorithms*, Comput. Optim. Appl., 8 (1997), pp. 5–19.